European Commission

Towards a European strategy on **business-to-government data sharing for the public interest**

Final report
prepared by the High-Level Expert Group on Business-to-Government Data Sharing

The High-Level Expert Group on Business-to-Government (B2G) data sharing was an independent expert group set up by the European Commission in November 2018. Its mandate ended in December 2019.

Contact: Alberto Gago-Fernandez (Policy coordinator - B2G data sharing)
*Email: CNECT-B2G-DATASHARING@ec.europa.eu*Luxembourg: Publications Office of the European Union, 2020

© European Union, 2020

Reproduction is authorised provided the source is acknowledged.

The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

For any use or reproduction of photos or other material that is not under the copyright of the European Union, permission must be sought directly from the copyright holders.

Print ISBN 978-92-76-11423-9 doi:10.2759/406717 KK-01-19-750-EN-C
PDF ISBN 978-92-76-11422-2 doi:10.2759/731415 KK-01-19-750-EN-N

Towards a European strategy on **business-to-government data sharing for the public interest**

Final report
prepared by the High-Level Expert Group
on Business-to-Government Data Sharing

FOREWORD BY ROBERTO VIOLA, DIRECTOR GENERAL FOR DG COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY, EUROPEAN COMMISSION

The European Union is undergoing a digital transition that is changing our societies and economies at an unprecedented speed. More than 33 zettabytes (ZB) ($\times 10^{21}$) of data was generated worldwide in 2018, a number that is set to grow to 175 ZB by 2025. This creates an extraordinary opportunity for Europe to use this enormous amount of data yet to be created and lead the data revolution on the world stage. It will deeply transform the way we communicate, live and work. It will be used to train the artificial intelligence (AI) of the future. Its insights will make us lead longer and healthier lives, develop more efficient solutions to tackle climate change, improve business operations and reduce energy consumption: the list is endless. Making the data economy work for Europe is, therefore, a top priority for the Commission.

One of the areas where data can make a difference is the delivery of public services. Access to the right data can help us, for example, radically improve public transport, make cities greener and cleaner, tackle epidemics and develop better policies by making them more evidence-based. Indeed, business-to-government (B2G) data sharing for the public interest can be a game-changer for improving general welfare.

We therefore asked a group of independent experts to reflect on how to ensure more of the data held by the private sector could be used by public administrations for the common good. I would like to thank the 23 experts who participated in this initiative. This report summarises their work and contains key recommendations that will contribute to making B2G data sharing a responsible, sustainable and scalable practice in the EU.

The report calls on the Commission, the Member States and all stakeholders to take the necessary steps — in our different roles — to make more data available and increase its reuse for the common good.

FOREWORD BY ALBERTO ALEMANNO, RAPPORTEUR OF THE HIGH-LEVEL EXPERT GROUP ON BUSINESS-TO-GOVERNMENT DATA SHARING

Data is increasingly regarded as an essential infrastructural resource for economic growth, innovation and the overall well-being of society. Thanks to our enhanced capability to collect, process and use data, we are able to know where the epicentre of an earthquake is, how to limit a pandemic such as Zika from spreading or even how to reduce pollution in cities. This data 'revolution' must be contextualised within a broader, 3-century-long attempt to neutralise irrationality in human decision-making by gaining increased access to information.

Yet, while virtually every organisation today, including small companies or grassroots movements, is a data entity, only a few of them (generally in the private sector) have collected vast amounts of data (be it personal or non-personal) and acquired a unique capability to make sense of such information. As a result, most of this much-prized data is in the hands of businesses, not of public authorities, with the latter lagging behind in embracing the power of data to inform their daily policies and service-delivery actions.

Hence the challenge — entrusted by the European Commission to our expert group — to explore the creation of an enabling environment for privately held data to be shared with (or at least be accessible to) public authorities in complying with their public-interest missions. For the time being, most of the efforts focus on getting the supply-side ready (e.g. by developing data-sharing models, de-risking data sharing or reskilling personnel) for B2G data sharing, but omit the need to sensitise and prepare the demand side (the public authorities). More critically, at a time in which reputation-based mechanisms transcend the mere financial rating and venture into state-driven 'social credit' scoring, the role that citizens play in the data-sharing equation remains limited and ancillary to most of the solutions at stake.

This report, which has benefited from the experience and dedication of 23 experts coming from different walks of life, strives to address both of these systemic shortcomings. By carving out a role to play for both the public sector and citizens, it offers an initial set of ethically-aware paths the EU might chart to break new ground in advancing and accompanying B2G data sharing.

The hope is that by following some of this expert advice the EU might become a global leader in fostering not only a market for B2G data access, but also a sensible, inclusive and participatory data culture through a set of viable, practicable and scalable welfare-enhancing solutions. Only a more cautious, humble and humanised approach to the unprecedented amount of data we produce and collect every day may pave the way for a future in which it is the human factor that defines our daily life experiences.

EXECUTIVE SUMMARY

The challenges our society currently faces, ranging from natural disasters to traffic congestion, are of a scale and complexity that traditional policy tools cannot always address. However, today's digital revolution offers several innovative tools that can be used by governments in providing for the general welfare of their people. Data is one of these, and its potential to not only provide for new scientific insights, but also to inform policymaking and deliver better public services to the people is invaluable. If used effectively, data can become an enabler for a better society and a more efficient public sector.

Yet, much of the potential for data and its insights to be used for the benefit of society remains untapped. Not only because the vast majority of data is in the hands of the private sector, but also because the public sector does not seem ready to realise the full potential of data. Due to organisational, technical and legal obstacles (as well as an overall lack of a data-sharing culture) business-to-government (B2G) data-sharing partnerships are still largely isolated, short-term collaborations.

This report provides a detailed overview of these barriers and proposes a comprehensive framework of policy, legal and funding recommendations to enable scalable, responsible and sustainable B2G data sharing for the public interest.

The high-level expert group explored the state of play across the EU, and identified the lack of governance structures and dedicated professionals in the field as well as certain economic barriers as key obstacles to the scaling-up of B2G data sharing. Furthermore, it noted the rapid emergence of sectoral data-sharing legislation in certain Member States, but not in others. This increases fragmentation within the EU's internal market, which in turn nurtures growing uncertainty on the rules and procedures governing B2G data sharing. To overcome this, the expert group recommends the following:

- Member States should put in place national governance structures that support B2G data sharing;
- a recognised data steward function should be created and promoted in both the public and private sectors. The European Commission should encourage the creation of a network of such data stewards, as a community of practice in the field;
- B2G data-sharing collaborations should be organised:
  - in testing environments ('sandboxes') for pilot testing ('pilots') to help assess the potential value of data for new situations in which a product or service could potentially be used ('use cases'),

  - via public-private partnerships.
- the European Commission should explore the creation of an EU regulatory framework providing a minimum level of harmonisation for B2G data-sharing processes;
- in acquiring privately held data for public-interest purposes, preferential conditions may apply in line with the updated B2G data-sharing principles.

Mechanisms are not yet in place to ensure the accountability, transparency and compliance with ethical principles when a B2G data-sharing collaboration is established. Furthermore, the general public is not empowered to share its data such that it can be used to address the societal challenges of its choice. Public awareness on the benefits of B2G data sharing is limited. Moreover, the public sector is slower to embrace the digital transformation as compared to the private sector, and public-sector workers often do not have the knowledge or skills to process data. In light of this, the expert group recommends that:

- all those involved are transparent on the B2G data-sharing collaborations in which they engage, including the data used and the impact of the collaboration;
- the general public is made aware of the societal benefits of data sharing and is involved in the choice of public-interest challenges to be addressed through such collaborations;
- the general public is encouraged to share their data for public-interest purposes and, to facilitate this, user-friendly data-donation mechanisms should be created and promoted;
- the European Commission explores whether to develop ethical guidelines on data use, including for the public interest, and taking into account the *Ethics guidelines for trustworthy AI*;
- Member States foster a data-literate public sector, by investing in the training and reskilling of policymakers and public-sector workers.

The expert group emphasises that trust between the private and public sectors as well as the general public is key if B2G data sharing is to become a reality. However, operational models and secure technical systems to enable safe and trusted data sharing are as yet underdeveloped. In addition, while B2G data sharing frequently requires cross-sectoral and/or cross-border datasets to be combined, datasets are often not interoperable or of sufficient quality. The expert group therefore recommends that:

- the European Commission and Member States explore incentivising mechanisms, such as public-recognition programmes, to increase B2G data sharing on a voluntary basis;

- the EU's Horizon Europe ([1]) and Digital Europe ([2]) funding programmes invest in the development and deployment of the technologies necessary for scalable, responsible and sustainable B2G data sharing (e.g. privacy-preserving technologies), the prioritisation of standards and the setting-up of pilots in regulatory sandboxes for specific societal challenges;
- the European Commission carries out studies to obtain further empirical evidence of the macroeconomic and social benefits of B2G data sharing for the public interest.

In addition, the High-Level Expert Group revised the Commission's principles on private-sector data sharing in B2G contexts (published in April 2018) and included one new principle on accountability and one on fair and ethical data use. These updated principles should form the backbone of B2G data sharing, whether required or voluntary.

Finally, the experts put forward 12 examples of existing good practices and five pledges to engage in new B2G data-sharing collaborations. The latter will serve as important use cases to test the revised principles and illustrate further good practice.

([1]) https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en.

([2]) https://ec.europa.eu/digital-single-market/en/news/digital-europe-programme-proposed-eu92-billion-funding-2021-2027.

CONTENTS

Chapter 1: B2G data sharing for the public interest . . . 13

1.1. Why B2G data sharing for the public interest? . . . 13

1.2. How data serves the public interest . . . 17

1.3. How privately held data can be accessed by the public sector . . . 19

1.4. What types of data may be subject to B2G data sharing? . . . 22

1.5. State of play of B2G data sharing for the public interest . . . 24

1.5.1. Organisational challenges . . . 25

1.5.2. Governance . . . 25

1.5.3. Absence of a data-sharing culture . . . 26

1.5.4. Operational and technical challenges . . . 27

1.6. Towards a European strategy on B2G data sharing for the public interest . . . 28

Chapter 2: Governance of B2G data sharing across the EU . . . 31

2.1. Issues . . . 31

2.1.1. Lack of structures to support B2G data sharing for the public interest . . . 31

2.1.2. Shortage of dedicated functions in the field . . . 33

2.1.3. Economic barriers preventing B2G data sharing . . . 33

2.1.4. Fragmented landscape for data-sharing rules . . . 35

2.2. Solutions . . . 37

2.2.1. National structures to support B2G data sharing . . . 37

2.2.2. Recognising and establishing champions of data sharing: data stewards . . . 37

2.2.3. Incentives for the private sector . . . 39

2.2.4. Create legal certainty in B2G data sharing for the public interest . . . 41

Chapter 3: Transparency, citizen engagement and ethics on B2G data sharing . . . 49

3.1. Issues . . . 49

3.1.1. Lack of transparency and awareness . . . 49

3.1.2. Citizen engagement in B2G data-sharing relations is underdeveloped . . . 51

3.1.3. Absence of a coherent ethical approach to B2G data sharing . . . 52

3.1.4. Need for digital skills to foster B2G data-sharing collaborations . . . 53

3.2. Solutions **54**
3.2.1. Ensure transparency and awareness on B2G data sharing 54
3.2.2. Foster and improve citizen participation 55
3.2.3. Develop a data ethics framework 56
3.2.4. Digital skills to foster B2G data sharing 58

Chapter 4: Operational models, structures and technical tools to facilitate trusted data sharing . . . 59
4.1. Issues **59**
4.1.1. Absence of operational models and dedicated structures to scale up B2G data sharing 59
4.1.2. Technical challenges in B2G data sharing 61
4.2. Solutions **67**
4.2.1. Scalable B2G data sharing and incentives for data providers 67
4.2.2. Investing in technical means for B2G data sharing 70

Key recommendations **75**

Revised principles on business-to-government data sharing for the public interest **79**

Experts' pledges **87**

Annex I: Data taxonomy **93**

Annex II: Good practices **95**

Annex III: Members of the High-Level Expert Group **107**

Annex IV: Mandate of the High-Level Expert Group **109**

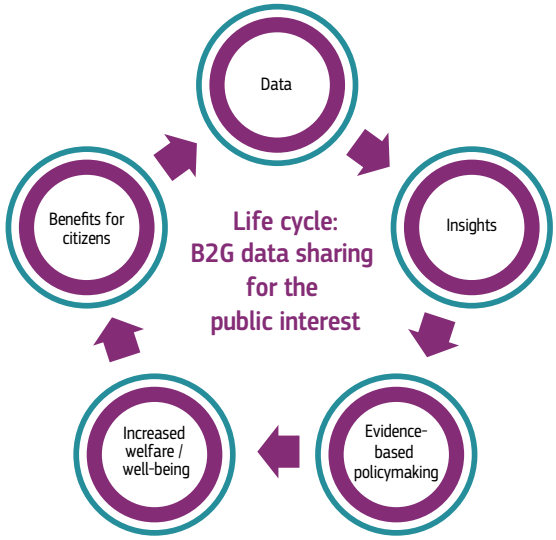
CHAPTER 1: B2G DATA SHARING FOR THE PUBLIC INTEREST

1.1. WHY B2G DATA SHARING FOR THE PUBLIC INTEREST?

Today a growing number of societal challenges (such as climate change, natural disasters, urban planning and pandemics) are not only extremely complex, but also interrelated. Data represents a key raw material to deal with such challenges.

The huge amount of data produced every day can reveal real-time information that is critical to understanding patterns of human behaviour and activities. In turn, these insights ultimately allow both the private and public sectors to take better decisions ([3]).

Data in general, and privately held data in particular, has a high potential to serve the general public interest by informing decision-making, providing for new scientific insights and resolving policy issues, thus enabling more-targeted interventions and improving public-service delivery, amongst other possibilities ([4]). This can also bring about significant savings for the public budget.



For example, data collected from mobile phones, social media, digital transactions, global positioning system (GPS) devices and other sensors provide an evidence source that, when acted upon, can provide for sustainable regional and urban planning, environmentally friendly transport and energy systems, saving lives in humanitarian crises or improving education.

([3]) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Towards a common European data space* (COM(2018) 232 final of 25.4.2018).

([4]) 'Enter the data economy. EU policies for a thriving data ecosystem', *EPSC Strategic Notes*, Issue 21, 11 January 2017, European Political Strategic Centre, 2017, (https://ec.europa.eu/epsc/publications/strategic-notes/enter-data-economy_en).

More specifically, data from sensors in cities can provide insights to predict tourist inflows, estimate pollution and understand traffic flows. Exchanges of data on transportation and cargo can ensure near-frictionless border control. This data can also serve to develop AI solutions that benefit society.

B2G data sharing and personal data

With the GDPR ([5]) and ePrivacy ([6]) legislation, the EU has put in place a solid and trusted legal framework for the protection of personal data. Any B2G data-sharing partnership involving personal data must be carried out in full compliance with this legislation.

Certain types of data, such as behavioural data (e.g. mobile phone records, GPS location or social media data (which can be crucial to better understand population movements, lifestyle changes, disease patterns and habits) are largely in the hands of the private sector. While, through the open data policy ([7]) the public sector is required to make its data available, there is no similar policy through which the public sector can access or reuse private-sector data.

Yet, the public sector is tasked with the responsibility of resolving societal challenges and ensuring the overall welfare of the general public. To achieve this goal, it would significantly benefit from becoming more data driven and cost-efficient. However, public-sector bodies ([8]) generally lag behind in the data revolution as compared to the private sector ([9]). Furthermore, they often cannot access such data or insights, for example due to the lack of a mandate.

There are already several successful examples of private companies and civil-society organisations ([10]) that have entered into

([5]) Regulation (EU) 2016/679 of the European Parliament and of the 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 (General Data Protection Regulation (GDPR)) (OJ L 119, 4.5.2016, pp. 1-88).

([6]) 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) (OJ L 201, 31.7.2002, pp. 37-47).

([7]) Notably, the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public-sector information (OJ L 172, 26.6.2019, pp. 56-83), also known as the 'Open Data Directive', aims to make more public-sector information available and reusable. It replaced the Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public-sector information (OJ L 345, 31.12.2003, pp. 90-96), also known as the 'PSI Directive'. Directive 2013/37 /EU of the European Parliament and of the Council of 26 June 2013 (OJ L 175, 27.6.2013, pp. 1-8) subsequently amended the latter.
The Tallin Ministerial Declaration on eGovernment states that 'eGovernment is significant for the development of the data economy and the Digital Single Market, especially for ensuring the secure and free movement of data as an enabler for digital innovation in Europe and for reducing the costs of and barriers to seamless functioning of the Single Market', (https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration).

([8]) In this report, 'public-sector bodies' refers to European, state, regional or local authorities or other bodies governed by public law.

([9]) For this reason, the Commission has put forward an EU eGovernment action plan 2016-2020 on accelerating the digital transformation of government, (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0179&from=EN).

([10]) A growing number of civil-society organisations such as non-governmental organisations (NGOs) and philanthropic foundations are collecting more and more data and, as such, they should be considered when discussing the B2G data-sharing phenomenon. As a result, the term 'business', as used in the term 'business-to-government (B2G)' in this report, refers not only to the private sector but also to a broader category, encompassing civil-society organisations, such as NGOs and philanthropic foundations.

collaborations with the public sector under non-commercial (e.g. corporate social responsibility (CSR) arrangements) or commercial terms([11]). Private sector data has been used, for instance, to improve disaster relief or traffic congestion in major cities.

For example, Telefónica, signed an agreement with the United Nations (UN) Food and Agriculture Organisation (FAO) in 2017 that envisaged data-related initiatives to help quantify the rate of forced migration in Colombia due to extreme drought because of climate change. The extreme drought caused farmers to lose their livelihoods and left them with no option other than to move to other places ([12]). Vodafone developed a pioneering programme to use aggregated anonymised mobile data to help governments track trends in population movements and control epidemics so as to prevent widespread outbreaks ([13]). In addition, Twitter collaborated with the United Nations Children's Fund (UNICEF) to track anti-vaccination sentiment in eastern European social media networks and to develop specific recommendations for improving messaging strategies across languages, platforms and conversation themes ([14]). In Amsterdam a public-private data task force ([15]) was established in 2017 to deploy data sharing for safety-related data in real-life situations. Its goal is to stimulate the exchange and sharing of safety-related traffic information between vehicle manufacturers, service providers and Member States, in an architecture that will ultimately allow cross-border exchanges to foster pan-European solutions and interoperability.

Yet, much of the potential for private-sector data and insights to be used by public-sector bodies to tackle societal challenges remains untapped. Recent developments (generally referred to as 'B2G data sharing for the public interest', but also 'data-driven social partnerships', 'data collaboratives' ([16]), 'data trusts' or 'data philanthropy' ([17])) are premised on an emerging consensus about the need to unlock privately held data to maximise the power of data to

([11]) For a list of incentives for private companies to engage in data sharing with the public sector, see Klein, T., and Verhulst, S., 'Access to new data sources for statistics: business models and incentives for the corporate sector', *OECD Statistics Working Papers*, No 6, 5 May 2017, OECD Publishing, Paris, 2017, (https://doi.org/10.1787/9a1fa77f-en).

([12]) 'FAO and Telefónica to boost use of cutting-edge digital technologies to assist farmers in developing countries', 12 February 2017, FAO News, 2017, (http://www.fao.org/news/story/it/item/1099923/icode).

([13]) Rana, N., and Majmudar, U., 'Data can be a force for social good: Nuria Oliver, Vodafone', *Responsible Future*, 4 October 2018, The Economic Times Blogs, 2018, (https://economictimes.indiatimes.com/blogs/ResponsibleFuture/data-can-be-a-force-for-social-good-nuria-oliver-vodafone).

([14]) Verhulst, S., and Young A., 'The potential of social media intelligence to improve people's lives', GovLab Report, 24 September 2017, The Governance Lab, 2017, (https://www.thegovlab.org/static/files/publications/social-media-data.pdf).

([15]) Data Task Force, on Data for Road Safety website (https://www.dataforroadsafety.eu/data-task-force); 'On our way towards connected and automated driving in Europe', 15 February 2017, Government of the Netherlands, Amsterdam, 2017, (https://www.government.nl/documents/leaflets/2017/05/18/on-our-way-towards-connected-and-automated-driving-in-europe).

([16]) Verhulst, S., and Sangokoya, D., 'Data collaboratives: exchanging data to improve people's lives', *Medium,* 22 April 2015, (https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a).

([17]) Kirkpatrick, R., 'A new type of philanthropy: donating data', *Harvard Business Review*, 21 March 2013, (https://hbr.org/2013/03/a-new-type-of-philanthropy-don).

deliver transformative change in a collaborative and inclusive way ([18]).

However, this presupposes the identification and qualification of what 'public interest' means. This concept allows for limitations on the exercise of fundamental rights or the exceptions ('derogations') to competition law that might be needed to foster data-sharing collaborations. While 'public interest' broadly refers to the welfare of individuals in society, its exact boundaries remain largely undefined, being heavily dependent on socioeconomic, cultural and historical factors. While some purposes (such as public health) are generally recognised as being in the public interest, other purposes (such as improved public services) might be less clear cut. Therefore, instead of trying to define public interest in a 'one-size fits all' manner, the expert group acknowledged the context-specific nature of this concept and identified instead a set of criteria to help qualify whether a given purpose is in the public interest. These are inspired, for example, by the concept of services of general economic interest that was developed in EU law and clarified by the Court of Justice of the European Union ([19]). Although the notion of public interest may include law-enforcement purposes, such purposes are beyond the mandate of the group. Needless to say, any B2G data-sharing collaboration should fully respect EU values and principles.

The value of the data economy in the EU

The value of the EU data economy was almost EUR 301 billion in 2018, representing 2.4% of gross domestic product (GDP). A recent European data market study ([20]) estimates that if the policy and legislative framework conditions around the data economy are put in place in time, the value of the data economy will increase to over EUR 829 billion by 2025, representing 5.8% of the overall EU GDP. Furthermore, the use of data increases the productivity of companies in all sectors of the economy and can help tackle societal challenges.

([18]) For a contrarian view, Elkin-Koren, N., and Gal, M., 'The chilling effect of governance-by-data on innovation', *University of Chicago Law Review*, Vol. 86, 30 July 2018, (https://ssrn.com/abstract=3213175).

([19]) Protocol No 26 to the Treaty on the Functioning of the European Union (TFEU) concerns Services of General Interest (SGEIs), but does not define the concept. The Commission has clarified the concept in its quality framework (COM/2011/0900 final, p. 3), where it explains that they are services that public authorities of the Member States at national, regional or local level classify as being of general interest and, therefore, subject to specific public-service obligations. The Court has established that the concept may apply to different situations and terms, depending on the Member State, and EU law does not create any obligation to formally designate a task or a service as being of general economic interest, except when such obligation is laid out in EU legislation (e.g. universal service in the postal and telecommunication sectors). Cases C-179/90 *Merci convenzionali porto di Genova* [1991] ECR I-5889, paragraph 27; Case C-242/95 *GT-Link A/S* [1997] ECR I-4449, paragraph 53; and Case C-266/96 *Corsica Ferries France SA* [1998] ECR I-3949, paragraph 45. For more information, (https://ec.europa.eu/competition/state_aid/overview/public_services_en.html).

([20]) *D2.6 Second interim report. The European data market monitoring tool: key facts and figures, policy conclusions, data landscape and quantified stories*, 28 June 2019, IDC Italia srl and The Lisbon Council, (http://datalandscape.eu/sites/default/files/report/D2.6_EDM_Second_Interim_Report_28.06.2019.pdf).

1.2. HOW DATA SERVES THE PUBLIC INTEREST

A dataset may become particularly valuable when it is (re)used together with other datasets, providing insights for decision-making processes or the development of services. Data is unlike any other resource ([21]). The non-rivalrous and non-excludable nature of data is the fundamental economic driver of socioeconomic welfare gains in data-sharing operations: many parties can use and reuse the same dataset for a variety of purposes and an unlimited number of times, without any loss in its quality or quantity ([22]). This results in substantial cost savings for society as data need be collected only once, and more than one innovative output can be produced with the very same set of data: including applications that the original data collector may not have envisaged. Data sharing can now become a reality thanks to today's technological developments, for example big data analytics and AI.

Economists consider data as an 'experience good' — in other words, its value is unknown until it has been used for a particular purpose. When used for a different purpose, its value may not be the same, in particular because the real value of data does not come from a single dataset, but from combining datasets from different sources. This may render cost-benefit analyses on the value of data challenging to undertake.

While a specific dataset may give only an incomplete or even a biased picture, combining the data from several providers may allow for more refined and accurate insights to be deduced, thus avoiding biased conclusions from non-representative data. For example, combining location data with rich socioeconomic profile data from social media services ([23]) may generate a better picture of the number of tourists in a city than that emerging from the dataset of a single credit card company. Therefore, by enabling datasets from different origins to be combined, data

([21]) *Data-driven innovation: big data for growth and well-being*, OECD, 2015, (https://www.oecd.org/sti/ieconomy/data-driven-innovation.htm).

([22]) 'Data-driven innovation: big data for growth and well-being', *OECD Publishing*, Paris, 2015, (https://www.oecd.org/sti/ieconomy/data-driven-innovation.htm); Duch-Brown, N., Martens, B., and Mueller-Langer, F., 'Data ownership, access and trade', *JRC Digital Economy Working Paper*, No 1, 2017, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf); Jones, C. I., and Tonetti, C., 'Non-rivalry and the economics of data', *Working Paper*, No 3716, August 2019, Standford Graduate School of Business, 2019, (https://www.gsb.stanford.edu/faculty-research/working-papers/nonrivalry-economics-data).

([23]) Verhulst, S., and Young, A., 'How the data that internet companies collect can be used for the public good', *Harvard Business Review*, 23 January 2018, (https://hbr.org/2018/01/how-the-data-that-internet-companies-collect-can-be-used-for-the-public-good).

sharing may produce strong social welfare benefits ([24]).

Using vessel traffic data for the production of official statistics

In 2018, Statistics Netherlands (CBS) and MarineTraffic [provider of ship-tracking and maritime intelligence] signed a memorandum of understanding to share vessel-tracking data and knowledge to improve the quality of existing statistics and foster the production of new statistical products useful to policymakers. The data was shared, at no cost, directly between the parties involved. It was anonymised and aggregated data of vessel activities, which had been collected through the automatic identification system terrestrial networks of receivers MarineTraffic owns and operates. The idea behind this pilot was to investigate the possibilities for using vessel-tracking data as a fast economic indicator and a proxy for international trade. ([25])

The socioeconomic value of a B2G data-sharing operation thus depends on the value of the decisions or public services produced with such data or the economic savings achieved after resolving a particular societal challenge. Benefits can also be related to citizen empowerment through increased transparency; addressing new problems that require multidisciplinary/multi-stakeholder solutions; and research and development. The use of data has great potential to benefit society, while the costs remain small in comparison ([26]). Thus if the potential of private-sector data is not fully realised, this could lead to missed opportunities to improve people's lives ([27]).

'... data as an infrastructural resource that can be used by an unlimited number of users and for an unlimited number of purposes as an input to produce goods and services.' ([28])

With these characteristics in mind, data emerges as an infrastructural resource ([29]) needed to tackle some of the most pressing challenges that we face as a species, ranging from population ageing, chronic diseases, epidemics and natural disasters to congested roads, to name but a few. For example, the lifestyle data necessary to tackle health-related challenges is often gathered by private

([24]) Duch-Brown, N., Martens, B., and Mueller-Langer, F., 'Data ownership, access and trade', *JRC Digital Economy Working Paper*, No 1, 2017, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf).

([25]) Excerpt from the good practice shared by Dimitris Zissis in Annex II.

([26]) The JRC concludes that while we cannot concretely estimate the value of these welfare gains for society we can safely assume that there are significant gains from data-sharing operations. Martens, B. and N. Duch-Brown (2020), The economics of Business to Government data sharing, JRC Digital Economy Working Paper 2020-04, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf).

([27]) Alemanno, A., 'Data for good: unlocking privately held data to the benefit of the many', *European Journal of Risk Regulation*, Vol. 9. No 2, June 2018, pp. 183-191, https://doi.org/10.1017/err.2018.34.

([28]) *Data-driven innovation: big data for growth and well-being*, OECD, 2015, (https://www.oecd.org/innovation/data-driven-innovation-9789264229358-en.htm).

([29]) *Data-driven innovation: big data for growth and well-being*, OECD, 2015, (https://www.oecd.org/innovation/data-driven-innovation-9789264229358-en.htm).

companies (e.g. mobility and consumption patterns). Data from both the mobile phone network infrastructure and GPS location can be helpful in modelling human mobility, which is valuable for managing public health or responding to natural disasters. As such, the combination of public- and private-sector data is considered pivotal in resolving some of the most pressing and long-standing societal issues, such as the challenges that the UN's sustainable development goals (SDGs) are designed to tackle ([30]).

To sum up, data can serve the public interest in at least five different ways:

1. to improve situational awareness, i.e. what do we know about a given phenomenon or reality;
2. to better understand the causes and variables behind the current situation;
3. to more accurately predict and forecast;
4. to run more rigorous impact assessments and evaluations (of any intervention) in order to better define the policy problem and identify the most effective policy options ([31]); and
5. to guide public management decisions taken either by humans or automated processes.

1.3. HOW PRIVATELY HELD DATA CAN BE ACCESSED BY THE PUBLIC SECTOR

Currently, various methods exist that allow private companies and civil-society organisations to share data, or make it available for reuse, with the public sector ([32]). If data access is conceptualised on a continuum from closed to open, the range of existing public-sector accessibility methods can be described according to where they are on such a spectrum, each making more or less data available to those authorised. Going from the 'most restricted' to the 'most accessible', the following can be presented as illustrative non-exhaustive examples (see also Section 4.1.1.1 for more detailed operational models and 4.1.2.1 for technical approaches ([33])).

([30]) 'Transforming our world: the 2030 agenda for sustainable development, A/RES/70/1', United Nations, 2015, (https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf).

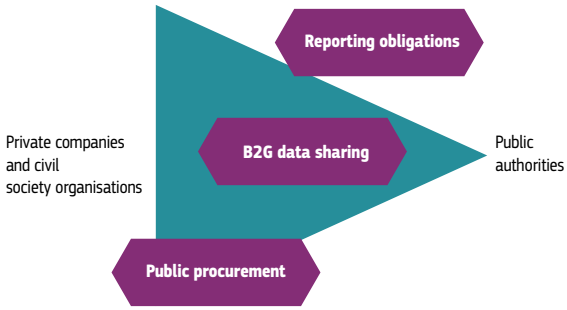
([31]) Verhulst, S., Young, A., and Winowatan, M., and Zahuranec, a., J., *Data collaboratives. Leveraging private data for public good*, May 2019, The Governance Lab (GovLab), (https://datacollaboratives.org/static/files/existing-practices-report.pdf).

([32]) It is important to note that data is not always handed over ('shared'), i.e. a data transfer takes place.

([33]) de Montjoye, Y., Gambs, S., Blondel, V. et al., 'On the privacy-conscientious use of mobile phone data', *Scientific Data*, No 5, 11 December 2018, (https://www.nature.com/articles/sdata2018286.pdf).

1. A data provider ([34]) conducts all the data analysis in-house and then shares with public authorities, or the general public, the insights that emerge from that analysis.
2. A data provider can host external researchers within its network to analyse data and report findings publicly.
3. A data provider works with other data providers to share data or insights derived from the data amongst themselves, and also with public authorities.
4. A data provider shares data with trusted third parties.
5. A data provider allows direct access to some of its datasets.

Through these mechanisms, data can be shared either on a voluntary basis or to comply with certain legal obligations. For instance, reporting obligations can mandate businesses to share certain data that is necessary for the public sector to carry out its functions on a continuous basis and for free ([35]). These reporting obligations can range from financial reporting, to environmental or social reporting. Sometimes they are used to monitor the implementation of, or compliance with, existing legislation. Generally, this data cannot be reused or combined with other datasets for other purposes. Moreover, in the absence of reporting obligations, some public administrations may acquire data through public procurement procedures (e.g. studies or data to inform policymaking) or as a by-product of the service procured (e.g. data generated while implementing a service contract). The introduction of data-sharing obligations as part of subcontracted services is not yet widely used by the public sector. However, some cities (e.g. Barcelona (Spain) and Eindhoven (Netherlands)) include clauses in their tenders that require companies to give the data produced during the implementation of such services ([36]) to the city authorities.



The public procurement of data as such (i.e. not as a by-product of a procured service) — whether raw, pre-processed or processed data, or data-driven insights — is often not the most appropriate or cost-efficient approach to acquire data. First, in several cases, access to numerous datasets from different companies over time is required in order to obtain

([34]) Data providers in B2G data interactions are typically private companies, but could also include civil-society organisations.

([35]) For certain reporting obligations, sometimes the public sector reimburses certain costs incurred by the business for the sharing of the data where such an obligation is put only on one individual company and this could affect its competitive position vis-à-vis other companies in the same sector.

([36]) Francesca Bria, interviewed by Thomas Graham for *Wired UK*: 'We are introducing clauses into contracts, like data sovereignty and public ownership of data. For example, now we have a big contract with Vodafone, and every month Vodafone has to give machine-readable data to city hall. Before, that didn't happen. They just took all the data and used it for their own benefit', (https://www.wired.co.uk/article/barcelona-decidim-ada-colau-francesca-bria-decode).

meaningful value and to ensure unbiased public policies ([37]). For example, if a national environment agency decides to study the increased mortality of bees, it will need data from a large number of sources, including data from beekeepers, pesticides sellers, land owners, agricultural data and weather data. Multiple parallel procurement procedures cannot possibly ensure that all of these datasets are accessed in time, not to mention the high cost of each individual procured dataset that in itself is not useful unless combined with other datasets. If one or several of those potential sources rejects access to its data there is a direct impact on the usefulness of all procured datasets, resulting in an important societal loss. Second, given that data is an 'experience good', its value in helping to solve the intended public purpose cannot necessarily be estimated *ex ante* ([38]). This inherently prevents data providers and the public sector from fixing a fair price for the data. Furthermore, data markets are currently underdeveloped or are in their infancy ([39]). In this context, there is no guarantee that a significant number of well-qualified, interested contractors would respond to the tender or that the tenders would be of a sufficient size to justify the resource-intensive procurement procedures. If there is no active market for the type of data that public-sector bodies seek to reuse, public procurement is a difficult endeavour. Finally, if the data sought is unique or very specific and there is only one provider ([40]), this provider could set pricing conditions beyond what is acceptable to the public sector. The provider's denial to give access to such data would cause social welfare losses ([41]).

For all of these reasons, the procurement of data is frequently an ineffective and inefficient mechanism to enable B2G data sharing for public-interest purposes.

Against this background, B2G data sharing for the public interest emerges as a novel form of collaboration between data providers and data users that can take shape along any of the five data accessibility mechanisms listed above.

Conceptually, data sharing fits in between the two previously mentioned accessibility instruments used by the public to gain access to privately held data: reporting obligations

([37]) This is either due to their inherent complex purpose or to the methodological need to ensure representativeness of the results.

([38]) The actual value of data is not based on the data itself but on the outcomes of the potential combinations of that data with other data sources for a specific purpose. Having this in mind, the usefulness of data for the intended public-interest purpose cannot be guaranteed *ex ante*, which makes public procurement procedures delicate as well. If that data is finally not useful, there is a clear societal loss. Furthermore, without knowing that usefulness, governments find it difficult to set a fair maximum price in the calls for tender.

([39]) Very few data market places exist today (e.g. Dawex, Nallian, Deutsche Telekom's Data Intelligence Hub) while data brokers usually act as intermediaries providing processing services when it comes to commercial offerings of data.

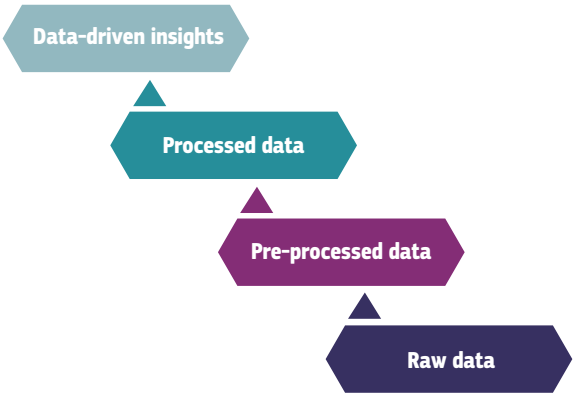
([40]) For example, a city authority may request access to private-room-reservation data from an online booking platform that operates in the city for sustainable tourism. Every booking platform has exclusive access to its own room-reservation data and therefore enjoys a monopolistic position.

([41]) This has happened on several occasions to the JRC, Martens, B. and N. Duch-Brown (2020), The economics of Business to Government data sharing, JRC Digital Economy Working Paper 2020-04, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf).

and public procurement. B2G data sharing offers great flexibility to both parties to the collaboration. As such, it can entail a mere one-off collaboration or a longer-term partnership with potentially different monetary and non-monetary compensation modalities. At the same time, B2G data sharing goes beyond reporting obligations per se, insofar as it provides additional help to public-sector bodies in their task of ensuring general welfare and tackling major societal challenges, without however entailing the constraints related to public procurement.

1.4. WHAT TYPES OF DATA MAY BE SUBJECT TO B2G DATA SHARING?

B2G data-sharing collaborations cover data that is already collected by all kinds of private companies (e.g. supermarkets, retailers, digital platforms, telecommunication companies) and civil-society organisations (NGOs, philanthropic foundations) for internal business purposes or for developing future products or services (e.g. personalised services). As such, B2G data sharing does not exclude personal data. Personal data must be collected, processed and anonymised in full compliance with the GDPR (see Annex I: Data taxonomy). Data portability coupled with appropriate consent or other legal grounds for processing allow for personal data to be included in the scope of a policy on B2G data sharing for the public interest and may open the door to data donation (with the specific consent of the individual person (the 'data subject')). For data that is subject to intellectual property rights (e.g. portraits, books) or database rights, the relevant legal framework needs to be respected ([42]).



When it comes to the level of detail and granularity of data required to render B2G data sharing meaningful for public authorities, highly aggregated data or insights could be appropriate or sufficient ([43]). However, for some purposes (e.g. statistics), the public

([42]) Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (OJ L 77, 27.3.1996, pp. 20-28), (whenever a right of intellectual property protects the data).

([43]) For example, in a study by the Spanish national office of statistics (INE) with three mobile operators, data was not shared but insights derived from the data were. Maqueda, A., 'El INE seguirá la pista de los móviles de toda España durante ocho días', *El País*, 29 October 2019, (https://elpais.com/economia/2019/10/28/actualidad/1572295148_688318.html).

sector would need to examine more-granular data to ensure, audit or verify that the insights provided by the private sector are not biased, discriminatory or altered in any way. A data taxonomy has been developed to categorise the different types of data that can be shared in a B2G data-sharing initiative (see Annex I: Data taxonomy).

Data taxonomy

In this report, the term 'data' refers to data at all levels of abstraction, from raw data to pre-processed, processed data and insights derived from the data, depending on the use case, the type of data, etc. (for further details, see Annex I: Data taxonomy).

Raw data: data collected from a source (e.g. numbers, instrument readings, images, text, videos, sensor data).

Pre-processed data: pre-processing includes, for example, cleaning, instance selection, re-sampling, normalisation, transformation, feature extraction and selection.

Processed data: data processing can be described as the manipulation of data in order to produce meaningful information.

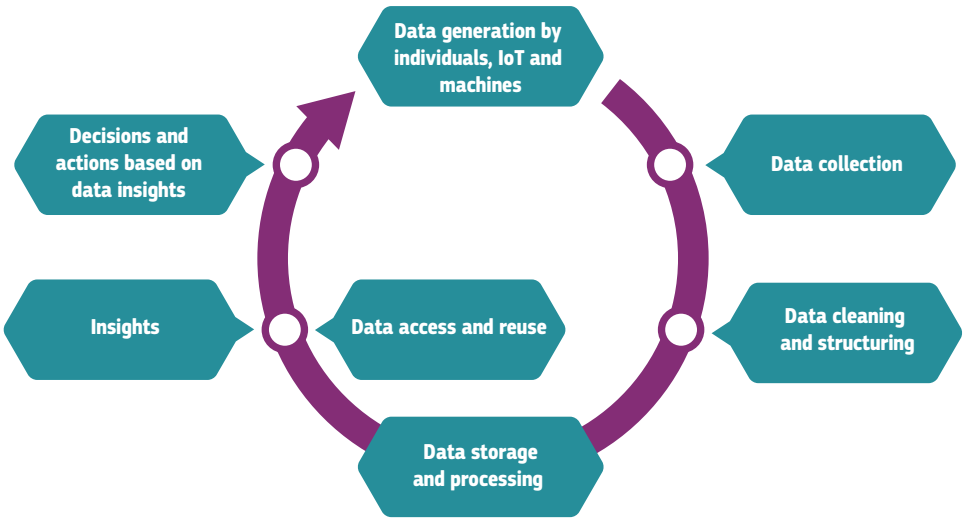
Data-driven insights: are generated by drawing conclusions from processed, analysed data.

All levels of abstraction need to be handled in full compliance with privacy legislation.

Data that may be rendered accessible to public authorities through B2G data sharing should not be reused for purposes related to law enforcement, such as countering financial crime or other law-enforcement driven endeavours. As such, should public-sector bodies receive this type of data from data providers in the framework of a data-sharing collaboration, they should abstain from using it for law-enforcement or taxing purposes against the data providers or individuals.

The focus of B2G data sharing is on datasets that the public sector either does not have access to or has no way to collect, in addition to situations where the sharing of such data would prevent the duplication of effort and investment in data collection and storage. Mobile phone data, data from social media companies or from Internet of things (IoT) devices, for instance, are usually unique datasets that are mainly collected by the private sector and can offer valuable insights that can be translated into policymaking. B2G data sharing should also take into account the investments made by data providers in data collection, storage and aggregation.

1.5. STATE OF PLAY OF B2G DATA SHARING FOR THE PUBLIC INTEREST



While studies have assessed the economic and social impact of the use of public-sector data for the public interest (e.g. open data) ([44]), similar studies are less numerous in relation to B2G data sharing. Despite the perceived value of B2G data sharing, the practice remains limited due to a variety of obstacles. As such, the full value of data for society remains unfulfilled. This section, after exploring what prevents privately held data from being shared, identifies what could be done about it by anticipating some of the major recommendations made in the following chapters of this report.

While there are common challenges to both sides of the data-sharing equation, some are specific to the data provider and others to the data user.

Overall, an entangled set of regulatory, economic, organisational, technical, social and ethical risks may dissuade potential partners from (or at least do not incentivise) entering into data-sharing collaborations.

Those risks manifest themselves all along the data life cycle: that is, all along the sequence of stages that a particular unit of data goes through from its initial generation or capture

([44]) For example, the study 'Creating value through open data: study on the impact of reuse of public data resources' prepared by Capgemini Consulting as part of the European Data Portal, 2015, (https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf).

to its eventual archiving and/or deletion at the end of its useful life ([45]). It is crucial to understand the nature of the barriers that separate currently feasible from unfeasible B2G data-sharing operations. In so doing, one may distinguish at least four different categories of obstacles to data collaborations: organisational, governance, cultural and techno-operational.

1.5.1. Organisational challenges

From an organisational perspective, a data provider must weigh the benefits against the costs and risks of a given data-sharing initiative. Significant cost factors (see Chapter 2) include finding a suitable data-sharing partner and negotiating a contractual arrangement, as well as the resources required to select, collect, prepare, clean and validate the data before it is shared ([46]). Investments may also be required in new data infrastructure that enables data sharing. As is common in many cost-benefit analyses, while the costs are tangible and straightforward to calculate, the value of the benefits stemming from the data sharing are more difficult to quantify for both parties to the data-sharing agreement (see Chapter 4). This is all the more so as some companies still lack well-defined key performance indicators (KPIs) to measure the success of their data-sharing initiatives ([47]). Likewise, it proves extremely arduous to measure the risks of not sharing data, be they reputational or commercial. Having said this, these organisational considerations have more to do with the inherent features of private companies and civil-society organisations that consist of inward-looking data infrastructures, made to extrapolate commercial (as opposed to public) value from their data (see Chapter 4). Along these lines, data sharing might require leadership support, data literacy, skills and awareness that together require a cultural change internal to any organisation so as to unleash all transformative effects of data. This appears to be equally true for data users, be they public-sector bodies (that might be unable to use the data they have already), NGOs or academia, insofar as their awareness, data literacy, skills (see Chapter 3) and overall capacity to prompt, enter and maintain a data collaboration are limited ([48]).

1.5.2. Governance

B2G data-sharing collaborations lack a governance framework that prompts, accompanies and governs their development.

([45]) While multiple versions of a data life cycle exist with differences attributable to variation in practices across domains or communities, each data life cycle model provides a high-level overview of the stages involved in successful management and preservation of data for use and reuse.

([46]) Shared data is typically obtained from an organisation's operational or production data. Often times, data needs to be enriched by adding documentation and metadata to improve discoverability by and usability for data users.

([47]) Capgemini Consulting, 'Cracking the conundrum: how successful companies make bid data operational', *Data Science Association*, 8 March 2015, (http://www.datascienceassn.org/content/cracking-data-conundrum-how-successful-companies-make-big-data-operational).

([48]) Hidalgo-Sanchis, P., and Verhulst, S., 'Opinion: The promises — and challenges — of data collaboratives for the SDGs', *Devex*, 7 January 2019, (https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082).

This includes the absence of dedicated structures, a legal framework, best practices and standard scenarios. In particular, one of the major uncertainties remains the applicable liability regime: determining who, if anyone, can be held responsible when inaccurate or biased data is shared and this may contribute to a 'wrong' or discriminatory decision by a public-sector body, possibly causing damage to the general public or organisations. Similar uncertainties exist in relation to the applicability of intellectual property (can the data provider actually share the data it processes?) and competition law (can a B2G data-sharing collaboration be unfair to competitors?) ([49]) regimes relevant to data sharing. The most immediate consequence stemming from the absence of such a data-governance framework is the emergence of a number of economic barriers that prevent B2G data sharing from flourishing. Consider the high entry costs related to entering into a data-sharing partnership. As the experience gained by each and every data collaboration is not shared, this leads to duplications of effort. As some Member States, aware of these bottom-up approaches and their limitations, are putting in place structures and procedures to provide guidance, they risk producing further costs and in particular barriers to sharing cross-border datasets, thus preventing scalability and sustainability. In the absence of a consistent ('coherent') approach, the lack of a governance framework manifests itself via a very diverse landscape of B2G data sharing across the EU, in which data sharing often does not happen because (without advisory structures) it can be a complex, time-consuming and overall uncertain process (see Chapter 2). Ultimately, a common language for the emergence of a data-sharing governance ecosystem remains missing in the EU.

1.5.3. Absence of a data-sharing culture

Due to their spontaneous nature and limited scale, B2G data-sharing collaborations are not yet sufficiently visible, transparent, nor are they scalable and easily repeatable processes. Public authorities, private companies, organisations and the general public are not always fully aware of the benefits of B2G data sharing. Public authorities lack clarity on what data is available and what it entails to create value from it. As a result, private companies, organisations and the general public might be less willing to share data that could otherwise be used to tackle societal challenges (see Chapter 3).

Some ethical considerations also prevent responsible data sharing from occurring, in particular in relation to the modalities of transparent data collection, use and reuse. In particular, by maximising the accessibility and use of data, data sharing inherently raises major concerns, such as those epitomised by the risk of reidentification and the consequences

([49]) See, on this point, the solution provided by the directive on open data and the reuse of public-sector information (OJ L 172, 26.6.2019, pp. 56-83).

flowing from data-protection legislation. B2G data sharing might either foster social development, knowledge and the flourishing of information societies or might help steer the design of current and future societies in the opposite direction ([50]). In other words, while there is something morally desirable about it, B2G data sharing may and should raise serious ethical questions (see Chapter 3, Section 3.1.3).

1.5.4. Operational and technical challenges

Security and privacy of the data are prerequisites for B2G data sharing to happen. From an operational and technical perspective, the availability of trusted technical systems that enable 'safe' B2G data sharing is currently limited ([51]). From the data provider's perspective, sharing data can entail a variety of risks, from the disclosure of sensitive commercial information, trade secrets, customers' personal information or the reidentification of a customer in breach of that customer's right to privacy ([52]). Data anonymisation ([53]), pseudonymisation ([54]) and aggregation ([55]) are a few of the most common techniques to protect personal data, yet their proper application is far from trivial ([56]). Established data-sharing mechanisms are typically sector specific (for example, under the Payment Services Directive 2), notably in terms of technical infrastructure (e.g. data maintenance, analytics, application programming interfaces (APIs)), or only available to data providers and users within a particular (commercial) ecosystem. As such, a model by which users access data (from raw granular data to insights derived from the data) is needed and each method — be it a download or an API — presents its pros and cons. This is particularly relevant insofar as sharing data that is not already open with a controlled audience may raise risks regarding information security. In the absence of this, data providers may be hesitant, if not totally afraid, to share their data. In addition, the limited trust currently existing between a given private company or civil-society organisation and the public-sector body at the time of the storage, access and processing

([50]) Taddeo, M., 'Data philanthropy and the design of the infraethics for information societies', *Philosophical Transactions of the Royal Society A*, Vol. 374, Issue 2083, 28 December 2016, (https://doi.org/10.1098/rsta.2016.0113).

([51]) With the establishment of the International Data Spaces Association, business and research take an active part in designing a trustworthy architecture for the data economy, (https://www.internationaldataspaces.org/our-approach).

([52]) Reidentification by entailing the breach of individual privacy highlights a tension between individual rights and B2G. With the establishment of the International Data Spaces Association, business and research take an active part in designing a trustworthy architecture for the data economy, (https://www.internationaldataspaces.org/our-approach).

([53]) The process of removing personally identifiable information from datasets so that people remain anonymous.

([54]) It replaces personally identifiable information with artificial, fictitious identifiers, to enable the data provider to reidentify the individual in case of need.

([55]) It changes the granularity of the dataset, replacing the data describing a set of individuals with aggregated statistical value that is relevant to the aims pursued by the data processing.

([56]) For a new initiative using differential privacy see 'Harvard's Institute for Quantitative Social Science and Microsoft announce a major collaboration to develop an open data differential privacy platform', *News*, The Institute for Quantitative Social Science, 26 September 2019, (https://www.iq.harvard.edu/news/harvards-iqss-announces-major-collaboration-microsoft).

of data further prevents those collaborations from happening.

Another set of technical challenges emerges from existing limitations in analysing the data for the problem at hand (e.g. potential discrimination in data-driven decisions, difficulties to access data in real time and hence to make decisions in real time, challenges derived from having to combine different datasets, correlation vs causality inference).

These operational and technical challenges also currently contribute to preventing B2G data-sharing collaborations from scaling into a thriving and sustainable ([57]) ecosystem (see Chapter 4).

1.6. TOWARDS A EUROPEAN STRATEGY ON B2G DATA SHARING FOR THE PUBLIC INTEREST

This introductory chapter has shown that getting the best out of the data we incessantly create is a relatively new societal need. In the past, we had to build major infrastructures to enable transportation, energy and postal services to become available. Nowadays, as data becomes an invaluable and inescapable asset not only for the market but also for the state and society, appropriate infrastructures need to be in place for the data and insights to benefit society. Yet, as illustrated in the following chapters, we are currently far from achieving this objective. B2G data sharing requires an overall set of actions towards the creation of an enabling environment for data collaborations to occur at scale. Each of the following chapters expands upon the major obstacles and challenges to B2G data sharing already identified and offers a set of recommendations aimed at overcoming them. The obstacles and their related recommendations are not presented here in order of priority, nor are they mutually exclusive.

Chapter 2, entitled **Governance of B2G data sharing across the EU**, recommends the creation of national structures and functions (e.g. data stewards) necessary for B2G data sharing. The expert group also recommends the promotion of digital skills, improved capacity building and the development of appropriate incentives in order to create a data-sharing culture cutting across the private and public sectors. In addition, the European Commission should explore the creation of a regulatory framework enabling the development of fast, responsible and sustainable B2G data sharing for public-interest purposes. This would provide legal certainty for businesses when sharing

([57]) In this context, 'sustainable' means creating a long-term thriving ecosystem.

data and facilitate the sharing of cross-border and cross-sectoral datasets.

Chapter 3, entitled **Transparency, citizen engagement and ethics on B2G data sharing**, recommends promoting and supporting an informed, inclusive and trusted data-sharing ecosystem. To attain this, the expert group recommends EU action in five areas: to ensure transparency on B2G data-sharing partnerships; to improve public awareness and participation in prioritising societal challenges; to advance data-donation mechanisms; to develop guidance on data ethics; and to foster specialised and bespoke digital skills.

Chapter 4, entitled **Operational models, structures and technical tools to facilitate trusted data sharing**, recommends the development of scalable B2G data sharing through incentives for data providers and public investment in technical means as well as in the emergence of trusted intermediaries. In addition, it calls for further empirical evidence to be obtained on the quantitative impact of B2G data sharing for the public interest.

The last section, in turn, gathers and formulates the key recommendations.

CHAPTER 2: GOVERNANCE OF B2G DATA SHARING ACROSS THE EU

2.1. ISSUES

In the absence of structures to support and oversee B2G data sharing, it remains a complex, time-consuming and overall uncertain process. Furthermore, several economic barriers, such as a lack of incentives to contribute to the public interest, high costs to enter into a B2G data-sharing collaboration and, in certain cases, monopolistic data pricing, hinder the emergence of B2G data sharing. Moreover, access to private-sector data for public-interest purposes is currently governed by a combination of national laws, some EU sectoral legislation as well as contractual arrangements. As a result, there is an increasingly fragmented landscape between and within Member States, as well as between and within sectors. Given these circumstances, the EU is failing to reap the full potential of B2G data sharing.

2.1.1. Lack of structures to support B2G data sharing for the public interest

Due to the rapid 'datification' of our societies, the public sector is becoming increasingly aware of the value of data in helping to address a wide range of societal challenges as well as to save resources for public administrations. At the same time, more and more individuals and actors in the private sector are recognising this potential. For this reason, B2G data sharing for public-interest purposes is developing across the EU and some initiatives to support the supply of private-sector data to public bodies are emerging at national, regional and local levels. However, the number and impact of these efforts remain limited to certain organisations or public-interest purposes. These initiatives are generally heavily dependent on specific individuals committing to this endeavour, which is often not part of their job description and is generally far from their organisations' business priorities and mission.

The main tool used today for B2G data sharing is contractual agreements. Yet, contracts are often phrased in a highly restrictive way vis-à-vis the actions to be performed on the data and may therefore limit its use or render big data analytics unworkable. For example, restrictive

language in clauses can prohibit actions such as merging, enriching, decompiling, structuring, cleansing, etc. ([58]). Companies may be reluctant to enter into data-sharing agreements due to the difficulties in tracking down and controlling the usage of their data. Furthermore, the complexity of setting up various contractual agreements when different sources of data are accessed increases the costs to the parties involved. Finally, there may be cases where the data provider's dominant position could give rise to unfair contractual conditions to access data ([59]).

Currently, B2G data-sharing collaborations are not evolving into sustainable initiatives, but take the form of standalone pilot projects. In particular, since the experience gained is generally not shared across pilots, new entrants into emerging partnerships invest time and money in structuring and tackling operational and legal issues that may already have been addressed by others. This is especially damaging in the case of emergencies, where time is critical and any time lost in structuring collaborations can result in a loss of lives. This situation hinders both public and private organisations' capacity to move away from limited experimentation based on ad hoc and voluntary cooperation to the systematic use of these new data sources. In the absence of a critical mass of experimentation and practice, these scattered projects will not translate into a European culture of data sharing that fosters experts and role models.

Some Member States, aware of these bottom-up approaches and their limitations, are putting in place structures and procedures to provide guidance and operational models, share good practice and foster responsible and sustainable B2G data-sharing collaborations. However, many other Member States are not. Furthermore, even amongst those Member States that are taking action, there is no consistent approach. This creates barriers to sharing cross-border datasets or when aiming to tackle transnational challenges, thus preventing scalability and sustainability.

This also creates a very diverse landscape of B2G data sharing across the EU, in which data sharing often does not happen because (in the absence of advisory structures) it can be a complex, time-consuming and overall uncertain process.

([58]) Out of 129 companies that responded to a survey devoted to data sharing in the EU, 41% of respondents pointed to unfair conditions in contracts as the main obstacles to data sharing, 54% of respondents identified legal uncertainty (what to lawfully do with the data), while 41% pointed to unfair conditions in contracts. Critically, 66% pointed to denial of access as the main obstacle. Everis Benelux, *Study on data sharing between companies in Europe*, 24 April 2018, carried out for the European Commission, (https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en); Bird&Bird, *'Data-related legal, ethical and social issues'*, August 2019, (https://www.twobirds.com/~/media/pdfs/eu-data-economy-legal-ethical--social-issues.pdf?la=en&hash=5E348656C2F87363DD3695F8F06C99E53CEA6B21).

([59]) Everis Benelux, *Study on data sharing between companies in Europe*, 24 April 2018, carried out for the European Commission, (https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en); Bird&Bird, *'Data-related legal, ethical and social issues'*, August 2019, (https://www.twobirds.com/~/media/pdfs/eu-data-economy-legal-ethical—social-issues.pdf?la=en&hash=5E348656C2F87363DD3695F8F06C99E53CEA6B21).

2.1.2. Shortage of dedicated functions in the field

The potential of data collaborations for public-interest purposes is also hindered by the absence of dedicated professional figures in this emerging sector. According to an International Data Corporation (IDC) report ([60]), there were 7.2 million data professionals in 2018, representing 4.3% of the EU workforce. This figure is set to rise to 13 million in 2025. Yet, according to the data professionals skills gap indicator, there could be over 1.5 million unfilled positions in the 27 EU Member States by 2025 ([61]).

In the field of B2G data sharing, this is even more challenging as currently there is an absence of authoritative figures that would make B2G data sharing become a systemic reality. As such, there is no vanguard community of public- and private-sector professionals with the relevant skills to work together, including data scientists, engineers and data-protection officers. In addition, to foster a growing number of data collaborations, a single point of contact is needed within each public, private and civil-society organisation. This point of contact should be able to specify the problem statement to be solved with data, identify the necessary datasets, handle the data from a technical perspective, create trust in both parties and ensure that the data is accessible and of an acceptable quality (e.g. free of errors, bias and inconsistencies).

2.1.3. Economic barriers preventing B2G data sharing

There are a number of economic reasons that currently prevent (or at least do not foster) the development of B2G data-sharing collaborations in the public interest ([62]).

In parallel to the identification of socioeconomic welfare gains, there are some economic barriers that prevent the B2G data ecosystem from emerging and flourishing. These appear particularly acute when data providers, notably private companies and organisations, are not in the business of directly selling their data or making them available to third parties. Yet, when

([60]) *D2.6 Second interim report. The European data market monitoring tool: key facts and figures, policy conclusions, data landscape and quantified stories*, 28 June 2019, IDC Italia srl and The Lisbon Council, (http://datalandscape.eu/sites/default/files/report/D2.6_EDM_Second_Interim_Report_28.06.2019.pdf).

([61]) *D2.6 Second interim report. The European data market monitoring tool: key facts and figures, policy conclusions, data landscape and quantified stories*, 28 June 2019, IDC Italia srl and The Lisbon Council, (http://datalandscape.eu/sites/default/files/report/D2.6_EDM_Second_Interim_Report_28.06.2019.pdf).

([62]) As explained above, 'public interest' refers to increased social welfare outcomes for society.

the social welfare benefits stemming from B2G data sharing exceed the private costs incurred by the data provider (i.e. data-production costs and possible adverse effects to the market), it would be in the interest of society to ensure that the parties go ahead with the data sharing.

The major economic barriers that distinguish feasible from unfeasible B2G data-sharing operations consist of the following:

- a lack of incentives for private firms to contribute to the production of public benefits ([63]);
- high *ex ante* transaction costs and perceived *ex post* risks; and
- monopolistic data suppliers and data pricing.

2.1.3.1. Lack of incentives

A lack of incentives makes private companies and organisations (who may fear a negative impact on them) reluctant to share data with governments. For example, some data providers fear that their data may be used to impose new regulations on their business ([64]). Others may fear that the sharing of such data would make them lose their competitive position in upstream or downstream markets ([65]) or compete with their current data-monetisation products. Finally, the benefits to society at large that are generated through a B2G data-sharing collaboration often cannot be directly monetised by the data supplier. This may hold private companies and organisations back from sharing data.

2.1.3.2. Entry costs

Another important economic barrier to data sharing is the high *ex ante* transaction costs related to finding a suitable data-sharing partner, negotiating a contractual arrangement, reformatting and cleaning the data, etc. Potentially interested public-sector bodies may not be aware of available datasets, or may not be in a position to work with them or understand their advantages and disadvantages. There are also perceived *ex post* risks related, for example, to uncertainties in the quality and/or usefulness of the data, the implementation of the data-sharing deal, ensuring compliance with the agreed conditions, the risk of data

([63]) Martens, B. and N. Duch-Brown (2020), The economics of Business to Government data sharing, JRC Digital Economy Working Paper 2020-04, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf).

([64]) For example, when the JRC sought to procure data from parcel-delivery firms about cross-border deliveries of e-commerce packages in the EU, several firms refused to provide the data because they feared they might be used by the Commission to impose new regulatory measures on the sector.

([65]) For example, mobility-as-a-service (MaaS) urban-transport platforms are a good example of conflicting incentives for private firms to share data with public authorities. MaaS platforms combine information about, and access to, multiple transport-service providers in a single interface. This requires private and public-transport operators to share data with a publicly operated platform. In doing so, these transport operators cooperate and compete at the same time: they compete over rides and market shares in transport while they collaborate to build a common data pool and service that can bring them all more rides. Bruno Carballa Smichowski shows that, although every transport-service provider has incentives to share data, provided that a critical mass of competitors do likewise, the cooperative-competitive dynamics of data sharing can lead to situations where some transport operators will not have an incentive to share their data because they may lose market share and revenue. Carballa Smichowski, B., 'Determinants of coopetition through data sharing in MaaS', *Management & Data Science*, 9 September 2018, (https://management-datascience.org/2018/09/09/determinants-of-coopetition-through-data-sharing-in-maas).

leaks to unauthorised third parties, the lack of experience in data sharing or processing and exposure of personal and confidential data.

2.1.3.3. Monopolistic data suppliers and data pricing

Although there is currently a lack of empirical evidence, there is a high probability that monopolistic data pricing occurs in B2G data sharing. For example, every booking platform has exclusive access to its own room-reservation data and is therefore in a *de facto* monopoly position. Thus, if a city authority requests access to private-room-reservation data from an online booking platform that operates in the city, this city could potentially face monopolistic data pricing. In some cases, the data seller may be in a position to price discriminate between a commercial buyer and a public authority, charging a lower price to the latter . For example, the Joint Research Centre (JRC) procured a unique dataset on online digital transactions from a private market-analysis firm. While this data has a high commercial value, the supplier agreed to sell it at a much lower price because it would not be used for commercial purposes([66]). In other cases, however, a public-sector body may be unable to buy data at a reasonable price, and the collaboration would therefore not go ahead ([67]).

2.1.4. Fragmented landscape for data-sharing rules

Some Member States have adopted horizontal, or even sector-specific, legislation concerning data sharing with the public sector.

Legislation in Member States: some examples

In France, the law for a digital republic allows the public sector to access certain (private sector) data of general interest ([68]). This law considers that data of general interest is:

- data that is held by a private person but linked to a public person (e.g. data on public procurement and concessions, data produced during the performance of a concession contract and data about public subsidies), and
- data necessary for official statistics ([69]).

In Finland, the Finnish forest act obliges forest owners to share information related to the management of the forest (such as forest utilisation, damage, nature management and forest characteristics) with the public sector ([70]).

Furthermore, the British digital economy act grants United Kingdom statistical authorities access rights to new sources of private-sector data and a right to be consulted where these sources make changes to their data or how it is collected and processed ([71]).

([66]) This led to a price negotiation between the two parties, above marginal cost however.

([67]) This has happened to the JRC on several occasions. Martens, B. and N. Duch-Brown (2020), The economics of Business to Government data sharing, JRC Digital Economy Working Paper 2020-04, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf).

([68]) French legislation, 'Loi No 2016-1321 du 7 octobre 2016 pour une République numérique', (https://www.legifrance.gouv.fr/affichLoiPubliee.do?idDocument=JORFDOLE000031589829&type=general&legislature=14).

([69]) This is subject to a decision by the minister for economics and consultation of the national council for statistical information.

([70]) Ministry of agriculture and forestry of Finland: Forest legislation in Finland, (https://mmm.fi/en/forests/legislation).

([71]) United Kingdom government legislation, 'Digital economy act 2017', (http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted).

In the absence of a harmonised framework, Member States, as they become increasingly aware of the potential value of data sharing, are likely to set up additional conditions to facilitate access to private-sector data for certain public-interest purposes, thus increasing fragmentation.

As a result, there is, and will continue to be, growing uncertainty when it comes to identifying not only the rules and procedures governing data sharing across borders and sectors, but also the operational models enabling the sharing of private-sector data (see Chapter 4). This increases the costs for companies engaging in data-sharing collaborations (e.g. legal compliance and technical requirements) which ultimately could tarnish the prospect of a market for data sharing and, more broadly, lead to no (or less) data sharing. In addition, in the absence of a consistent approach to establishing a link between the public interest pursued (as well as the conditions for reuse) and the type of data needed to fulfil that public interest (see Annex I: Data taxonomy), an uneven playing field is emerging. This is one of the reasons why B2G collaborations are not scaling up into sustainable collaborations and partnerships.

In particular, due to the lack of a common, consistent approach to B2G data sharing, a risk of fragmentation exists ([72]):

i. between Member States: this can harm the functioning of the single market, as companies and organisations may have to respect as many different B2G data-sharing legal frameworks as there are Member States. This has already been observed in the statistical sector and led to harmonisation initiatives at the EU level ([73]);

ii. between sectors: due to fragmented practices and rules within sectors, the full value that could be derived from combining datasets from different sectors is not reaped.

This fragmentation is particularly problematic insofar as many of the societal challenges which could be addressed by B2G data sharing require cross-border and cross-sectoral datasets (e.g. climate change, transport or containment of epidemics). Yet the current uncoordinated approach deters such potential for data sharing in the EU.

([72]) The risk of fragmentation also exists between different levels of government within each Member State. However, in line with the subsidiarity principle, this must be addressed at national level.

([73]) Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (OJ L 87, 31.3.2009, pp. 164-173).

2.2. SOLUTIONS

The expert group recommends the promotion and recognition of data stewards as well as the development of digital skills and capacity building in order to ultimately create a data-sharing culture that cuts across the private and public sectors and fosters the principle of reciprocity. In addition, national support structures necessary for B2G data sharing should be created. To overcome the increasingly fragmented landscape for B2G data sharing in the EU, the expert group recommends that the European Commission explore the creation of an EU regulatory framework, including a set of data-sharing requirements and safeguards for the use of data for the public interest. Furthermore, B2G data-sharing collaborations could be organised in sandboxes for pilots to help assess the potential value of data for new use cases, or take the form of public-private partnerships (PPPs).

2.2.1. National structures to support B2G data sharing

Member States should have in place structures to support B2G data sharing. These structures could be a body (or bodies) tasked with assisting public-sector organisations and private companies or civil-society organisations in entering into new data-sharing partnerships and facilitating the sharing of good practice. Over time, such structures could become trusted third parties between the public and private sectors, by bringing the relevant players together. Each Member State would decide whether an existing entity would be appointed to take on these additional responsibilities or a new body created.

In addition, they could oversee responsible B2G data-sharing practices and ensure that when a public-sector body uses data obtained from the private sector, it does so legally and responsibly, without causing harm to the general public or the private-sector partner(s). Member States could also designate within their governance structure an appropriate body (building upon both public- and private-sector expertise) to provide guidance on how to deal with initial dispute resolutions, for example through mediation and complaint procedures. These functions could be mandated to the abovementioned body or to a different structure. In both cases, disputes should be resolved by the established jurisdictional mechanisms in the Member States, respecting the principle of procedural legal autonomy.

2.2.2. Recognising and establishing champions of data sharing: data stewards

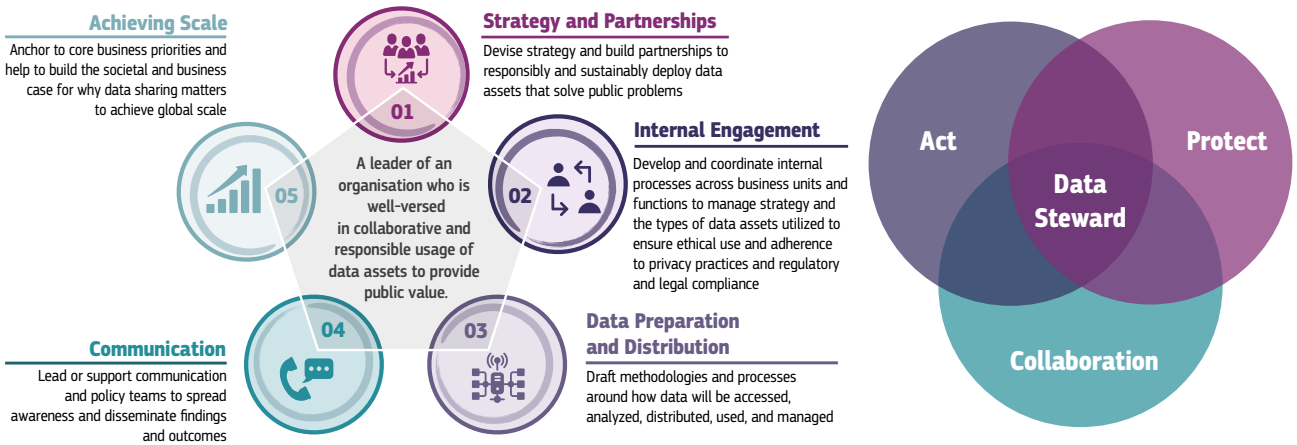
A key success factor in setting up sustainable and responsible B2G partnerships is the existence, within both public- and private-sector organisations, of individuals or teams that are empowered to proactively initiate, facilitate and coordinate B2G data sharing

when necessary. As such, 'data stewards' should become a recognised function ([74]). A data steward should have the required expertise and authority to look for opportunities for productive collaborations or to respond to external requests for data. The primary role of the data steward is to systematise the process of partnering and help scale efforts when there are fledgling signs of success. Some of these tasks might already be carried out by one or more individuals within an organisation, such as a chief data officer, open data officer or chief digital officer. It would be beneficial to group some of these functions together with additional functions in the data steward role.

Data stewards could be integrated into companies and other private-sector organisations, as well as in public-sector bodies and NGOs, to ensure transparency on who needs to be contacted to carry out (and be in charge of) B2G data-sharing collaborations. They would discern, for example, what the purpose of the collaboration is, how to assess proportionality and the appropriate compensation model (e.g. market price, preferential treatment or free of charge). They should seek a win-win collaboration, but also ensure that the public-interest purpose is fulfilled through value/risk assessments. They would ensure that the right processes are in place in their organisations, and that those processes are ethically sound.

As with every new profession, establishing and implementing data stewards is challenging. Today, many who are already fulfilling the functions of a data steward are operating in isolation. Many are simply not sure of the best course of action or common practices on how to establish B2G data-sharing partnerships. There is also no forum where these professionals can learn from each other or exchange good practices. Therefore, the EU should encourage the creation of a network of data stewards, as a community of practice.

The EU institutions should set the example and help other public authorities to follow suit.



([74]) The use of the term 'data steward' should not be confused with the term as used in the context of data management/governance, where it has mostly an internal focus. While the technical data-related tasks of both functions are similar (data quality, standards, metadata, etc.), a data steward in the context of B2G data sharing has additionally an external function for data collaborations and a coordinating role to align all stakeholders around the objectives of the data collaboration. DAMA-DMBOK, Data management body of knowledge, Technics Publications, New Jersey, 2017.

The figure below illustrates the five functions of data stewards ([75]).

2.2.3. Incentives for the private sector

To overcome the economic barriers, appropriate incentives should be in place to encourage private companies and civil-society organisations to share their data. These could include, for example, tax incentives, investment of public funds to support the development of trusted technical tools or recognition schemes for data sharing (see Section 4.2).

Requiring access to certain datasets in specific situations (see Section 2.2.4) does not necessarily imply that the data should be shared free of charge. Companies incur costs to make data available, including time and resources for preparing the data, adapting it to the specific request and building the infrastructure necessary for the transmission of such data (e.g. APIs). For B2G data sharing to become sustainable, these initiatives should be mutually beneficial.

The level of compensation for acquiring data should, in principle, be linked to the effort and investment required by the private company or civil-society organisation for making it available. More specifically, it could be linked to the type of data shared (raw, (pre-) processed, insights: for more information see Annex I: Data taxonomy), the frequency of access (one-off, periodic, continuous) and the potential risk to the data provider (e.g. lost business opportunities, potential competition with the commercial data offerings). Non-monetary benefits (such as tax incentives, improved government services or public-recognition programmes to enhance a company's reputation) might reduce the level of monetary compensation that companies would expect for their data.

The expert group considered four types of compensation for the public sector to acquire privately held data:

- free of charge;
- marginal costs for dissemination;
- marginal costs for dissemination + fair return on investment (ROI);
- market price.

An alternative model of free of charge + tax incentives for the private partner could also be considered. However, insofar as tax regimes differ between Member States, this model cannot be harmonised at EU level, but some Member States may want to adopt such a system.

In acquiring privately held data for public-interest purposes, some preferential conditions may apply for public-interest purposes that have a clear social benefit that justify the resources needed to make the data accessible (see revised principles at the end of this report). For a number of purposes, for example responding to natural disasters or humanitarian crises, data would be shared for free. This approach is already the norm in the CSR programmes of several companies.

([75]) Verhulst, S., 'The three goals and five functions of data stewards', *Medium*, 19 September 2018, (https://medium.com/data-stewards-network/the-three-goals-and-five-functions-of-data-stewards-60242449f378).

Mastercard's data for good initiatives

Mastercard's center for inclusive growth ([76]) advances sustainable and equitable economic growth and financial inclusion around the world. The centre builds on Mastercard's expertise, data analytics, technology and partnerships. Subject to robust privacy, security and data-protection controls, the centre's data for good initiatives solves systemic problems using real evidence, e.g.:

- data fellows programme: research partnerships with leading economists and scholars to create actionable insights, based on aggregated and anonymised data,
- data science for social impact: collaborative initiatives, for example with the Rockefeller Foundation to reduce information inequality and advance social good,
- programmes and pro-bono engagements: leveraging Mastercard's expertise to enable entrepreneurs and workers to grow, thrive and succeed in the new economy.

Mastercard recently launched its data responsibility imperative establishing six core principles ([77]) to guide the collection, use and management of data including the principle of social impact, i.e. companies should use data to identify needs and opportunities to make a positive impact on society.

Data for the production of official statistics is governed by EU and Member State legislation in line with UN principles ([78]). When data is collected from businesses in a particular sector, private companies are not remunerated as a matter of principle ([79]). However, when statistical offices carry out pilot studies for developing new methods to extract or produce statistics out of data sources, certain operational costs could be covered. The development of B2G data sharing is an opportunity for businesses and national statistical offices to reduce the overall administrative burden through access and reuse of new types of data, in a cost-efficient and innovative manner.

([76]) For more information about Mastercard's center for inclusive growth, please visit https://www.mastercardcenter.org/about-us#our-mission.

([77]) 'Mastercard establishes principles for data responsibility', *Press Releases*, Mastercard, 24 October 2019, (https://newsroom.mastercard.com/press-releases/mastercard-establishes-principles-for-data-responsibility).

([78]) Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (OJ L 87, 31.3.2009, pp. 164-173). 'European statistics code of practice', Eurostat, (https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice). 'Fundamental principles of national official statistics', United Nations Statistics Division, (https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx).

([79]) Statistical authorities never reimburse statistical questionnaires that are sent to companies for data. Owing to the specific characteristics of big data, the need for exceptional contributions or specific incentives in return to data sharing for statistical purposes (e.g. preparation of data) would need to be explored on a case-by-case basis.

2.2.3.1. The role of sandboxes and PPPs

All B2G data sharing can ultimately be understood as a form of public-private partnership (PPP). However, from a practical point of view, sandboxes and specific PPPs can be put in place for specific public-interest purposes or to target specific challenges related to B2G data sharing.

The use of sandboxes could be a way to pilot B2G data-sharing collaborations to assess the potential value of certain data for the public interest. These would also help to develop good practices that enable parties to overcome legal or technical challenges related to data sharing. For public-interest tasks that are managed by a combination of public-sector bodies and private companies or civil-society organisations or for public-interest purposes that are completely outsourced, PPPs could ensure a sustainable flow of data from the private company or civil-society organisation to the public-sector body. By creating stable public-private structures, the involvement of all stakeholders within the data-sharing ecosystem, including end-users and the general public, can be ensured. Any PPP or sandbox should be governed by the B2G data-sharing principles and should comply with the obligations flowing from the open data directive and from competition law.

2.2.4. Create legal certainty in B2G data sharing for the public interest

To address some of the challenges outlined above, in particular the growing risk of fragmentation in this nascent sector, the expert group recommends that the Commission explore the creation of an EU regulatory framework to enable and facilitate B2G data sharing for public-interest purposes. This would be subject to an in-depth impact assessment, in which various policy options would be explored on the basis of their economic, social and environmental impacts (quantified costs and benefits whenever possible) ([80]).

This framework would establish a minimum level of harmonisation to create converging B2G data-sharing and data-collaboration practices, thus ensuring a consistent approach between Member States and sectors while maintaining a certain degree of flexibility for Member States. It would allow datasets and insights to be shared more easily across borders and sectors from business sources to public authorities, making it easier to address cross-border challenges.

Having recognised the need for a harmonised approach to data sharing, the EU has already contributed to putting in place the conditions for more data openness. This proposed B2G framework would build upon and complement the existing and emerging instruments of the European data economy. The EU has taken

([80]) *Better regulation: guidelines and toolbox*, European Commission, (https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en).

action to facilitate the sharing of privately held data in business-to-business (B2B) contexts in some sectors ([81]). B2B data-sharing frameworks take data openness yet a step further than any B2G data-sharing collaboration. Given the special non-rivalrous nature of data, the structures and processes put in place by private companies and civil-society organisations to share data with public authorities could help the former in developing their B2B data markets.

Clear EU-wide conditions regarding B2G data sharing for the public interest would facilitate the scaling-up from pilot projects to sustainable partnerships. A regulatory framework would, in particular, create the foundation for responsible, trustworthy and mutually beneficial data-sharing practices, support the development of standards and interoperability, and reinforce the acceptability of ethical data sharing by the general public.

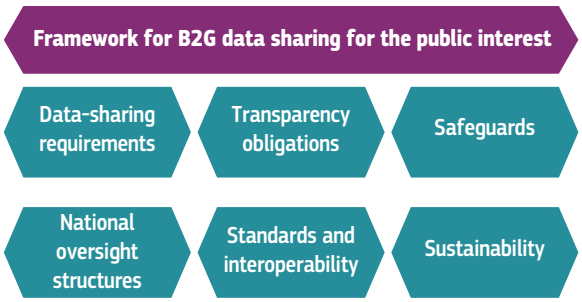
This regulatory framework would be without prejudice to the applicable legal frameworks for personal and non-personal data (e.g. GDPR, ePrivacy directive and the free flow of non-personal data regulation) and intellectual property (IP) rights. This framework could also help discern the responsibilities and liabilities of the parties to a B2G data-sharing collaboration, according to the technical means used for the data sharing (e.g. whether the data is transferred to the public-sector body, the company database is queried remotely or the data is stored in a third-party data platform). It should take into account concerns related to the quality and security of the data as well as cybersecurity of the IT systems. It should also ensure that the competitive position of private companies and civil-society organisations, or their value chains, is not undermined and that B2G data sharing does not distort competition. Ensuring a level playing field can be especially important for small and medium-sized enterprises (SMEs) and start-ups.

The regulatory framework would contain six essential pillars:

1. data-sharing requirements,
2. transparency obligations,
3. safeguards to ensure accountability, protect privacy, and protect the interests of data providers,
4. national oversight of B2G data-sharing structures or functions,

([81]) For example, bank account information via the Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC (OJ L 337, 23.12.2015, pp. 35-127). Metering and electricity consumption information via the (future) directive on common rules for the internal market in electricity (COM(2016)864/F2). Repair and maintenance information via Regulation (EU) 2015/758 of the European Parliament and of the Council of 29 April 2015 concerning type-approval requirements for the deployment of the eCall in-vehicle system based on the 112 service and amending Directive 2007/46/EC (OJ L 123, 19.5.2015, pp. 77-89). Information on medicinal products and clinical trials via the Commission Regulation (EC) No 729/2004 of 15 April 2004 concerning the classification of certain goods in the combined nomenclature. Chemical properties of substances via the Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.

5. common standards aimed at ensuring interoperability across borders and sectors, thereby lowering the costs for data-sharing partners (see Chapter 4),
6. sustainability.



The recommended regulatory framework would apply to all involved in a B2G data-sharing collaboration: private companies (including platforms, civil-society organisations that operate in the EU, collect data from the general public in the EU or have a legal representative in any of the EU Member States) and public-sector bodies. It should also take into account the specific situation and characteristics of SMEs and start-ups.

Obligations flowing from this framework would only apply to data that has already been collected, for example for internal business purposes or for developing current or future products and services (e.g. personalised or targeted advertising). It will not create any new obligation on data providers to gather additional data. Furthermore, public-sector bodies would not seize private-sector data because businesses might continue to monetise the same data in existing or future B2B data markets, given the specific characteristics of data, namely being a valuable non-rivalrous and infrastructural good.

2.2.4.1. Data-sharing requirements

Required sharing of data and insights

The EU faces a number of cross-border challenges in building a thriving data-driven economy. Access to private-sector data may facilitate or even render certain solutions possible. EU-wide public-interest purposes such as environmental protection, cross-border emergencies (e.g. natural disasters or health epidemics), statistics or the delivery of certain public services (e.g. cross-border health care) therefore need trustworthy and stable channels for cooperation. To bring legal certainty for ongoing public needs and the required prompt action in the event of a public emergency, private companies would be required to share the necessary data (ranging from raw data to insights (see Annex I: Data taxonomy). Obligations could also be envisaged for certain data that is scarce (e.g. data deficits), unique (e.g. in the hands of a single company), needed by public authorities to ensure compliance with existing laws (e.g. local short-term rentals) or for cross-border datasets that would help solve cross-border challenges. This framework would also recognise that, in some cases, it would be more cost-efficient and preferable for private companies and organisations to share the insights instead of the raw data (provided that those insights are representative and not biased).

B2G data sharing is not yet a widespread, sustainable practice. There is some evidence that in the absence of obligations for certain public-interest purposes, private companies

may either resist disclosure or, should disclosure occur on a voluntary basis, share data or insights on their own terms ([82]). This may further constrain or limit the action of a public-sector body to solve a certain societal challenge. In other words, the nature and scope of the needed data may require the introduction of detailed legal obligations that would ensure the veracity of results and independence of public-sector action. Finally, required data sharing does not necessarily imply that the data should be shared free of charge.

Flexibility for Member States and specific sectors

A regulatory framework would set the conditions for a minimum level of harmonisation within the EU. In line with the subsidiarity principle, decisions and actions should be taken at the most appropriate level, taking into account the diversity between Member States and between sectors. As such, in addition to the recommended required sharing of data and insights, Member States would have the flexibility to make data sharing mandatory for purposes that are particularly relevant to their national or local priorities. Likewise, specific sectors could go beyond this minimum level of harmonisation provided at EU level, as is already the case for the statistical ([83]) and transport sectors. In the latter, Delegated Regulation No 886/2013 supplementing the intelligent transport systems (ITS) directive provides for an obligation to share road safety-related information with the public sector and other private operators ([84]).

To assess whether data sharing should be required for a given use case, the following criteria could guide the decisions of public authorities, thereby providing a harmonised decision-making procedure applicable in each Member State:

- likelihood of the benefits (capability to act on the insights, technical maturity, economic viability);
- intensity of the likely benefits (number of people affected, money saved);
- likelihood of harms when sharing the data (ability to identify persons/groups, harm to the data provider and reuser, low quality of the data, potential impact of fundamental rights);
- intensity of the likely harms when sharing the data (number of people

([82]) Scassa, T., 'Sharing Data in the Platform Economy: A public-interest argument for access to platform data', *UBC Law Review*, Vol. 50, No 4, 2017, pp.1017-1071.

([83]) Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (OJ L 87, 31.3.2009, pp. 164-173).

([84]) Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to data and procedures for the provision, where possible, of road safety-related minimum universal traffic information free of charge to users (OJ L 247, 18.9.2013, pp. 6-10).

affected, money lost, fundamental rights affected) ([85]);

- immediacy/urgency ([86]) of the situation;
- potential harm of the non-use of data; and
- no other possibility to have reasonable access to the data.

Producing relevant statistics on the collaborative economy

Within the digital economy, the collaborative economy is a fast-growing phenomenon where the services of platforms give customers a greater choice and create new entrepreneurial opportunities for businesses and the general public. This includes new offers of accommodation.

European statistics on tourist accommodation are collected via surveys to hotels with more than 10 bed places. Respondents typically spend 5-60 minutes to fill in the survey, depending on the level of automation. In order to capture the growing impact of the collaborative economy, it would be necessary to also include accommodation with fewer than 10 bed places. However, this would require setting up a survey for about 100 000-150 000 respondents and would imply high costs for the (mainly small) entrepreneurs.

The alternative is to collect such information from intermediate platforms, using their strong IT and data infrastructure. This would shift the reporting burden from micro-enterprises and SMEs to relatively few platforms. In fact, they are already being requested to provide ad hoc information to city councils or regional authorities about the number of visitors and length of stays. Because of a lack of coordination, each request is different and time-consuming and the platforms would benefit from a more coordinated approach and the availability of reliable statistics.

Voluntary sharing of data and insights

Data sharing for certain types of public-interest purposes could take place on a voluntary basis. However, to mitigate the risk of fragmentation stemming from such ad hoc collaborations, these voluntary collaborations should be guided by a set of governing principles intended to guarantee mutually beneficial, sustainable and responsible collaborations. These principles would be inspired by the principles of the EU regulatory framework, which may also govern voluntary B2G data sharing.

([85]) These criteria were used by the expert group to identify examples of required and voluntary use cases for the public-interest examples that are presented and discussed in this report.

([86]) The criterion includes situations which need to be dealt with in the short term to resolve medium- or longer-term objectives (e.g. urban mobility, climate change or chronic diseases).

2.2.4.2. Transparency on data for the public interest

As new use cases for B2G data sharing continue to emerge in parallel to greater societal expectations, the need for transparency, accountability and oversight of this collaborative space grows. The regulatory framework could include a number of transparency obligations on both the supply side of the equation (those that have the data) and on the demand side (those that need the data).

When it comes to the private sector, nobody really knows what data companies have. This suggests an inherently asymmetrical starting point when embarking on a data collaboration. Transparency could be key in shedding light on the various types and characteristics (e.g. quality, size, variability, representativeness or timeliness) of the data that is available, but is not tapped into due to lack of observability and awareness. In this respect, private companies and civil-society organisations should be transparent on the type of data available for B2G data sharing.

Likewise, the public sector should be transparent not only on how it engages with the private sector, why it needs private-sector data, what has been done with the data and the results obtained, but also on why it chose to embark on a given data-sharing collaboration with a specific private company or civil-society organisation and under what conditions. The public sector should provide feedback to the data supplier on how the data has been used. This would also ensure the public sector's accountability towards the private company or civil-society organisation, and society as a whole.

These transparency obligations would also help build public trust and societal acceptability of B2G data sharing (see Chapter 3).

2.2.4.3. Safeguards to protect privacy, accountability and interests of the businesses

Safeguards should be put in place to enable the development of trustworthy data-sharing practices and avoid the potential misuse of data (e.g. government surveillance or monitoring through the accumulation of the individuals' digital footprints). These include accountability (e.g. to minimise negative unintended consequences on results obtained or data used) and data-quality provisions (e.g. privacy by design and security by design) as well as provisions to ensure, where relevant, the informed consent of data subjects when public-sector bodies need to access granular personal data. Finally, provisions for fair, secure and ethical data use should be envisaged too.

The data accessed within the framework of B2G data sharing will not be used for law-enforcement purposes against private companies and organisations or individuals ([87]). Safeguards will be put in place not to overrule other obligations on data minimisation and data retention flowing from EU or Member States law.

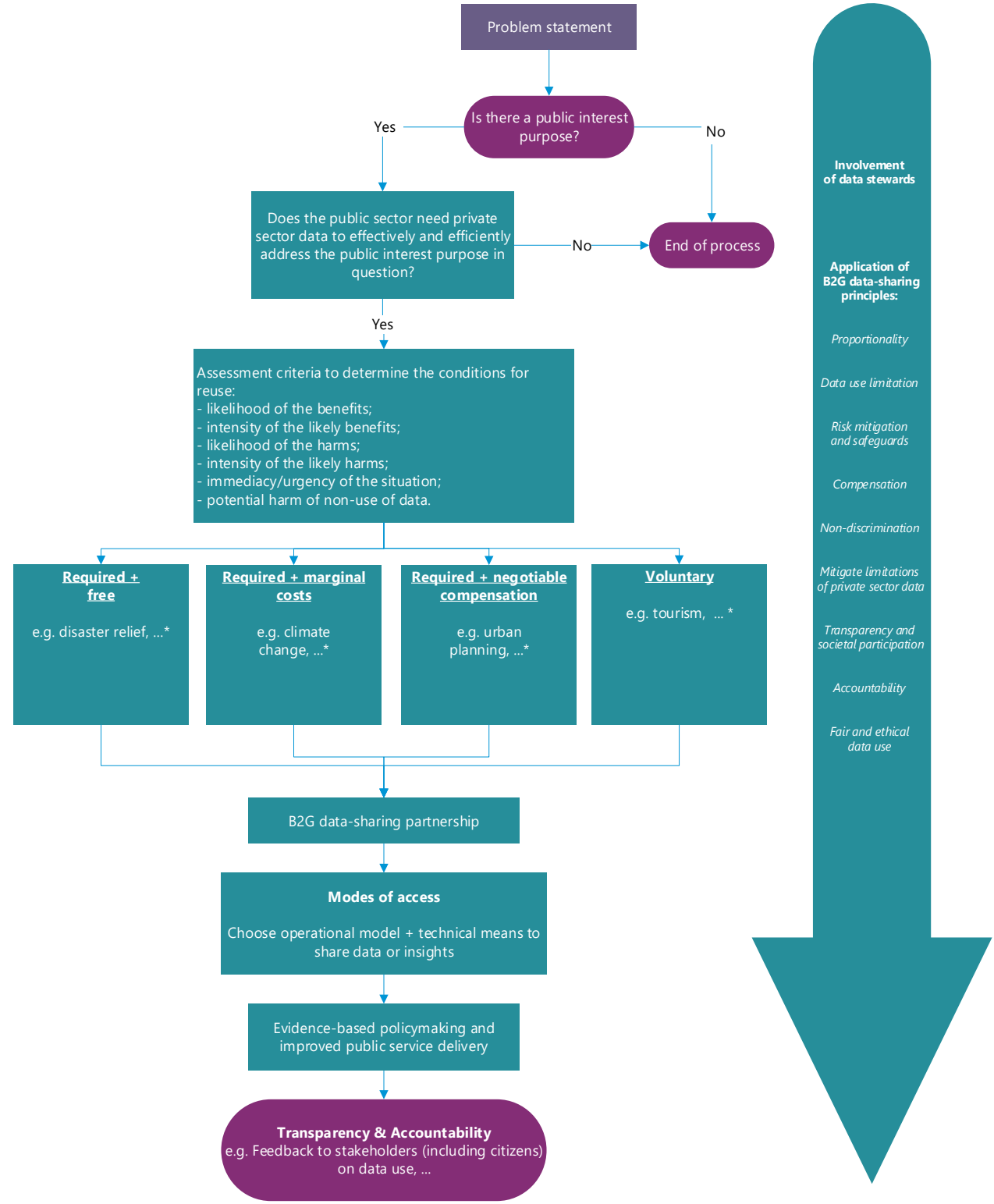
These safeguards would build on the revised principles on B2G data sharing for the public interest (see revised principles at the end of this report). They should also be embedded in the previously discussed governing principles for voluntary data sharing and complement the regulatory framework where necessary.

An EU B2G data-sharing regulatory framework should be future-proof and not crystallise the *status quo* in terms of categories of public interest that would trigger B2G data sharing or negatively impact the development of standards and interoperability. It should also not be stifled by the creation of burdensome regulations, although methodologies and processes (either at the EU or at the national level) should contain appropriate controls and safeguards for the general public and for private / public / civil-society organisations. Finally, it should ensure that European companies are not put at a disadvantageous position vis-à-vis companies from outside the EU.

This policy and regulatory framework would help create a culture of data sharing: a culture which should encourage PPPs through a variety of operational models and collaborations (see Chapter 4). These partnerships should be supported by capacity-building initiatives, research investments, development of digital skills, instruments for public engagement and other initiatives. The next chapters explore how further, and equally important, measures and investments should be put in place so that B2G data sharing becomes a sustainable, responsible and systemic activity.

([87]) Other legal frameworks exist for accessing data for law-enforcement purposes against companies and individuals. For example, the proposals on cross-border access to electronic evidence, (https://ec.europa.eu/info/policies/justice-and-fundamental-rights/criminal-justice/e-evidence-cross-border-access-electronic-evidence_en).

Business-to-Government data sharing for the public interest: step by step



* NB: examples are illustrative; they should not be considered as a binding list.

CHAPTER 3: TRANSPARENCY, CITIZEN ENGAGEMENT AND ETHICS ON B2G DATA SHARING

3.1. ISSUES

Due to their spontaneous nature and limited scale, B2G data-sharing collaborations are not yet sufficiently visible and transparent, nor scalable and easily replicable processes. In addition, public authorities, data providers and the general public are not always aware of the benefits of B2G data sharing, and — when they are — they do not necessarily have the skills to act on the data. As a result, private companies and civil-society organisations as well as the public might be less willing to share data that could otherwise be used to tackle societal challenges. Some ethical considerations also need to be addressed to ensure responsible B2G data sharing.

3.1.1. Lack of transparency and awareness

As B2G data-sharing collaborations for public-interest purposes emerge and evolve, there is a growing need to ensure a certain level of openness on such initiatives (see also Section 2.2.4.2). Currently, due to a variety of factors predominantly linked to their spontaneous nature, such collaborations are not yet fully transparent.

According to the World Economic Forum, transparency in data sharing 'involves giving stakeholders meaningful ways to understand relationships, intent and outcomes. To achieve this, individuals need accessible and understandable information on how relationships are structured on how data is being used. Additionally, transparency requires that organisations have the capacity and oversight to ensure that all outcomes from a data collaboration are accurate and that biases (intentional or unintentional) are not systemic.' ([88])

([88]) World Economic Forum, 'Data collaboration for the common good: enabling trust and innovation through public-private partnerships', *Insight Report*, April 2019, (http://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf).

In B2G data-sharing collaborations, the public sector bears the responsibility to be transparent not only on how it engages with the private sector, but also on why it chooses to engage with a specific data provider. It also needs to identify a clear value proposition once it decides to call for B2G data sharing. Currently, there is no reporting on the type of data that has been used to address specific public-interest purposes and the outcome of a B2G data collaboration. This transparency will allow control and oversight by societal players and make B2G data sharing responsible in the long term. Unless properly informed on what has been done with the data they provide, private companies and civil-society organisations do not always trust that public authorities will fully protect their interests and will be accountable for the handling of their data.

In order to boost the emergence of more B2G data-sharing collaborations, private companies and civil-society organisations need to be transparent towards the public sector about the kind of datasets available for B2G data sharing. This could mainly focus on observed data, not inferred data.

In practice, such transparency is often not achieved. This is because there is no established practice or framework providing guidelines on how, and to what extent, openness regarding B2G data sharing should be established.

This lack of transparency contributes to the wider issue of today's low level of awareness of the potential value of B2G data sharing — amongst private and public entities, but also society at large. In the absence of more widespread published information on the results achieved through B2G data sharing, and on its potential to tackle a wide range of societal challenges, the number of public-sector bodies that engage in new use cases will remain limited. The public is particularly exposed to such a risk and yet their cooperation is key. Without appropriate information and knowledge, they are less likely to share their data in B2G contexts or to support such types of collaborations.

In addition, lack of transparency might lead to misinformation, which can damage the reputation of the parties involved in data sharing. This happened in 2019 when the Spanish national office of statistics (INE) leveraged insights from three mobile operators to better understand the mobility of a large percentage of the Spanish population ([89]). Despite the fact that no personal data was shared with the INE, sensationalistic and inaccurate news coverage raised concerns amongst the public that their real-time position was shared without their consent ([90]).

([89]) See news coverage: (https://elpais.com/economia/2019/10/30/actualidad/1572467240_977794.html), (https://www.lavanguardia.com/vida/20191030/471289662660/guia-evitar-operador-movil-datos-estudio-ine-movistar-orange-vodafone.html) and (https://www.20minutos.es/noticia/4039296/0/ine-rastrear-movil-operadoras-afectadas-como-evitar).

([90]) Oliver, N., 'El INE no sabe dónde estás ahora mismo', *El País*, 19 November 2019, (https://elpais.com/tecnologia/2019/11/18/actualidad/1574070599_681970.html).

3.1.2. Citizen engagement in B2G data-sharing relations is underdeveloped

As data collaborations are already taking place and being experimented with, it is important that the general public is aware of the possibilities linked to B2G data sharing for the public interest. Only when the public is fully aware of the potential benefits will it be willing to contribute to and facilitate its development.

The need for active public involvement manifests itself at all stages in the life cycle of a B2G data-sharing collaboration: from identifying public-interest purposes that can and need to be tackled with the help of private-sector data, to receiving feedback on the impact of the use of such data. This lack of active involvement can also reduce the public's trust in the public sector in certain situations.

In a 2017 public consultation, there was general support for actions that enable the public to have access to and share health data across borders. Some 93% of respondents agreed that 'the public should be able to manage their own health data' while 81% of respondents believed that 'sharing of health data could be beneficial to improve treatment, diagnosis and prevention of diseases across the EU' ([91]). However, many people are unaware not only of the amount and type of data that they produce (e.g. call records, data from digital services, in-vehicle data, health-care records), but also neglect that they have certain rights over their data. Under the GDPR ([92]), individuals are entitled to transfer their data from any private company or civil-society organisation that has legally collected it to public authorities, without asking the consent of the former ([93]). This right is exercised in a limited number of piloting initiatives in some EU cities. However, due to the limited number of mechanisms in place that empower people to actively store, donate or manage their data, this possibility is not used to its full extent.

Mobile territorial lab

The mobile territorial lab is an urban living lab in Trento, Italy, where the general public can access all its mobile personal data, control who has access to it and even participate in a personal-data market experiment ([94]).

([91]) For more information, see the 'Synopsis report of the public consultation on Digital transformation of health and care in the context of the Digital Single Market', 25 April 2018, (https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-digital-transformation-health-and-care-context-digital).

([92]) Article 20 of the General Data Protection Regulation (GDPR) (OJ L 119, 4.5.2016, pp. 1-88).

([93]) Villani, C., *For a meaningful artificial intelligence: towards a French and European strategy*, *Report* from a parliamentary mission from 8 September 2017 to 8 March 2018, (https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.)

([94]) Statiano, J., Lepri, B., Oliveira, N., et al., 'Money walks: a human-centric study on the economics of personal mobile data', *UbiComp '14* Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, Washington, 1-17 September 2014, pp. 583-594.

3.1.3. Absence of a coherent ethical approach to B2G data sharing

The collection, processing and use of data according to a commonly shared ethical framework is critical to ensuring trust in the data ecosystem. It is essential to the acceptability of data sharing and the willingness of data subjects to share data that concerns them. Not all ethical considerations have been incorporated into binding law or into existing business models. For example, existing legislation does not address potential ethical risks related to the use of non-personal data (e.g. linking certain non-personal data back to an individual). Furthermore, technology developments may outpace regulatory frameworks ([95]), which could widen a legal void on data ethics. While there are some national initiatives that aim to ensure the ethical use of data, such actions are not yet replicated across the EU.

Data ethics seal

Denmark offers an example of conceptualising and operationalising data ethics. On the one hand, six principles of data ethics (self-determination, dignity, responsibility, explainability, equality and justice, development) lay the ground for the conceptual assessment of data ethics. On the other hand, the prototype of a data ethics seal shows how a government can design a tool to help companies operationalise data ethics ([96]).

Data ethics is a relatively new field, which is constantly evolving as new use cases for data sharing arise and the decision-making process becomes more and more data-driven. In this context, the high-level expert group on AI presented, in April 2019, the *Ethics guidelines for trustworthy artificial intelligence* that put forward some requirements that AI systems should meet in order to be deemed trustworthy ([97]). However, while the ethics of AI are closely linked to data ethics, the latter is much broader and considers the entire data life cycle. There are several dimensions to consider in this context.

There is an ethical responsibility to share data if this can, for example, save lives or contribute to other public-interest purposes such as helping reduce pollution in cities. Conversely, the decision (or merely the act) not to share such data or not to properly analyse data received and apply the insights obtained could have implications that could negatively affect society as a whole. From that, the need for accountability on the use and non-use of data arises.

In addition, there is a risk of data bias, notably that data could be used to the benefit of certain social groups — or to the detriment of others — if the collection, storage, processing or transfer of data by private and public players is not done in accordance with a commonly shared ethical framework. There is also, at each stage of the data life cycle, a risk that datasets are misused.

([95]) Romanoff, L., 'Privacy and ethics in international public sector', presentation given at the Workshop on Data Ethics as representative of UN Global Pulse, 26 September 2019, Directorate-General for Communications Networks, Content and Technology, Luxembourg.

([96]) The Expert Group on Data Ethics, 'Data for the benefit of the people — Recommendations from the Danish Expert Group on Data Ethics', November 2018, (https://eng.em.dk/media/12190/dataethics-v2.pdf).

([97]) High-Level Expert Group on Artificial Intelligence, *Ethics guidelines for trustworthy AI*, 8 April 2019, (https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai).

European cities recognise citizen data as a public asset

Recognised as a public asset by the Knowledge Society Forum, data generated in public spaces should be used and managed by society, as a whole. The cities of Barcelona, Edinburgh, Eindhoven, Ghent and Zaragoza led a group of 81 cities to become more socially responsible by using data and developed the 'citizen data' principles. The principles will serve as guidelines for cities in their activities and will help shape the future of EU urban policies for smarter, competitive, more liveable and more democratic cities ([98]).

In summary, the ethical implications concerning the collection of data should be thoroughly considered before, during and after a data collaboration. As discussed by Luciano Floridi, ethical standards are not good in themselves, nor are they sufficient to determine morally good outcomes, but they are likely to facilitate morally good actions ([99]).

Therefore, if public and private players, as well as the public, are to fully engage in and maximise the potential of B2G data sharing, it is important that ethical issues are not only addressed but also embedded in daily collaborations, so that trust in B2G data sharing can be built.

3.1.4. Need for digital skills to foster B2G data-sharing collaborations

The lack of digital skills necessary to carry collaborations through, in either the public or private sectors, emerges as another obstacle that often impedes B2G data sharing. To the extent that the private sector is quicker to embrace digital transformation than the public sector, the latter is set to face the greater challenge.

Those willing to engage in a B2G data-sharing collaboration should be well positioned and able to identify, for example, the dataset(s) needed to achieve a specific public-interest purpose. While the public authorities should then be able to analyse the dataset and work on the insights, the data provider should make sure that data is accessible and used for the agreed purpose. As pointed out above, technical tools and methodologies are critical in implementing B2G data sharing. In this regard, the need for the players involved in B2G data sharing to acquire digital skills in order to make use of these tools and methodologies is critical.

Finally, unless there are professionals with the appropriate digital skills, public-sector bodies will not be able to reap the benefits from the use of private-sector data and develop a more evidence-based and efficient public sector.

([98]) *Eurocities principles on citizen data: 10 data principles for the common good*, developed by the Knowledge Society Forum, March 2019, (http://www.eurocities.eu/eurocities/documents/EUROCITIES-principles-on-citizen-data-10-data-principles-for-the-common-good-WSPO-BAXFRH).

([99]) Floridi, L., *The fourth revolution, how the infosphere is reshaping human reality*, Oxford University Press, Oxford, 2014.

3.2. SOLUTIONS

If B2G data sharing is to be used to its full potential, an informed and trusted data-sharing environment needs to be designed, encouraged and facilitated by all those involved. To create such an environment, the expert group recommends action in five areas: to ensure transparency on B2G data-sharing partnerships; to improve public awareness and participation in prioritising societal challenges; to advance data-donation mechanisms; to develop a data ethics framework, and; to foster specialised and bespoke digital skills.

3.2.1. Ensure transparency and awareness on B2G data sharing

If B2G data sharing is to happen sustainably and responsibly, B2G data-sharing collaborations should be transparent about their existence and about their operation. More specifically, public authorities should make information available regarding data-sharing agreements, to the extent that they do not jeopardise the interests of the private company or civil-society organisation. To the extent possible, public bodies should also be transparent on the data that has been used, the algorithms applied as well as the results of B2G data-sharing collaborations in a proportionate way. Finally, where appropriate and feasible, they should act on insights provided by B2G data sharing.

In order to more easily identify potential B2G data collaborations, private sector players would need to ensure more transparency towards public-sector bodies on the type of data they collect and reuse. This transparency could be achieved through the transparency provisions of the EU framework recommended above.

Measuring economic resilience to disasters with financial data

Banco Bilbao Vizcaya y Argentaria (BBVA) data and analytics collaborated with the UN global pulse on a project using financial-transaction data to understand how people behave before and after natural disasters ([100]). BBVA shared data with the UN global pulse for free through the 'question and answer' (Q & A) model ([101]).

In addition, other dissemination models could be envisaged to increase awareness and transparency of B2G data collaborations. First, good practices on B2G data sharing that help improve people's lives and resolve societal challenges should be made publicly available. The Commission has already taken steps to publish good practices on

([100]) Elena Alfaro et al. *Measuring economic resilience to natural disasters with big economic transaction data,* Cornell University, 2016, (https://arxiv.org/abs/1609.09340).

([101]) de Montjoye, Y., Gambs, S., Blondel, V. et al., 'On the privacy-conscientious use of mobile phone data', *Scientific Data*, No 5, 11 December 2018, (https://www.nature.com/articles/sdata2018286.pdf).

data sharing by financing the creation of the Support Centre for Data Sharing ([102]). To reinforce this, an official 'good practice' label could also be envisaged. This could increase awareness and understanding of the potential of private-sector data for the public interest and encourage players to engage in B2G data sharing. Additionally, private sector players could use these labels and publicity for their own marketing and reputation campaigns (e.g. the 'good practice' label could qualify as an inclusion criteria for the Dow Jones Sustainability Index).

3.2.2. Foster and improve citizen participation

EU action should strive to make B2G data sharing better publicly understood, as well as more socially accepted and human-centric, by actively engaging the public as a potential source of data and by involving them in the choice of societal challenges that can be addressed through B2G data collaborations.

First, public authorities should allow the general public to have a say in the choice of societal challenges that should be addressed and, more specifically, the public-interest purposes that might justify data sharing. This could be done at EU, national or local level. Public online consultations, such as the GovLab's *The 100 questions initiative* ([103]), could provide a way to fully involve the public in the decision-making process on B2G data sharing.

Second, mechanisms should be put in place that facilitate the implementation of the portability rights provided by the GDPR ([104]). For example, the European Commission could fund the development of data-donation schemes or personal-data spaces to empower individuals to take control of their data by exercising their rights under the GDPR, namely giving and revoking consent, requesting data access and erasure of data and requesting data portability. This action would be especially useful in the field of health or for the development of sustainable cities ([105]).

Finally, public authorities should establish feedback mechanisms with the public to inform them of what has been achieved with their data.

([102]) Support Centre for Data Sharing will include model contracts and technical guidance on data sharing both in B2B and B2G contexts, (https://eudatasharing.eu/homepage).

([103]) The GovLab, *The 100 questions initiative*, (http://www.the100questions.org).

([104]) For example, the aim of MyData, a global movement that started in Finland, is to develop national (internationally scalable) interoperability model for personal data management. More on MyData Finland, (https://mydata.org/finland).

([105]) Villani, C., 'For a meaningful Artificial Intelligence: towards a French and European strategy', *Report* from a parliamentary mission from 8 September 2017 to 8 March 2018, (https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf).

'Self data territorial' in La Rochelle, Nantes and Lyon

The people of La Rochelle, Nantes and Lyon can share their data with the government for the public interest (including research) while still keeping control of it. Nantes uses such data to develop an energy transition for the city. La Rochelle aims to collect data for improving mobility services and public transportation. Lyon aims to help socially excluded families and to simplify the lives of those who do not speak French.

3.2.3. Develop a data ethics framework

Ethical considerations should guide the actions of all the players involved in data sharing for the public interest. The framework recommended in Chapter 2 must recognise the ethical responsibility of all involved to contribute with their data when lives are at stake (e.g. natural disasters, health epidemics) or to achieve a healthier and more sustainable society. Moreover, data stewards in private companies and civil-society organisations could take into account these ethical considerations in their CSR strategies to further foster B2G data sharing for the public interest and, at the same time, benefit from reputational gains (see Chapter 4).

The European Commission should publish ethical guidelines for the handling of data, from collection to reuse, in full compliance with human rights and against the risks of misuse and non-use of data. In doing that, the Commission might draw inspiration from and complement the guidelines on the ethical use of AI, the FATEN principles or the work carried out by the UN development group in 2017. The latter resulted in a risk-management tool taking into account fundamental rights, by setting principles for obtention, retention, use and quality control for data from the private sector, and by establishing common principles to support the operational use of big data to advance work on the SDGs ([106]).

Such EU ethical guidelines for data are important for both personal and non-personal data. It has been proven that from non-personal data, it is possible to infer attributes as personal as some dimensions of personality, level of education or interests ([107]). This element is essential to understand the implications of the use of algorithms to model, or even influence, human behaviour at the individual level, as was made clear in the 2018 Facebook/Cambridge Analytica scandal ([108]). Certain attributes and characteristics (sexual orientation, religion, etc.) should remain in the private sphere and should not be or inferred or used by AI systems unless the person expressly decides otherwise.

([106]) UN development group, *Data privacy, ethics and protection. Guidance note on big data for achievement of the 2030 agenda*, 2017, (https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf).

([107]) Park, S., et al., 'When simpler data does not imply less information: a study of user profiling scenarios with constrained view of mobile HTTPS data', *ACM Transactions on the Web*, 2017, (http://www.nuriaoliver.com/papers/TWEB_2017.pdf).

([108]) Euroactive news reports about Cambridge Analytica (https://www.euractiv.com/topics/cambridge-analytica/).

The EU has assumed some leadership in this with the recent implementation of the GDPR. For example, automated decision-making is regulated under the GDPR to counter the absence of meaningful human contact (for example in services operated exclusively by chatbots) or transparency rights to counter being secretly measured, analysed, profiled, oriented or subliminally influenced via algorithms.

FATEN: Ethical principles

Fairness: B2G data-sharing frameworks might need to address potential biases in the data, or provide information about such biases to the public-sector bodies. Given that B2G data sharing impacts on several areas such as health, climate change or transport, a constructive exchange of resources and knowledge between the private, public and social sectors should be encouraged and developed to achieve their maximum potential of application and competitiveness.

Accountability and autonomy: there should be clear accountability regarding the consequences of decisions taken and actions implemented because of B2G data sharing. Human autonomy should never be violated through the use of data.

Trust and transparency: generally, trust emerges when three conditions are met: (1) competence regarding the specific task that the trust will be placed onto; (2) reliability, that is, sustained competence over time; and (3) honesty and transparency.

Equality and beneficence: any B2G data-sharing project should aim to have a positive impact on society, with sustainability, diversity and veracity. B2G data sharing should contribute to increasing equality by enabling more stakeholders to benefit from the existence of data.

Non-maleficence: B2G data-sharing initiatives need to minimise their potential maleficence or potential negative consequences ensuring the security, reliability, robustness of the data and the processes that analyse it, apply a principle of prudence and always preserve privacy ([109]).

Future technologies and technical means to operationalise B2G data sharing need to be designed with these ethical guidelines in mind. In addition, these guidelines should ensure that insights from B2G data sharing exclude any kind of bias that could discriminate against any societal group. The European Commission can target funding from the Digital Europe Programme to create digital innovation hubs and structures where stakeholders can figure out how to ethically use data as well as deploy technical tools for ethical data sharing (enabling data sharing in a safe and ethical manner).

([109]) Oliver, N., 'Governance in the era of data-driven decision-making algorithms', in González, A., and Jansen, M., *Women Shaping Global Economic Governance*, CEPR Press, pp. 171-180, July 2019, (http://www.nuriaoliver.com/papers/Oliver_FATEN.pdf).

Considerations about data ethics will take a primary role in the establishment of B2G data-sharing collaborations as a tool to fill in the gap between the current legislation on data protection and the need to comply with human rights when data is used for the public interest.

3.2.4. Digital skills to foster B2G data sharing

EU funding programmes should support the development of the technical and operational capacity of the public sector to use and act on data. The European Commission could support, among other things, educational programmes, lifelong and inter-generational learning initiatives and training courses.

The expert group, in particular, considers that the EU must support training programmes for the reskilling of policymakers and public-sector workers, in order to enable them to acquire the practical skills needed to leverage the potential of data. However, public administrations should stimulate and incentivise the 'going digital' of policymakers and public-sector workers by also praising those that make efforts to increase the level of digitisation of public authorities. Furthermore, the expert group recommends that the EU support the creation of recognised professional figures in the field, such as data stewards ([110]).

Data-Pop Alliance and data literacy

The Data Literacy pillar in Data-Pop Alliance ([111]) has developed a framework and tools to establish core competencies towards becoming data literate. They are defined around the '4 building blocks of data literacy': context and concepts, design and strategy, methods and tools, and ethics and engagement. These building blocks form the basis of professional training workshops.

([110]) The main functions and roles of a data steward have been described in Section 2.2.2.

([111]) Data-Pop Alliance is a global coalition on big data and development created by the Harvard Humanitarian Initiative, MIT Connection Science, and Overseas Development Institute that brings together researchers, experts, practitioners and activists to promote a people-centred big data revolution through collaborative research, capacity building and community engagement. Since February of 2016, Flowminder Foundation has also joined Data-Pop Alliance, (https://datapopalliance.org).

CHAPTER 4: OPERATIONAL MODELS, STRUCTURES AND TECHNICAL TOOLS TO FACILITATE TRUSTED DATA SHARING

4.1. ISSUES

One of the major barriers hindering the scaling-up of B2G data-sharing collaborations is the absence of operational models and dedicated, trusted, technical systems that enable the safe and trusted sharing of data. Such systems are essential to avoid risks in sharing data (e.g. disclosing sensitive commercial or personal information). From a technical perspective, datasets are not yet interoperable, and data quality and representativeness are not maximised.

4.1.1. Absence of operational models and dedicated structures to scale up B2G data sharing

4.1.1.1. Current operational models do not scale up

While B2G data-sharing collaborations are currently predominantly limited to one-off pilots, they are not limited to a 'one-size-fits-all' format. Instead, they take very different forms according to the different contexts and needs pursued. The most common types of operational models for B2G data sharing are explained in the box below (which can relate to different levels of data accessibility, as pointed out in Section 1.3).

Not all of these models are scalable nor do they serve the same purposes. As such, each of them presents different features.

- **Prizes and challenges**: data holders make data available to qualified applicants who compete to develop new apps or discover innovative uses for the data. This type of data sharing typically involves sharing one specific dataset that was collected at some point in the past for the purposes of the challenge (e.g. a 'hackathon').
- **Data pooling**: data holders group together to create 'data pools' with shared data resources.
- **Trusted intermediaries**: to address a lack of trust to share data, a third party can be tasked to process the shared data to achieve a specified objective without any of the participating parties having access to its full set. The intermediary could be an individual acting as a trustee

or organisations offering, for example, a data-sharing platform.

- **Data philanthropy/donorship**: data holders provide their data to public authorities at no cost, generally in the case of emergencies and/or humanitarian crises. It can be considered a form of corporate social responsibility (CSR).
- **Research data partnerships**: they are characterised by the pursuit of mutual interests and are often found in academia and research-intensive sectors such as the pharmaceutical industry and transport. They are typically subject to non-disclosure agreements.
- **Application programming interfaces (APIs)**: they make data accessible to third parties for testing, data analytics and product development.
- **Intelligence products**: refer to companies sharing data, generally aggregated, that are used to build an application (app), dashboard, report or other tool to support a public-interest purpose.
- **Intelligence sharing**: shared (and often aggregated) private-sector data is used to build a tool, dashboard, report, app or other technical device to support a public interest objective ([112]).

Partnerships and trusted-intermediary models could be scalable, but they need initial investments to grow and thrive and, ultimately, become sustainable and trusted. Prizes and hackathons are not scalable by design as operational models, but they could be useful to help identify how particular datasets could serve a public-interest purpose. Data philanthropy and donorships, although not scalable, can be crucial to make B2G data sharing happen quickly in crisis situations. In other cases, data partnerships are needed to set up structures that ensure continuous data sharing to tackle public-interest purposes such as climate change or mobility in cities.

These operational models need to be supported by a proper governance framework and, in many cases, complemented by the right incentives to encourage private companies and civil-society organisations to engage in B2G data sharing.

4.1.1.2. Lack of incentives and organisational structures

In order to share data with public-sector bodies for the public interest, private companies and public-sector bodies would need to put in place dedicated sharing structures. These consist of both logistical infrastructures (e.g. data maintenance, analytics, APIs) and business-suitable processes and skilled professionals. Setting up these technical and organisational

([112]) Huyer, E., and Cecconi, G., and Capgemini Invent, *Analytical Report 12: Business-to-Government Data Sharing*, European Data Portal, March 2019, pp. 15-18, (https://www.europeandataportal.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf); S. Verhulst, A. Young and P. Srinivasan, *An introduction to data collaboratives*, The GovLab, (http://datacollaboratives.org/static/files/data-collaboratives-intro.pdf).

infrastructures requires an initial investment in terms of financial, technical and human resources, as well as a cultural change. To make up for those costs, there are currently no satisfactory well-identified monetary and non-monetary compensation models for data providers (see Section 2.2 for more on compensation models).

In addition, data providers may fear potential difficulties in justifying any data sharing under preferential conditions to their shareholders, as the impacts will benefit society as a whole and may be difficult to quantify ([113]).

Win-win for Eindhoven and Amber Mobility

Eindhoven has signed a letter of intent with electric-car-sharing firm Amber Mobility to gain data-driven insights that will help shape new mobility policies and implement more efficient public services and solutions. The municipality will grant a parking permit discount for all paid parking areas to Amber Mobility in exchange for access to the data the company collects on the use of shared cars within the city.

4.1.1.3. Lack of quantifiable economic and societal gains of B2G data sharing

Some scientific studies assess the economic and societal impact of the use of certain types of data for the public interest (e.g. open data) ([114]). However, scholars find it notably difficult to quantify the benefits obtained from good practices on B2G data sharing, beyond any qualitative assessment. There is no in-depth macroeconomic quantitative assessment yet of what would be the gains and costs for society of B2G data sharing ([115]). There is, however, a growing consensus on the need to prioritise the development of a more solid evidence-based 'business case' on the benefits derived from sharing and the costs related to not sharing, as well as on a clear description and motivation of what is public interest, to move B2G data sharing from a good idea in theory to a good practice.

4.1.2. Technical challenges in B2G data sharing

4.1.2.1. Lack of trusted data sharing

One of the major barriers preventing B2G data sharing for the public interest is the absence of trusted technical systems that enable safe data sharing, and first of all safe access. While a growing number of private companies and civil-society organisations consider data as an

([113]) Huyer, E., and Cecconi, G., and Capgemini Invent, *Analytical Report 12: Business-to-Government Data Sharing*, European Data Portal, March 2019, pp. 15-18, (https://www.europeandataportal.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf).

([114]) While multiple versions of a data life cycle exist with differences attributable to variation in practices across domains or communities, each data life-cycle model provides a high-level overview of the stages involved in successful management and preservation of data for use and reuse.

([115]) Martens, B. and N. Duch-Brown (2020), The economics of Business to Government data sharing, JRC Digital Economy Working Paper 2020-04, (https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf).

asset, they also see sharing it as a source of potential, multiple liabilities. Sharing data can entail risks, for example, if accessing certain data would disclose sensitive commercial information, leak the data provider's trade secrets or even open the door to cyberattacks.

Additionally, sharing data could hurt companies, if it competes with an existing data-monetisation product, or if accessing certain data would disclose customers' personal information or ease the reidentification of a customer and, thus, violate his/her right to privacy. Pseudonymisation might not be sufficient to preserve people's privacy, particularly if different datasets are combined ([116]). Moreover, personal attributes (e.g. sexual or political orientation, socioeconomic status, education level or even mental health) might be inferred from non-personal data via machine-learning algorithms ([117]).

Security and privacy of the data are thus prerequisites for B2G data sharing to happen. Ensuring security and privacy implies the existence of trust between the private company or civil-society organisation and the public-sector body at the time of the storage, access and processing of data.

OPAL

The Open Algorithm project ensures secure, privacy-conscious and audited access control to privately held data, through the development of APIs. It can give access to public entities to allow authorised queries against data that never leaves the premises of the data holder. However, the representativeness of the insights gathered through the data queries will depend on how representative the data itself is. For this reason, more research is necessary to allow from the analysis of different databases consecutively, ensuring representative insights via federated learning.

Today, OPAL has evolved into a socio-technological innovation with pilots in Colombia and Senegal. OPAL aims to unlock the potential of private-sector data for public-good purposes. It uses state-of-the-art privacy-preserving technology and a participatory governance system with an ethical-oversight body and capacity-building activities ([118]).

Today, several types of cybersecurity systems ([119]) exist to prevent cyberattacks. However, as they have different ways to transfer the data securely, set up the back

([116]) de Montjoye, Y., Gambs, S., Blondel, V. et al., 'On the privacy-conscientious use of mobile phone data', *Scientific Data*, No 5, 11 December 2018, (https://www.nature.com/articles/sdata2018286.pdf).

([117]) Oliver, N., 'Governance in the era of data-driven decision-making algorithms', in González, A., and Jansen, M., *Women Shaping Global Economic Governance*, CEPR Press, pp. 171-180, July 2019, (http://www.nuriaoliver.com/papers/Oliver_FATEN.pdf).

([118]) OPAL website: (https://www.opalproject.org).

([119]) For example: *Homomorphic Encryption Applications and Technology* (CORDIS), (https://cordis.europa.eu/project/rcn/194171/factsheet/en); and *Privacy-Enhancing Cryptography in Distributed Ledgers* (CORDIS), (https://cordis.europa.eu/project/rcn/213144/factsheet/en).

office and store the data, this increases the *silo-ing* of data. There are some piloting initiatives, for example in the automotive sector, which allow access to data while keeping levels of cybersecurity high. However, these have not been replicated in other sectors.

Data taskforce for mobility

In February 2017 a data taskforce between Member States, vehicle manufacturers and service providers was set up with an active data-sharing mechanism in place. This model brings the data out of the vehicles and provides a mechanism for a safe and secure way of sharing data, among other things, via a web server from each manufacturer. Data is dispatched further to and exchanged with third parties, including public authorities. Thanks to the model, third parties do not need to retrieve the data from the vehicle system itself. Neutral servers have also been added so that data can be dispatched from multiple manufacturers via intermediaries ([120]).

When it comes to accessing data, different approaches have been identified (see box in Section 4.1.1.1 for more on operational models and Section 1.3 on general accessibility methods). However, these are not necessarily operational yet in real-world conditions. Technically speaking, data could be accessed as follows.

- Via remote access by a trusted intermediary or directly by the interested party (e.g. researchers, data scientists working for the government) to the database only for such authorised (third) parties under the supervision of the data controller. While the remote-access model would enable near real-time data analysis for confirmatory or applicative analysis (including training of machine-learning algorithms), it is difficult to scale to a large number of users or databases.
- Via a Q & A mechanism by which (third) parties run authorised algorithms via APIs on the data controller's database to obtain certain insights. The question-and-answer mechanism can ensure high levels of data protection and it can be used by many (third) parties. However, it might not be suitable when there is a need to combine data from different sources.
- Via limited release of certain data under certain conditions stipulated in a contractual agreement. This model would only apply to exploratory projects that are either in the research phase or in an early development stage. Thus, the scalability of this model would be low. Examples are the prizes and challenges and data philanthropy initiatives mentioned in the box in Section 4.1.1.1.

([120]) The extended vehicle model is standardised through ISO (ISO 20077-1). More details can be found in Annex II: Good practices. Detailed info in 'Access to vehicle data for third-party services', *ACEA Position Paper*, 2 December 2016, (https://www.acea.be/publications/article/position-paper-access-to-vehicle-data-for-third-party-services); and on *Car data facts*, (https://cardatafacts.eu/extended-vehicle-concept).

- Via pre-computed indicators and synthetic data that is highly aggregated and disconnected from the raw data that provides certain insights on the data. This model would have a high scalability, but it would require the data provider to invest in building certain sharing infrastructures such as APIs (examples are the intelligence products and intelligence sharing types mentioned in Section 4.1.1.1).

In addition, other technologies can boost trust during the processing of the data. On the one hand, several promising designs for secure databases have been suggested in recent years, including trusted user inferfaces and distributed or federated learning algorithms operating in different databases. This allows insights to be extracted without the data needing to leave the data provider's infrastructure. On the other hand, cryptographic techniques could support computing on encrypted data, such as fully and partially homomorphic encryption[121] or secure multiparty computation[122]. These techniques permit collaborative engagement in computations with untrusted parties, while obtaining some guarantees as to the integrity of the output.

These new techniques have reached a stage of relative maturity. Unfortunately, their practical ('concrete') implementation still highlights many limitations in the real-world setting, for example regarding the practicality and simplicity of the data sharing and the lack of performance, efficiency and flexibility. Additionally, the majority of these techniques are capable of protecting data only when that data is 'at rest', while lacking the capacity to process and protect streaming data (real time, continuously flowing), such as data produced by IoT devices.

Differential privacy, for example, is particularly promising in the context of data sharing. It could enable statistical analyses of datasets that might have personal data without singling out whether any data of individuals was included in the original dataset ([123]). The most studied use case is a statistical query release, which obtains quantitative answers that have limited random noise (e.g. how many people in the dataset live in Belgium?). Differentially private algorithms can tackle many of these queries approximately so that the same conclusions can be reached without accessing the granular data ([124]). Beyond these simple statistical queries, today there are examples

([121]) For more information about homomorphic encryption, please visit https://en.wikipedia.org/wiki/Homomorphic_encryption.

([122]) For more information about multiparty computation, please visit https://en.wikipedia.org/wiki/Secure_multi-party_computation.

([123]) The guarantee of an algorithm applied to a dataset that is differentially private is that its behaviour does not change when an individual is included or excluded from the dataset. Note that this guarantee would hold for any individual and any dataset. Hence, regardless of the specific details of an individual's data (even if such an individual is an outlier), the guarantee of differential privacy should still hold. Nissim, K., et al. 'Differential privacy: A primer for a non-technical audience', *Privacy Law Scholars Conf.*, 2017, (http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience); Dwork, C., and Roth, A., 'The algorithmic foundations of differential privacy', *Foundations and Trends® in Theoretical Computer Science*, Vol. 9, No 3-4, 2014, pp. 211-407.

([124]) Obviously, the larger the dataset, the higher the guarantees of differential privacy because as the number of individuals in a dataset increases, the impact of any single individual on the aggregate statistical results diminishes.

of differentially private algorithms in game theory, machine learning and statistical estimation, all of which are of relevance to B2G data sharing for the public interest ([125]). However, the deployment of differential privacy in the real-world setting and for other use cases is still challenging and would need further investment.

Finally, technologies such as distributed ledger or blockchain can provide a new way of handling digital data, such as recordkeeping, storing and transferring. This can be done in combination with cryptographic technologies to ensure the confidentiality of the data. Distributed ledger technologies are gaining momentum in financial technology *(Fintech)* with the development of cryptocurrencies. Yet they still need to be deployed in real-world conditions and to be used also in other sectors. Limitations such as energy consumption and inefficiency would also need to be tackled.

4.1.2.2. Lack of interoperability

With the continuously growing amount of data, interoperability is increasingly a key issue in exploiting its full value when combining data from different data sources, reusing it for multiple purposes and across sectors.

There is generally no standardisation of different levels (be it at metadata schema, data-representation formats or licensing conditions of data) enabling broad data integration, data exchange and interoperability with the overall goal of fostering innovation based on data. This refers to all types of data, including both structured and unstructured data, and data from different domains, e.g. agriculture, health or transport. Multilingual data or metadata descriptions can also pose problems, as reusers need to understand what the data is about in order to process it.

A common standard for the referencing of data is clearly required. In the open data world, candidates for a common standard are the Application Profile for Data Portals in Europe (DCAT-AP), the Connecting Europe Facility (CEF) Context Broker open-stack-based specification and open standards APIs. In the private sector, there is currently no candidate for a common standard, and because of that, the reuse of private-sector data remains limited and is a complex endeavour.

Mapping out existing relevant standards for a number of big data areas would be highly beneficial. Moreover, it might be useful to identify European clusters of industries dealing with sufficiently similar activities to develop data standards. In particular, data provenance and licensing (for example the potential of machine-readable licences) need to be addressed and encouraged.

In the area of licensing, there are no standardised contracts, governance framework or data-sharing structures that can guide B2G data sharing. This absence of legal interoperability is an important barrier due to the legal costs of initiating such sharing.

([125]) Ruggles, S., et al., 'Differential privacy and census data: implications for social and economic research', *AEA Papers and Proceedings*, 2019, pp. 403-8, (http://users.hist.umn.edu/~ruggles/Articles/Privacy.pdf).

4.1.2.3. Data quality and potential bias

Data quality must be fit for purpose in a given process. Higher or lower data quality might be required depending on the use case. In some domains data-quality measures exist (e.g. the financial sector, meteorology, physical sciences, medicine), ensuring the right data is used for the intended purpose. However, there is currently no common and horizontal system by which data quality can be measured. Normally, data quality is based on the measures adopted by data providers to ensure the systematic and standardised documentation and availability of data they wish or need to share. Unfortunately, any data-quality intervention requires the investment of the data producer. For this reason, the benefits obtained from the sharing of such data need to outweigh the data providers' investment.

Moreover, using private-sector data to inform policymaking can pose some challenges related to the representativeness of the results. In fact, the data collected by a private company or civil-society organisation is representative only of their customer bases. Therefore, the data may be inherently biased and only represent a partial view of reality. To avoid this, public-sector bodies would need to access data from other sources via aggregation to get a critical and representative mass of data that can effectively contribute to resolving a public-interest purpose.

Other challenges include:

- lack of real-time access which is needed for some scenarios,
- technical difficulties derived from the need to combine data from different data sources,
- interpretability of machine-learning models applied to the data,
- causality versus correlation inference, and
- lack of accurate training datasets that could be used in supervised training models and/or validate data-driven models.

4.2. SOLUTIONS

The expert group recommends the development of incentives as well as public investment to encourage participation in B2G data-sharing pilots or sandboxes. Furthermore, Horizon Europe and Digital Europe Programme funding should be channelled into prioritising standards to enable broader data sharing as well as into high-impact B2G data sharing aiming to create sustainable and mutually beneficial data partnerships. Horizon Europe funding should also target the development of technologies that ensure the secure and trusted transfer and processing of data operations in real-world conditions. Finally, further empirical evidence should be obtained on the quantitative impact of B2G data sharing for the public interest.

4.2.1. Scalable B2G data sharing and incentives for data providers

4.2.1.1. Investment, via Horizon Europe and the Digital Europe Programme, in research on how to use data for the public interest and the development of data-donation schemes

The expert group recommends increasing investments from both Horizon Europe and the Digital Europe Programme in the previously identified operational models for B2G data sharing. This would help B2G data sharing to scale up.

The expert group recognises that for a B2G data-sharing collaboration to be initiated, there needs to be certainty over the type of data needed to resolve a particular societal challenge. For instance, current research has shown the potential of mobile-call-data records for purposes such as understanding migratory patterns, the spread of diseases or population density. The expert group recommends investing in projects that would develop infrastructures that could be made available in subsequent projects or by other parties. These projects would need to focus on those areas where data can have the highest impact and to identify how B2G data sharing can scale up and become sustainable in those impactful areas.

Investments should be increased for the development of data prizes, hackathons or data challenges that would motivate researchers to work on solving pressing societal issues with the help of data. These initiatives would carry inherent learning and dissemination effects, by allowing researchers to discover new uses of data as well as models that could inform future B2G data sharing.

Further investments should also be used to support the setting-up of big data innovation hubs, incubators and sandboxes where both private companies and civil-society organisations, researchers and public-sector bodies can experiment with different types of data in a safe manner and develop sustainable collaborations.

The Project DECODE

DECODE ([126]) provides tools that put individuals in control of whether they keep their personal information private or share it for the public interest. In doing so, it tries to address people's concerns about a loss of control over their personal information on the internet. This experimental project explores how to build a data-centric digital economy where data that is generated and gathered by the public, the internet of things (IoT) and sensor networks are available for broader communal use, with appropriate privacy protections. As a result, innovators, start-ups, NGOs, cooperatives, and local communities can take advantage of that data to build apps and services that respond to their needs and those of the wider community.

Private companies and civil-society organisations as well as the general public could decide to donate their data for a particular societal cause. Data donation and philanthropy is mainly exercised by private companies and civil-society organisations to help address humanitarian crises and natural disasters and by the general public to improve transport and mobility in the cities. The expert group recognises the importance of data-donation models and recommends increasing investment in setting up schemes or infrastructures that allow for private companies and civil-society organisations as well as individuals to donate their data for the public good.

4.2.1.2. Setting up incentives and organisational structures for B2G data sharing

The expert group recommends the creation of mutually beneficial data collaboratives in order to foster their sustainability. B2G data-sharing partnerships can become sustainable only where the business models on which they are based are fair to all parties involved. The group recommends applying preferential treatment to public bodies which access privately held data for those public-interest purposes that have a clear societal benefit that and justifies the effort needed to make the data accessible (see Chapter 2, Section 2.2.3).

Some piloting B2G data-sharing activities could become part of CSR programmes ([127]). CSR could help establish the business processes necessary for B2G data sharing within the private companies. This would allow for more consistent and transparent management of B2G data sharing, also from the point of view of data providers. Under this model, data stewards could design, follow and supervise the process leading to the establishment of a B2G data-sharing collaboration. Moreover, if B2G data sharing becomes an element of CSR, the public sector could also grant related tax reductions (see Chapter 2, Section 2.2.3 for other tax incentives).

The expert group furthermore recommends developing a number of non-monetary incentive schemes (such as tax incentives) at

([126]) For more information on the DECODE project: (https://www.decodeproject.eu/what-decode).

([127]) This might not be enough to make B2G data sharing sustainable. Nonetheless, it is important to promote it.

EU, national and local level whereby private companies and civil-society organisations could stand to gain in other ways. The expert group recommends developing recognition schemes for trusted partners in B2G data sharing. This would increase the positive reputation of private companies and organisations that help public authorities increase general welfare. Such recognition could qualify as a criterion for a company to be included in the Dow Jones Sustainability Index (for example through recognition of the role of a chief data steward in private companies and civil-society organisations).

LUCA: the Telefónica data unit

The mission of LUCA is to offer the experience and capabilities acquired through the Telefónica transformation process to all other private and public organisations. However, LUCA is more than that. The data experts in this unit believe big data and AI can be of great use to society's development. Therefore, they use Telefónica connectivity data along with other external data sources to give the world back the value of data and contribute to the UN sustainable development goals (SDGs) for 2030. LUCA has established agreements with some of the most relevant organisations and NGOs in the world to create projects that contribute to the social well-being of the people most in need and comply with the SDGs ([128]).

The expert group also recognises that the public sector could bear some of the costs of the B2G data sharing by investing in the development of sharing infrastructure. With this infrastructure in place, various players could access data under similar conditions.

Finally, the expert group recommends targeted investments to support the emergence of intermediaries that, over time, would become trusted partners ([129]). A trusted intermediary could be a statistical office, a data broker from the private sector or an NGO, for example. Trusted intermediaries are part of an organisational structure that is sometimes necessary for data providers to enter into B2G data sharing. Intermediaries enable responsible B2G data sharing between two previously unknown parties. The investments should be directed at helping to further develop this type of player and identifying where the opportunities are for resolving societal challenges. After the initial investments, trusted intermediaries are expected to become a scalable and sustainable model in the long run.

([128]) For more information on Telefónica's LUCA: (https://luca-d3.com/data-for-good).

([129]) *Data trusts: lessons from three pilots*, Open Data Institute Report, 2019, (https://docs.google.com/document/d/118RqyUAWP3WIyyC04iLUT3oOobnYJGibEhspr2v87jg/edit#).

Flowminder: a trusted intermediary

Flowminder Foundation is a non-profit organisation based in Stockholm. Its mission is to improve public health and welfare in low- and middle-income countries by providing data as global public goods. The organisation works with governments, intergovernmental organisations and NGOs. It collects, aggregates, integrates and analyses anonymous mobile operator, satellite and household survey data. These operations enable Flowminder to map the distributions and characteristics of vulnerable populations. In turn, this helps save lives by providing analyses and decision support to all relevant parties all around the world ([130]).

4.2.1.3. Supporting evidence gathering in B2G data sharing through impact assessments

The expert group recommends investing in the development of methodologies to quantify the public value that is derived from implementing B2G data sharing. Such assessments should take into account both the economic and social benefits (e.g. return on investment, benefits to businesses from a better-functioning society) at a particular point in time, as well as include the costs that such collaborations would entail. However, this is methodologically challenging insofar as the costs and benefits of B2G data transfers are, by definition, attributed to different parties: the private costs are borne by the data provider while the social benefits are enjoyed by society as a whole or a selected target group. Ideally, impact-assessment studies should help bring clarity not only on the value of B2G data sharing but also on the value of data more broadly.

4.2.2. Investing in technical means for B2G data sharing

4.2.2.1. Developing trusted B2G data sharing

Privacy-preserving technologies are set to be the key enabler of any B2G data sharing for the public interest, as they will mitigate many of the existing concerns of data providers. They are an essential condition to unlock the huge potential that the use of people's data can have to tackle societal challenges (see Chapter 3). The expert group recommends prioritising Horizon Europe investments in improving the practicality, user friendliness and availability of solutions similar to the Q & A model via APIs due to its high potential for scalability. A Q & A model enables an authorised party to make queries to different databases and receive an insight that is a combination of all the individual federated insights (e.g. queries to federated systems). It should also be a priority to invest in federated learning. These technologies could build on the approach to the OPAL project ([131]) privacy risks and ensure that the current limitations

([130])For more information on Flowminder: (https://web.flowminder.org).

([131])See example box in Section 4.1.2.1.

on the representativeness of the insights are overcome.

Furthermore, the expert group recommends targeting Horizon Europe investments to make available in real-world conditions technologies that ensure the secure and trusted transfer and processing of data. This includes improving the practicality, user friendliness and availability of solutions for performing computations between those who are not trusted parties ('untrusted parties'). Equally important is to invest in the maturity and availability of solutions for multilevel secure databases ([132]) and secure multiparty computation ([133]). For example, technologies such as multiparty computation or homomorphic encryption consume a lot of energy and need many computation resources at their current stage of development and they cannot be utilised with a large number of datasets. Investments are needed to develop technologies with a smaller energy footprint. At the same time, they should allow the processing of large datasets from various parties in a secure way. Furthermore, investments should focus on making these solutions operational not only with data at rest but also under data-streaming conditions and with historical datasets.

Further targeted funding is necessary to innovate through pilots and make data technologies work at speed, in an efficient way. For example, there is a need to apply current cryptographic technologies to real-world situations. Their usefulness has been proven in some piloting initiatives, yet the full deployment of these technologies across sectors and Member States needs to be boosted both through Horizon Europe and the Digital Europe Programme investments.

Further support for open source solutions is also necessary. Trustworthy encryption systems are based on sound mathematics analysed by experts, that have stood the test of time. The transparency provided by open source software enables trust in security solutions.

Horizon Europe should support research on the application of differential privacy and an assessment of whether this technology could be used for B2G data sharing. In particular, applied research on data-preparation methods through differential privacy should be prioritised. Testing, deployment and experimentation with B2G data-sharing use cases using differential privacy ([134]) could be funded under the Digital Europe Programme.

([132]) Thuraisingham, B., 'Multilevel secure database management system', *Encyclopedia of Database Systems*, 2019, pp. 1789-1792, (https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_230).

([133]) Lindell, Y., 'Secure multiparty computation — a seasoned technology with strong foundations', *Blogpost Unbound*, 14 July 2019, (https://www.unboundtech.com/secure-multiparty-computation-seasoned-technology-strong-foundations).

([134]) For a new initiative using differential privacy see 'Harvard's Institute for Quantitative Social Science and Microsoft announce a major collaboration to develop an open data differential privacy platform', *News*, The Institute for Quantitative Social Science, 26 September 2019, (https://www.iq.harvard.edu/news/harvards-iqss-announces-major-collaboration-microsoft).

Horizon Europe and Digital Europe Programme investment in the deployment of blockchain ([135]) in real-world situations should be considered so that private companies and civil-society organisations can exercise control over the data and the access to such data. By their decentralised nature, blockchain technologies have the potential to make processes more democratic, transparent and efficient.

As far as cybersecurity is concerned, the expert group recommends investing in state-of-the-art and secure computing infrastructure for both public-sector bodies and private companies and civil-society organisations. Helping them develop secure sharing infrastructures (such as APIs, and capacity building) would allow the use of pre-computed indicators and synthetic data to be shared between untrusted parties.

Investments in these technologies can reduce *ex ante* transaction costs and *ex post* risks that currently hold back the development of B2G data sharing for the public interest.

4.2.2.2. Development of common standards

Standardisation at different levels (such as metadata schemes, data-representation formats and licensing terms) is essential to enable broader data integration, data exchange and interoperability with the overall goal of fostering data-driven innovation. This standardisation would apply to all types of (multilingual) data, including both structured and unstructured data, and data from different domains.

The expert group recommends that the Digital Europe Programme invest in the development of common standards for data, metadata, representation and standardised transfer protocols. Building on existing EU programmes, initiatives and working groups, such as the CEF and ISA^2 ([136]) programmes and the multi-stakeholder platform for IT standardisation, the expert group recommends prioritising those standards that are most generally used over creating new ones. The chosen standards should then be further developed, possibly in cooperation and with the support of a European standardisation body. Agreeing on a (set of) common standard(s) and promoting this among the Member States will substantially improve the interoperability among data catalogues and the data exchange between Member States and private companies and civil-society organisations.

The recommended EU framework would serve to reduce transaction costs, as it would facilitate the sharing of experience and thereby reduce the duplication of effort. It would also help bring legal interoperability that can be coupled with the development of creative

([135]) Blockchain is a neutral technology with the potential to redefine the 'rules of the game' in terms of data validation, sharing and access. Given the theoretical incorruptible nature of the ledger, blockchain can be a good 'storage place' to ensure the consistency of the non-sensitive elements of a data-sharing initiative.

([136]) ISA^2: Interoperability solutions for public administrations, businesses and citizens: (https://ec.europa.eu/isa2/home_en).

commons licences ([137]) for data that is shared voluntarily.

4.2.2.3. Ensuring data quality and the representativeness of results

For data to be auditable and trustworthy for as many parties as possible, a methodology must be found. In particular, this must enable the objective measurement of data quality via some provable and comprehensible metrics. The development of a metadata reference system at EU level describing the quality of data (e.g. with indicators of accuracy, veracity, consistency, representativity, credibility or conformance) could be envisaged. This system could be implemented through a quality label for data that companies could attach to specific datasets. The quality label would provide information on when, where and how the data was gathered, its recommended use cases, potential bias and, where relevant, information regarding the subjects' demographics and consent. This would enable greater transparency and accountability ([138]).

Only adherence to standards would help increase data quality. To implement this sustainably, measures need to be in place all along the data life cycle. For this purpose, data quality improvement would need to be considered as a process rather than a one-off measure.

Data-quality indicators

The European Data Portal has indicated three aspects a public authority should evaluate to assess the quality of any dataset before it is published: content quality, timeliness and consistency ([139]).

Content quality — datasets should be complete and accurate.

Timeliness — data must be up to date and must be published as soon as an update is available.

Consistency — standards must be used, and published data must be consistent in terms of equal quality and continuity over time.

Furthermore, the Digital Europe Programme should support the development of data infrastructures or spaces allowing public-sector bodies to access data from various sources, thus reducing the risk of biases in their B2G data sharing. It can also target investments at companies that build infrastructures to make data available for the public interest. These spaces could be developed in two ways: through a virtual or distributed platform, or

([137]) For more information about creative commons licences, please visit https://en.wikipedia.org/wiki/Creative_Commons_license.

([138]) Gebru, T., et al., *Datasheets for datasets*, Cornell University, 2019, (https://arxiv.org/abs/1803.09010).

([139]) 'Open Data Quality', as defined and assessed by the European Data Portal, focuses on the measures adopted by data providers to ensure the optimal discoverability of their data sources, the currency and completeness of the available metadata and data, the monitoring of the compliance with standards as well as the quality of deployment of the published data. High quality level enables data dissemination and reuse that goes beyond the publication of data in proprietary and/or non-machine-readable formats, through means of open and machine-readable formats, accompanied by stable and verified links for download. For more info please see: (https://www.europeandataportal.eu/en/providing-data/goldbook/preparing-data).

a physical platform. In the former, investments can be targeted at the development of APIs and other sharing technologies that ensure interoperable and timely access to data, while data is safely stored in distributed servers and databases. Alternatively, investments could be targeted at the development of physical platforms that store and grant access to the data to authorised parties. The management of this physical infrastructure may be attributed to a public, private or civil-society organisation. The feasibility and usefulness of either of the two options would depend on each sector or Member State.

The Finnish forest centre: A data broker

The Finnish forest centre is a state-funded organisation covering the whole country. It is tasked with promoting forestry and related livelihoods, advising landowners on how to care for and benefit from their forests and the ecosystems therein, collecting and sharing data related to Finland's forests and enforcing forestry legislation. The Finnish forest centre obtains data from the private sector through legislation, by purchasing it and also for free from the general public or private sector players via crowdsourcing solutions. Its Metsään.fi-eServices ([140]) offer the latest information directly to forest owners on their properties. Specifically, people who own forest property in Finland can conduct business related to their forests from the comfort of their own homes through this portal. It indeed connects owners with related third parties, including providers of forestry services. Moreover, the portal saves service providers the cost and effort of visiting sites to obtain the latest data on which to base plans. It also contains up-to-date contact details of forest owners.

([140])Finnish forest centre: (https://www.metsakeskus.fi/en/metsaanfi-eservices-forest-owners).

KEY RECOMMENDATIONS

The European Commission and Member States should consider data as a critical public infrastructure for the EU's future and, as such, take measures to facilitate the use of privately held data for the public interest.

Private, public and civil-society organisations should cooperate to ensure a more varied and regular use of data for the public interest to avoid any harm that might occur from its non-use.

GOVERNANCE OF B2G DATA SHARING ACROSS THE EU

- All Member States should have governance structures in place tasked with overseeing and giving advice on responsible B2G data-sharing practices.
- To support this framework, private, public and civil-society organisations should create and promote the function of a data steward. Furthermore, the European Commission should encourage the creation of a network of such data stewards, as a community of practice in the field.
- B2G data-sharing collaborations could be organised in sandboxes for pilots to help assess the potential value of data for new use cases, or take the form of PPPs.
- The European Commission should explore the creation of an EU regulatory framework to facilitate public-sector reuse of privately held data for the public interest. This framework should include data-sharing requirements, transparency requirements and safeguards, without imposing new obligations on the private sector to gather additional data.
- In acquiring privately held data for public-interest purposes, preferential conditions (including, in some cases, free-of-charge conditions) may apply for the public sector in line with the updated B2G data sharing principles. Potential data-sharing requirements do not necessarily imply that data will be shared for free.

TRANSPARENCY, CITIZEN ENGAGEMENT AND ETHICS ON B2G DATA SHARING

- Public, private and civil-society organisations should be transparent on the B2G data collaborations in which they engage, including the data used and the impact of the partnership.
- Member States and the European Commission should make B2G data sharing more citizen centric by ensuring public awareness of the social benefits of data (e.g. data-literacy programmes) and by involving the general public in the choice of societal challenges that should be addressed. The general public should be encouraged to share their data for the public-interest purposes of their choice. To enable this, Member States should create and promote user-friendly data-donation mechanisms.
- The European Commission should develop ethical guidelines on the use of data, including for the public interest, and, where relevant, taking into account the ethical AI guidelines.
- Member States should increase the readiness and the operational capacity of the public sector to use and act on the data, for example by investing in the training, education and reskilling of policymakers and public-sector workers.

OPERATIONAL MODELS, STRUCTURES AND TECHNICAL TOOLS TO FACILITATE TRUSTED DATA SHARING

- The European Commission and Member States should create incentives for B2G data sharing and mechanisms that ensure the public recognition of private companies and civil-society organisations that engage in B2G data sharing.
- The European Commission should, in particular through the Digital Europe Programme and Horizon Europe, support:
  - the creation of a light governance structure to prioritise standards for private- and public-sector data to lower the transaction costs of B2G data sharing and ensure interoperability;
  - the development of pilot B2G data-sharing partnerships in sandboxes for specific societal challenges and the improvement of the practicality, user friendliness and availability of technological solutions for B2G data sharing.

- Through Horizon Europe, the European Commission should fund the development and deployment of technologies needed to implement B2G data sharing at scale and in a responsible and sustainable way. Specifically, the European Commission should fund proposals on privacy-preserving technologies, security technologies and access control technologies.
- The European Commission should carry out studies to obtain further empirical evidence of the macroeconomic and societal benefits of B2G data sharing for the public interest.

REVISED PRINCIPLES ON BUSINESS-TO-GOVERNMENT DATA SHARING FOR THE PUBLIC INTEREST

On 25 April 2018, the European Commission published a set of principles on business-to-government (B2G) data sharing in its communication *Towards a common European data space* and accompanying staff working document *Guidance on sharing private-sector data in the European Data Economy*. In these documents, it defined six principles that could support the supply of private-sector data to public-sector bodies under preferential conditions for reuse.

As part of its mandate, the expert group was tasked to evaluate these principles. While acknowledging their pioneering role in capturing the major issues surrounding the operation of B2G data sharing, the expert group formulates here a number of suggestions to make them even more comprehensive and, ultimately, impactful.

For each principle, the first box shows the principle as it is currently published in the abovementioned documents. Each principle is then followed by the key changes suggested by the expert group. These suggestions are incorporated in the coloured boxes that contain the proposed revised principles. Moreover, the expert group recommends including two new principles: one on accountability and one on fair and ethical data use.

PROPORTIONALITY

Current text:

a) **Proportionality in the use of private-sector data:** Requests for supply of private-sector data under preferential conditions for reuse should be justified by clear and demonstrable public interest. The request for private-sector data should be adequate and relevant to the intended public-interest purpose and be proportionate in terms of details, relevance and data protection. The cost and effort required for the supply and reuse of private-sector data should be reasonable compared with the expected public benefits.

Key changes suggested by the expert group are as follows.

- *Include the need for a balancing test, in which the public interest should be balanced against the interests of other stakeholders, such as industry and individuals.*
- *Clarify that proportionality refers to 1) the detail of the requested private-sector data needed with regard to the public interest pursued, as well as to 2) the cost and effort required for the supply and use of private-sector data for the public-interest benefits pursued and the risks of harm if the data is not used.*

Revised text:

a) **Proportionality in the use of private-sector data:** Requests for the supply and use of private-sector data should be justified by clear and demonstrable public interest. The potential benefits of the public interest pursued should be reasonably balanced against the interests of other stakeholders. The requested private-sector data should be necessary, relevant and proportionate in terms of detail (e.g. type of data, granularity, quantity, frequency of access) with regard to the intended public interest pursued. The cost and effort required for the supply and use of private-sector data should be reasonable and proportionate to the public-interest benefits pursued, the interest of other stakeholders and the risks of harm if the data is not used.

PURPOSE LIMITATION

Current text:

b) **Purpose limitation:** The use of private-sector data should be clearly limited for one or several purposes to be specified as clearly as possible in the contractual provisions that establish the business-to-government collaboration. These may include a limitation of duration for the use of this data. The private-sector company should receive specific assurances that the data obtained will not be used for unrelated administrative or judicial procedures; the strict legal and ethical provisions governing statistical confidentiality in the European Statistical System could serve as a model in this regard.

Key changes suggested by the expert group are as follows.

- *Change the title 'purpose limitation', which is more a privacy principle, to 'data-use limitation'.*
- *Include the notion that, in certain cases, there can be decisions that require data sharing.*

- *Include the concept of data-use rights that clarify what can be done with the data by the public-sector body in full respect of existing legislation, including privacy, intellectual property and database laws, and contractual obligations to which the private sector may be bound.*
- *Data-use rights established between the parties must be respected by established exemptions where the public-sector body is subject to access to documents legislation.*
- *The data obtained may be further used only for compatible purposes to the extent necessary and proportionate.*
- *Clarify that the public sector should be able to combine the private-sector data with other data sources.*

Revised text:

b) **Data-use limitation:** The business-to-government collaboration agreement or the decision that requires data sharing should clearly specify the intended public-interest purpose or purposes as well as the data-use rights (e.g. stipulating what can be done with the data, time-limitation period).

Collaboration agreements should respect existing legislation, including privacy, intellectual property and database laws, and contractual obligations to which private and civil-society organisations may be bound. Data-use rights established between the parties must be respected by established exemptions if the public-sector body is subject to access to documents legislation.

The data obtained may be further used only for compatible purposes to the extent necessary and proportionate; the strict legal and ethical provisions governing statistical confidentiality in the European Statistical System could serve as a model in this regard.

The public sector should be able to combine the private-sector data with other data sources.

DO NO HARM

Current text:

c) **"Do no harm":** Business-to-government data collaboration must ensure that legitimate interests, notably the protection of trade secrets and other commercially sensitive information, are respected. Business-to-government data collaboration should allow companies to continue being able to monetise the insights derived from the data in question with respect to other interested parties.

Key changes suggested by the expert group are as follows.

- *Change the title, as 'do no harm' is a term that is frequently used in a humanitarian aid context, where it relates to not harming the beneficiaries of humanitarian aid.*
- *Include the general consideration that the risks of using private-sector data should be taken into account and mitigated, including the risk of not using private-sector data.*
- *Stress that private companies and civil-society organisations should not be held liable for the quality of the data in question or its use by public authorities for public-interest purposes.*
- *Clarify that public-sector bodies may not use private-sector data for commercial purposes or to compete with a company that has a similar offering.*
- *Include a number of safeguards on the use of private-sector data to protect stakeholders' rights regarding privacy, data security and non-discrimination, and especially those of the data subjects.*

Revised text:

c) **Risk mitigation and safeguards:** The risks, including damage due to the request for and use of private-sector data, should be taken into account and mitigated. Business-to-government data collaborations must ensure that legitimate private-sector interests, notably commercially sensitive information such as trade secrets, are respected. They should allow private companies or civil-society organisations to continue to be able to use and monetise the private-sector data in question as well as derived insights to their benefit. Private companies and civil-society organisations should not be held liable for the quality of the data in question or its use by public authorities for public-interest purposes. Public-sector bodies may not use private-sector data for commercial purposes or to compete commercially with a company that has similar offerings.

The risk of not using private-sector data in relation to tackling well-defined societal challenges should also be taken into account.

Business-to-government data-collaboration agreements or decisions should contain appropriate safeguards as regards the use of private-sector data in order to protect the rights (e.g. privacy, data security, non-discrimination) of stakeholders, in particular the individuals whose data is used.

CONDITIONS FOR DATA USE

Current text:

d) **Conditions for data reuse:** business-to-government data-collaboration agreements should seek to be mutually beneficial while acknowledging the public-interest goal by giving the public-sector body preferential treatment over other customers. This should be reflected in particular in the level of compensation agreed, the level of which could be linked to the public-interest purpose pursued. Business-to-government data-collaboration agreements that involve the same public authorities performing the same functions should be treated in a non-discriminatory way. Business-to-government data-collaboration agreements should reduce the need for other types of data collection such as surveys. This should reduce the overall burden on citizens and companies.

Key changes suggested by the expert group are as follows.

- *Clarify that the level of compensation agreed should be determined following the other principles, in particular the assessment of proportionality in principle a) and the risk mitigation and safeguards in principle c).*
- *Create a separate principle e) on non-discrimination, as it does not fit very well under the current 'conditions for data-use principle' and clarify that it addresses both the private sector and the public sector.*
- *Insert the notion that private companies and civil-society organisations may have a legitimate expectation that, where appropriate, the 'once only' principle will apply in order to avoid receiving repetitive requests for the same data.*
- *Name the remaining principle d) 'Compensation', which captures its content well.*

Revised text:

d) **Compensation:** Business-to-government data-collaboration agreements should seek to be mutually beneficial, while acknowledging the public-interest goal by giving the public-sector body preferential treatment. This should be reflected in the level of compensation agreed, the level of which should be determined taking into account the other principles, in particular the assessment of proportionality in principle a) and the risk mitigation and safeguards in principle c).

e) **Non-discrimination:** In business-to-government data-collaboration agreements, the private sector should treat public authorities that perform similar functions or are addressing the same public-interest purpose in a non-discriminatory way in equivalent circumstances.

Private companies and civil-society organisations may have a legitimate expectation that, where appropriate, the 'once only' principle will apply in order to avoid receiving repetitive requests for the same data.

Public entities should treat private companies and civil-society organisations that have similar datasets in a non-discriminatory way, respecting the commercial and competitive level playing field in and across sectors.

MITIGATE LIMITATIONS OF PRIVATE-SECTOR DATA

Current text:

e) **Mitigate limitations of private-sector data:** To address the potential limitations of private-sector data, including potential inherent bias, companies supplying the data should offer reasonable and proportionate support to help assess the quality of the data for the stated purposes, including through the possibility to audit or otherwise verify the data wherever appropriate. Companies should not be required to improve the quality of the data in question. Public bodies, in turn, should ensure that data coming from different sources is processed in such a way to avoid possible "selection bias".

Key changes suggested by the expert group are as follows.

- *Clarify what private-sector data quality means by giving some characteristics, such as type, granularity, accuracy, timeliness or format.*
- *Specify that private companies and civil-society organisations should not be required to improve data quality at no cost. Indeed, if a public-sector body specifically requests for the private-sector data to be improved/prepared, the private company or civil-society organisation is entitled to receive fair compensation for this effort.*
- *State that public-sector bodies, when processing data from different sources, should avoid biases in general. This comprises different types of bias, such as 'selection bias'. In this context, information on the customer base of private companies or civil-society organisations is highly relevant.*

Revised text:

f) **Mitigate limitations of private-sector data:** To address the potential limitations of private-sector data, including potential inherent bias, private companies and civil-society organisations should offer reasonable and proportionate support to help assess its quality for the stated purposes (e.g. type, granularity, accuracy, timeliness, format), including the possibility to verify the data, wherever appropriate. Private companies and civil-society organisations should not be required to improve data quality at no cost.

TRANSPARENCY AND SOCIETAL PARTICIPATION

Current text:

f) **Transparency and societal participation:** business-to-government collaboration should be transparent about the parties to the data-sharing agreement and their objectives. Public bodies' insights and best practices of business-to-government collaboration should be made publicly available as long as they do not compromise the confidentiality of the data.

Key changes suggested by the expert group are as follows.

- *Address public feedback and societal participation in a second paragraph of the principle. This refers, for example, to empowering the general public to identify public-interest purposes that can and need to be tackled with the help of private-sector data. It can also involve initiatives in which private individuals are asked for their consent to transfer their data from a private company or civil-society organisation to a public-sector body, as described in Chapter 3 of this report.*
- *Add a second type of transparency, by encouraging public bodies also to be open, where possible, on the data that has been used, the algorithms applied, as well as the results of business-to-government collaborations, including the relation to subsequent decision-making and the impact on individuals.*
- *Add a transparency obligation of public bodies to the private companies and civil-society organisations, allowing the latter to understand which particular public interest has been advanced with the use of their data and how, and cases where the data has not been used.*

Revised text:

g) **Transparency and societal participation:** Business-to-government data collaborations should be transparent about the parties to the collaboration and their objectives. Where possible, public bodies should also be transparent on the data that has been used and the algorithms applied, as well as on the results of the collaboration, including the relation to subsequent decision-making and the impact on individuals. Moreover, public bodies should ensure *ex post* transparency to the private companies and civil-society organisations on which particular public interest has been advanced with the use of their data and how, and cases where the data has not been used. Good practices should be made publicly available.

Whenever relevant, public bodies should ensure that mechanisms are in place to stimulate public feedback and societal participation, without compromising the confidentiality of the private-sector data.

ACCOUNTABILITY

The expert group, furthermore, proposes to include an additional principle on accountability with regard to responsible use and sharing of data by all players in a B2G data-sharing collaboration, in compliance with these principles. The accountability principle also includes the legitimate expectation that public-sector bodies truly act upon the insights that are generated by the B2G data-sharing collaboration.

h) **Accountability:** All partners in a business-to-government data-sharing collaboration should be accountable for using and sharing data in a responsible and lawful way and be able to demonstrate compliance. To this end, they would need dedicated 'data stewards' that have a mandate to handle accountable data-use and sharing activities and ensure compliance.

Public-sector bodies should, to the extent possible, act upon the insights provided by the business-to-government collaboration.

FAIR AND ETHICAL DATA USE

Finally, the expert group also suggests including an additional principle on fair and ethical data use. This principle stresses the need to ensure an ethical, legitimate, fair and inclusive way of using data.

i) **Fair and ethical data use:** Data should be shared and used in an ethical, legitimate, fair and inclusive manner, with full respect for the choices made by individuals on how their data can be used. Public bodies should ensure that data coming from different sources is processed in such a way as to avoid possible biases, including "selection bias". The use of the data and the objective of the collaboration should be in line with the public task of the public-sector bodies.

EXPERTS' PLEDGES

URBAN LOGISTICS DATA SHARING

The Association Française de Supply Chain, Club Demeter and Institut du Commerce (ASLOG) and the European technology platform to develop a comprehensive strategy for research, innovation and market deployment of logistics and supply chain management innovation in Europe (ALICE) will cooperate to collect data about urban logistics in the Engagement Volontaire pour une Logistique Efficiente (Evolue) initiative to optimise flows and generate a mutualisation of resources ([141]).

This initiative is led by the three major supply-chain associations in France and counts on the participation of shipping companies, carriers and cities. The data should provide information about the type of goods transported, journeys (time, origin, destination and route), volumes, weight, number of parcels, type of vehicles and the intensity of the emissions of the latter.

This initiative will also contribute to the objective of having zero-emissions urban logistics by 2050. It is expected that pilots with cities will start in 2020.

The sharing of logistics data will benefit cities, the general public and retailers because deliveries will be better planned and, thus, the number of vehicles entering specific parts of the cities reduced. This will help reduce traffic congestion and increase the number of parcel deliveries at each vehicle stop. Furthermore, this will allow the environmental footprint of urban deliveries to be measured and reduced.

([141]) Expert group member Laurent Cytermann has brought this pledge to the attention of the expert group. However, he is not involved in the initiative nor is he a representative of the sector.

BIG DATA, ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS (BIDA) AND THE UN CLIMATE CHANGE CONFERENCE (COP25): REQUEST FEEDBACK ABOUT B2G DATA SHARING

Richard Benjamins and Juan Murillo Arias

The Spanish observatory for big data, artificial intelligence and data analytics (BIDA) ([142]) studies the possibility of a business-to-government (B2G) data sharing initiative to provide future insights on climate change to policymakers.

BIDA consists of around 20 large enterprises and public bodies and is a forum for sharing AI and big data experiences between peers ([143]). It currently has two working groups, one on B2G data sharing and another on AI ethics. The B2G data-sharing working group is looking into the possibilities to combine public and privately held data (duly anonymised and aggregated) of its members into a common data lake to provide access to recognised climate change experts and data scientists. BIDA believes that this would be the first time business data would be shared for the common good on such a large scale. Applying AI and machine learning to this unique data set has the potential to uncover so-far-unknown insights about the relation between economic activities and potential measures to reduce climate change. The value of B2G data sharing to help solve large societal challenges is well recognised by relevant players such as the European Commission ([144]) and TheGovLab ([145]). However, this practice is currently mostly happening in pilot mode. One of the key success factors for B2G data-sharing initiatives is that, from the beginning, potential final users are involved and commit to putting the system in operation if the results of a first pilot are successful.

BIDA has taken the opportunity of the Climate Change Summit in Madrid (COP25 ([146])) to start conversations with policymakers and climate change experts to evaluate the opportunity of this unique initiative.

([142]) Observatorio Español sobre big data, inteligencia artificial y data analytics (BIDA): (https://aeca.es/observatorio_bida).

([143]) BIDA members: (https://aeca.es/observatorio_bida/composicion-bida/miembros_permanentes).

([144]) Commission appoints Expert Group on Business-to-Government Data Sharing: (https://ec.europa.eu/digital-single-market/en/news/commission-appoints-expert-group-business-government-data-sharing).

([145]) The GovLab: (http://www.thegovlab.org).

([146]) UN Climate Change Conference, December 2019: (https://unfccc.int/cop25).

REGIONAL EUROPEAN DATA SPACE

Nuria Oliver

Data-Pop Alliance, a global not-for-profit collaboration of Harvard and Massachusetts Institute of Technology (MIT) institutes, will seek to co-develop a 'regional European data space' with top global universities in Europe and the United States, private companies, civil-society organisations, governmental and intergovernmental institutions, and other interested parties, replicating the regional data space approach that has been successfully developed in Latin America since 2013. This will entail investing and raising funds to develop and deploy collaborative research, data literacy and strategic support activities to leverage private-sector data and AI for social good in a responsible, sustainable and systemic manner. Formal efforts will start in early 2020 with an initial 3-year strategic plan, with a central node in Barcelona and connections to other European cities (Alicante, Paris, Brussels, Berlin, Trento, etc.) and links to nearby regions (Middle East and North Africa, Sub-Saharan Africa) to help Europe lead the fourth industrial revolution.

Key research topics will include social cohesion (inequality, trust), criminality, migration, climate change resilience and sustainable data sharing, building on our research portfolio to date, to shed light on complex social processes and objectives that are core to Europe's values.

Capacity-building and data-literacy efforts will focus on empowering public-sector workers on the one hand and vulnerable and/or under-represented groups on the other hand (migrants, persons with disabilities, children, women and girls) with appropriate awareness and skills to serve as agents of positive social change.

Strategic support will aim to provide technical assistance and guidance to organisations willing to develop and deploy responsible and sustainable data and AI strategies, either internally or externally, to reap the benefits of the fourth industrial revolution's potential while avoiding its major pitfalls.

To that end, the Data-Pop Alliance will build on its extensive network and expertise in the field, investing EUR 100 000 in initial funding while seeking to raise EUR 2-3 million over 3 years.

THE B2G DATA STEWARDS NETWORK

Stefaan Verhulst

In the next 120 days, a steering committee comprised of several members of the B2G data sharing expert group and supported by the GovLab will seek to scope (and subsequently implement) the contours, needs, feasibility and design features of a data stewards network.

Among the members there are the following.

Stefaan Verhulst (the GovLab), who will act as coordinator of the steering committee, Nuria Oliver (Vodafone Institute), Richard Benjamins (Telefónica), Ioana Stoenescu (Roche), Helena Koning (Mastercard), Virpi Stenman (Finnish forest centre), Linda Austere (Finance Latvia Association), Juan Murillo Arias (BBVA), Dimitris Zissis (MarineTraffic) and Milan Petkovic (Phillips).

The GovLab will seek to hold one or more B2G data stewards camps or workshops to further the design features and prototype possible organisational and governance structures. Furthermore, the GovLab will seek additional funding and partners to subsequently implement the data stewards network according to the scoping exercise.

Among the areas that will be considered further are the following.

- Need and incentives: what are or could be the incentives of data stewards/corporations to join the network?
- Metrics: what would success look like? Organisationally? Societal impact? Market shaping impact? Raising the bar?
- Governance: What should the governance structure look like for the network? What can we learn from other models?
- Organisation: How to structure the network? Where to locate it (in existing associations or a new organisation?) What innovative models of organisations would work here?
- Sustainability and funding model: How to ensure long-term sustainability? What should be the funding model?

ENABLING A PERSONALISED HEALTH-CARE FUTURE BY ADOPTING COMPREHENSIVE GENOMIC PROFILING (CGP) AS A STANDARD IN CLINICAL PRACTICE

Ioana Stoenescu

Roche Pharma Romania, Foundation Medicine, Roche Group and the Timis county hospital of Romania are planning to broaden patient access to integrated personalised health care for lung cancer patients. These partners aim to transform the current 'one-size-fits-all' Romanian health-care system into one that is personalised, health-care driven and patient centred.

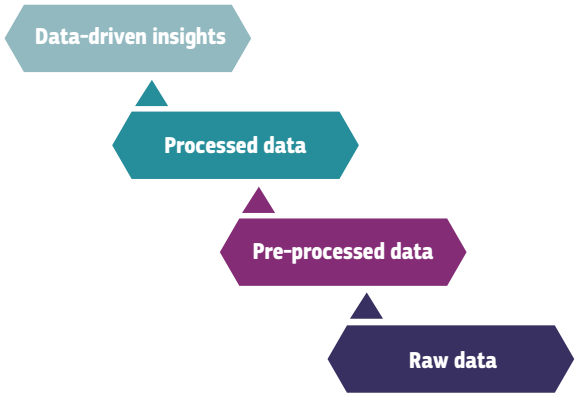
Lung cancer represents a major health burden in most developed countries. The majority of patients have either locally advanced or metastatic disease at diagnostic, and only 20% are detected at an early stage of the disease where surgery is still possible. As a result, the 5-year survival is below 15%. CGP could most effectively determine all the mutations present in a patient's tumour with a single test and determine the two potential indicators of the patient's likelihood to respond to immunotherapies.

Foundation Medicine will analyse the genomic data collected to identify the genomic alterations. This data will, then, be shared with oncologists that will analyse these insights together with the electronic health records of the patient (processed genomic data, time to diagnostic vs current testing alternatives, additional actionable mutations identified vs current testing alternatives). All this data will be collected over a 12-month period and anonymised, and oncologists will then be able to extract relevant insights for the clinical practice. The findings will be published by the investigators, and can be used for future clinical-decision support.

Lung cancer patients will have a more accurate and detailed diagnostic and personalised therapeutic treatment. This would ultimately prolong their survival rate and reduce unnecessary treatment-related toxicities. In turn, oncologists will have the information they need to take data-driven decisions on the personalised treatment. The health authority in Romania will reduce bureaucracy and allocate more resources to those patients for which immunotherapies work. Finally, Roche Group and Foundation Medicine can improve their reputation and raise awareness about their technologies.

ANNEX I: DATA TAXONOMY

In principle, all types of digital content can be captured under the term 'data'. This includes personal and non-personal data. For the purpose of this report, the following high-level data taxonomy maps out and systematises the major broad data categories that exist and are of relevance to B2G data-sharing collaborations. The data taxonomy, as shown in the figure below, consists of four levels of abstraction:



1. **Raw data**: also known as primary data, is data collected from a source (e.g. numbers, instrument readings, images, text, videos, sensor data). Raw data has not been subject to any operation or any analysis. In particular, raw data has not been subject to any other manipulation by a software program or a human researcher, analyst or technician.

2. **Pre-processed data**: data pre-processing is an important step in any data-centric application or service. Data preparation and validation can take a considerable amount of time. It includes, for instance, cleaning, instance selection, re-sampling, normalisation, transformation, feature extraction and selection. It may also include the creation of so-called metadata that provides information about the collected data, e.g. descriptive, structural, administrative, reference or statistical metadata.

3. **Processed data**: data processing can be described as the manipulation of data in order to produce meaningful information. It may involve various practices including, for instance, aggregation/de-aggregation (through which data is combined along a particular dimension, e.g. temporally, geographically; identification/de-identification) that entails combining/ removing personal identifiable information in the data ([147]). Data can be also processed by analysing it,

([147]) Any personal-data collection, processing or movement should be in full compliance with the GDPR (OJ L 119, 4.5.2016, pp. 1-88) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Anonymisation and pseudonymisation are techniques that can be used to protect the privacy rights of individual data subjects and allow organisations to balance this right to privacy against their legitimate goals.

by processing the data to carry out certain tasks, such as classifying it into different categories, clustering it in meaningful groups, making predictions, building models of the underlying data generation process ([148]).

4. **Data-driven insights**: they are generated by drawing conclusions from processed, analysed data. They are produced to understand what is going on with a particular situation or phenomenon. They can then be used to make better decisions and drive change. Data-driven insights consist of the new knowledge created as a result of data analysis processes as described in the previous broad category of processed / analysed data ([149]).

The B2G data-sharing process can take place at any level of this data taxonomy.

([148]) There are other processing practices.
- Sorting: arranging items in some sequence and/or in different sets.
- Summarisation: reducing detail data to its main points.
- Feature extraction: computing features from the data, which are typically the input to the data-driven machine learning/analysis algorithms.
- Combination/fusion: combining pre-processed data from different data sources to tackle a specific problem.

([149]) Dykes, B., 'Actionable insights: the missing link between data and business value', *Forbes*, 26 April 2016, (https://www.forbes.com/sites/brentdykes/2016/04/26/actionable-insights-the-missing-link-between-data-and-business-value/#573a240551e5).

ANNEX II: GOOD PRACTICES

FINNISH FOREST-DATA ECOSYSTEM

The forestry sector generates yearly around EUR 8.4 billion of Finnish GDP and it dominates the bio-economy sector in Finland.

The Finnish forest centre (FFC) is a public body that operates under the steering of the ministry of agriculture and forestry, Finland. It gathers data either by purchasing it from the private sector (e.g. airborne-laser-scanned data) or by receiving it for performing statutory responsibilities, i.e. data-sharing activities enforced by legislation. Finally, the FFC can also receive data for free from the general public or private-sector players via crowdsourcing solutions.

The majority of forest-data-related services is available on Metsään.fi, a portal through which people who own forest property in Finland can conduct business related to their forests from the comfort of their own homes. The portal connects owners with relevant third parties, including providers of forestry services. This makes it easy to commission management work and to be in touch with forestry professionals. The data (personal and non-personal, such as laser-scanned (point-cloud) data, canopy-height model (CHM), aerial photography, micro-compartment data) is normally anonymised. If it is not, there is a specific registration procedure, based on legislation, to complete to obtain the information.

The FFC is currently building a data platform to support private-sector application and business-solution-development purposes. The first use case is related to more cost-efficient timber material management and timber procurement. Another use case is related to the dynamic forest-management planning, which will produce better services for private forest owners. Private-sector players should develop these use cases.

In the near future, a cloud-based forest-data platform will be established to support data sharing between the FFC and other data sources, especially for the open forest-data services. The forest-data platform will combine data from multiple public- and open-data sources such as environmental and weather data. The new forest-data platform will enable data fusions and supports GeoJSON queries.

The Finnish open forest-data system was launched in March 2018. Since then, the available amount of open forest data has increased from 0.36 terabytes (TB) ($\times 10^{12}$) to 0.38 TB, and the amount of downloaded data has exceeded 10.5 TB. The ultimate goal of the open forest-data system is to provide a basis for developing innovative tools and solutions for forestry and other forest-related activities.

BBVA AND UN GLOBAL PULSE COLLABORATE TO MEASURE ECONOMIC RESILIENCE TO DISASTERS WITH FINANCIAL DATA

As temperatures climb and oceans warm, regions will face drastic changes and will be increasingly affected by natural disasters. In the face of climate change, how could measuring economic activity in real time be used to design feedback loops into disaster preparation, recovery and reconstruction programmes and policies?

To find out, BBVA data and analytics has collaborated with the UN global pulse and is throwing its weight behind a project using financial-transaction data to understand how people behave before and after natural disasters. The BBVA shared data with the UN global pulse through the Q & A model.

The project looked at the economic impact of Hurricane Odile on the Mexican state of Baja California Sur (BCS). Findings showed that, at the household level, people spent 50% more than usual on items such as food and gasoline in preparation for the hurricane.

The project analysed point of sale (POS) payment and automatic teller machine (ATM) cash-withdrawal data (anonymised and aggregated) from more than 100 000 BBVA clients. Data analytics were employed to derive proxy indicators of the economic impact and market resilience of people in the region. The UN global pulse defined the main questions, which were refined through an iterative process with the team at the BBVA that finally extracted the relevant indicators. It is relevant that this was the first time that this was done, so a sort of API for extracting the relevant aggregated indicators (if standard) could be an interesting solution upon an activation agreement.

In the aftermath of Hurricane Odile, economic activity decreased across the region. It took 2 weeks for POS and 1 week for ATM transactions to bounce back to normal. Measuring the level of transactions exactly 30 days after impact, the research showed that 30% fewer POS transactions and 12% fewer cash withdrawals were registered, compared to a normal period.

'This type of real-time quantitative data on how people prepare for disaster could be used to inform proactive, targeted distribution of supplies or cash transfers to the most vulnerable, at risk populations,' said Miguel Luengo-Oroz, chief data scientist at UN global pulse.

'With this project, we have created a replicable and evidence-based approach to understanding vulnerability. New insights can help authorities improve community resilience, which benefits the vulnerable and is also good for business continuity,' added BBVA expert Elena Alfaro.

This new partnership lays the groundwork for developing tools and approaches needed to scale up B2G data sharing and to explore their potential to inform humanitarian aid and relief efforts. Furthermore, it shows how a private company can improve its reputation and boost its CSR by contributing to the achievement of a public-interest purpose.

‘KNOW YOUR CUSTOMER’ (KYC) UTILITY

In the context of growing anti-money-laundering (AML) and combating-the-financing-of-terrorism (CFT) awareness worldwide, implementation of ‘know your customer’ (KYC) utility requirements raises economic and legal concerns. Indeed, it is demanding, considering regulatory risk, operational cost and customer (user) experience.

The shared KYC utility is a customer-due-diligence tool operating as a (non-personal, personal, raw, processed, insights) data repository where financial institutions as well as other obligated entities, public institutions and companies feed information and the outcomes of customer due diligence available for the identification and prevention of potential ML/FT risks and other financial crimes.

In the context of Latvia’s *Moneyval* evaluation (2018), the cabinet of ministers of Latvia adopted an action plan to address some problems such as including weak compliance culture and lax practices, scarce resources and tools available to players in the real economy, mainly SMEs. Finance Latvia was tasked, in cooperation with the ministry of finance and the ministry of environmental protection and regional development, with proposing a conceptual framework to enable data-sharing partnerships. They proposed to create a legal framework to enable businesses to launch shared KYC utilities, which operate like a credit bureau (providing part of the information necessary in a loan-issuing process). As the purpose of data processing stems only partially from legislation (KYC duty is not a legal requirement to all, rather a practice of prudent risk management), data protection is the prime limitation. It is possible not to process personal data in a KYC utility. However, as the utility focuses on corporates and individuals associated to corporates or individuals (politically exposed persons (PEPs)) already known from public registers, the risk is partially mitigated. It is also proposed to manage proportionality concerns via differing depth of the right to access and granularity of outputs (depending on the legal entitlement, data and insights, or risk indicators). Finally, to ensure GDPR compliance, a licensing regime for shared KYC utilities should be considered as a condition for reuse.

The policy framework developed will result in amendments to legislation enabling the information sharing and development of technical solutions (as a service by interested third-party providers, e.g. fintechs, special-purpose vehicles -SPVs) as well as analysis of legality, necessity and proportionality of the inherent limitations to a person’s rights (of being included in a shared KYC utility).

The solution is currently only developed for use in Latvia, due to the legal limitations of cross-border data sharing.

LABOUR-MARKET TRANSPARENCY (‘*TRANSPARENCE DU MARCHÉ DU TRAVAIL*’)

In France, labour-market intermediation services are provided both by a national public employment service, named *Pôle emploi*, and private companies (known as ‘job boards’), e.g. Monster, LeBonCoin, HelloJob, Indeed. For some time, *Pôle emploi* tried to collect as many job advertisements as possible by itself, in a logic of competition with private job boards. However, it changed its strategy in 2013 when it started to engage in sharing job advertisements with the job boards. The initiative was called labour-market transparency (*transparence du marché du travail*).

The data, shared through contractual arrangements and free of charge, consists of job advertisements containing information about the company, description of the job, qualifications required and contact data of the company. Job boards share all their job advertisements with *Pôle emploi* which, before choosing which ones to publish on its website, checks that the advertisements comply with the law forbidding discrimination and that they do not already appear on its website.

In 2016, there were 65 partners engaged in the initiative. At that time, *Pôle emploi* had aggregated on its website about 550 000 job advertisements, nearly 400 000 more than it could have collected by itself. This allows the public employment service to focus more on assisting jobseekers and providing added-value services than on collecting advertisements. Additionally, jobseekers can find nearly all the advertisements on a single website, which makes job searches easier. Finally, job boards increase traffic on their website because *Pôle emploi* redirects jobseekers to their websites when it publishes the job advertisements on its website.

MOBILITY DATA FOR OPERATING ENGLAND'S HIGHWAYS INFRASTRUCTURE

Highways England (United Kingdom) oversees over 4 300 miles (6 920 km) of road. The specific motorways under Highways England's jurisdiction ([150]) carry one third of all traffic by mileage and two thirds of all heavy goods traffic in England. Thus, these roads form a crucial support for economic productivity of the country.

Previously, data collection was arduous, time-consuming and expensive. Now, LUCA ([151]) has been working closely with Highways England to use mobile data to gather more accurate, faster and cheaper data.

Through a public tender with a allocated budget (and a voluntary negotiated compensation) that Telefónica won, Highways England got 24/7 access to mobility insights based on a specific full year of data. Updating the year requires a contract extension.

Highways England has access to anonymised data from the 4 billion network events created every day by O2 ([152]) customers. This data yields valuable insights for their modelling and infrastructure planning.

Thanks to this collaboration, Highways England was able to improve their efficiency and collect better data through using LUCA Transit. This product reduced the data-collection time period from 6 months to 7 days, a massive savings in labour hours. This resulted in a saving of millions of pounds each year on data-collection costs.

([150])Highways England website: (https://www.gov.uk/government/organisations/highways-england/about).

([151])LUCA website: (https://www.luca-d3.com).

([152])O2 website: (https://www.o2.co.uk).

MAPPING EU FISHING ACTIVITIES USING SHIP-TRACKING DATA

Knowledge of fishing activities is useful not only for fisheries research but it is also a key element for policymakers to plan activities at sea (maritime spatial planning) or to assess the impact of introducing new marine protected areas. Within this context, MarineTraffic collaborated with the European Commission's Joint Research Centre (JRC, BlueHub) by sharing its data and expertise in a big data analytics project aimed at better understanding fishing activities at sea through analysing large amounts of vessel-tracking data for the public interest.

The data used in this analysis originated from the automatic identification system (AIS) ([153]) terrestrial networks of receivers and contained information on the time, position, direction and speed of individual vessels of over 15 metres length. The data was shared directly between the teams involved. MarineTraffic provided data at no cost and under purpose-limitation conditions, and this was combined with other sources for enrichment. The processing of AIS historical data, from September 2014 to September 2015, led to the creation of the first map of EU fishing activities at European scale. Following data cleaning and classification, anonymised and aggregate results were transformed into density maps ([154]).

Several research institutes, universities, NGOs, international organisations and private-sector institutions have requested the map, since it is being used to further advance the fisheries-science community.

([153]) International Communications Union, *M.1371: Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band,* 2014, (https://www.itu.int/rec/R-REC-M.1371/en).

([154]) Available openly at: (https://data.jrc.ec.europa.eu/dataset/jrc-fad-ais1415).

NORDIC SMART GOVERNMENT

The ambition of Nordic smart government (NSG) is to make life simpler for small and medium-sized businesses (SMEs) in the Nordic region by changing the use of business data and making it available for public and private players in real-time. This will increase innovation and growth in the Nordic region and create a smoother everyday life for SMEs by greatly reducing their administrative burden. This programme is one of a kind in the Nordic region and it will create value for businesses, public authorities and society by sharing data across the region in an automatic, intelligent and secure manner. It was launched by the Nordic ministers of business in May 2018, and is supported financially by Nordic Innovation.

Annual reports are required to ensure transparency and trust in the market. Today, the situation of businesses may change rapidly, but annual reports do not reflect sudden changes. However, to many users, this is still the best data available. In addition, government regulation and administration inflict heavy administrative burdens upon the SMEs because of the reporting requirements.

The collaboration explores real-time access to business transactions from businesses' enterprise resource planning (ERP) systems: to replace reporting and enhance the sharing of data between private parties. The aim is to establish machine-readable interfaces to real-time data from companies' supporting systems and assess how data can be captured from the source.

The NSG will define the requirements for an ecosystem with a service environment with multiple data-based services made possible due to integrated digital solutions. Standardisation will enable data to be shared automatically. The data is already in the SMEs ERP systems, thus the aim is not to provide any governmental database. The focus is on creating consistency between the systems for efficient use and reuse of data. This means that the digital systems and solutions must apply standardised interfaces (APIs), which make them able to share data automatically.

The ecosystem provides real time, detailed, structured and historical data on demand, and thus serves the different data needs for both business and government decisions. Furthermore, the ecosystem will facilitate the development of new data-based products and services that can create value for both public and private players.

Smart government is a new concept for data exchange that may enhance the degree of digitisation and use of data in businesses and the public sector. Creating an ecosystem that is conducive to data sharing may also enhance B2G data sharing.

ROMANIAN BREAST CANCER MOLECULAR-PROFILING TUMOUR BOARD: THE STORY OF A SUCCESSFUL PUBLIC-PRIVATE PARTNERSHIP (PPP) IN THE E-HEALTH SECTOR

In Romania the presence of women with advanced breast cancer who were still in good physical shape yet had exhausted other therapeutic options stressed the need to use new molecular-profiling technologies, such as the comprehensive genomic profiling technology (CGP), to identify new personalised treatments. However, the CGP had not yet seen a large-scale adoption in the local clinical practice in Romania because of its high economic costs.

In the field of cancer research, CGP is a key driver of transformation in personalised cancer treatment. It utilises next-generation sequencing (NGS) to identify all classes of alterations across hundreds of genes known to drive cancer. Therefore, these tests are applicable for use across any type of cancer.

The private company Roche partnered with the Romanian public Institute of Oncology 'Prof. Dr Ion Chiricuta' Cluj-Napoca (IOCN) to provide CGP technology free of charge. The raw genome-sequencing data of each patient is collected and processed by Foundation Medicine (an asset of Roche), which shares the insights from the data with IOCN. The oncologists at IOCN then select the eligible patients and discuss the outcomes within the Romanian breast tumour board. This is the first molecular-profiling tumour board in Romania that assesses high-throughput genomic analysis for breast cancer patients with metastatic disease at presentation.

Moreover, the anonymised and aggregated genome-sequencing data, and the newly identified treatment options, are also introduced in an electronic health records (EHR) platform owned by ICON. Every 12 months, a statistical analysis is performed on the existing data in order to extract relevant insights. Finally, the findings are published by the investigators, and can be used for future clinical-decision support as well as health-care policymaking.

SAFETY-RELATED TRAFFIC INFORMATION

Vehicle-generated data has a huge potential to help achieve public-interest purposes, such as road safety.

During the high-level meeting on connected and automated driving of 15 February 2017, in Amsterdam, a public-private data task force was established to deploy data-sharing for safety-related data in real-life situations (local-hazard warning, incident management, infrastructure maintenance, etc.). Its goal is to stimulate the exchange and sharing of safety-related traffic information between industry (vehicle manufacturers and service providers) and Member States, in an architecture that would ultimately allow cross-border exchanges to foster pan-European solutions and interoperability.

The data task force has initiated a proof of concept (PoC) to validate and test general principles of data sharing, access and use, primarily of vehicle-generated data.

The traffic participants providing ITS services provide the data. Service providers and vehicle manufacturers receive and send data from and to the Member States, which have national access points that function as either actual data platforms or meta-databases with links to the data. The data is about events/incidents such as anti-lock braking system (ABS) detecting traction loss: this information is created in the vehicle and available at the original equipment manufacturer (OEM) backend and/or service provider, or events/incidents detected by multiple vehicles: data of a higher quality (duplicated and cleaned) and possibly cross-brand available on an aggregation server (neutral server and/or road operator's platform).

Service providers or road authorities can aggregate and process the data from individual vehicles to establish services to end-users. Road authorities can qualify the data received and add an additional intelligence layer, to be able to send back to vehicles and end-users higher qualified data and services, either directly or via service providers. In the current PoC, all participants in the task force have agreed not to require any compensation for processing data from any other participant.

The data can be shared from vehicle manufacturers' extended vehicle system that includes a cloud-based server from where the data can be sent to third parties. The International Organisation for Standardisation (ISO) is setting the framework for the data architecture from the vehicle side (ISO 20078, ISO 20077), but the data can be formatted in different ways depending on the current systems at different vehicle manufacturers. Service providers such as HERE and TomTom will use their current tools to aggregate and process the data from several vehicle manufacturers and provide the data in a way that allows an easy usage for safety-related end-user services.

As this is an ongoing PoC, we do not have any results yet. The intention is to use the PoC to assess and ideally eliminate areas of uncertainty prior to full deployment of an agreed way of making use of vehicle-generated data for traffic safety-related services.

SMART MOBILITY INFO AND TICKETING SYSTEM LEADING THE WAY FOR EFFECTIVE E-MOBILITY SERVICES (SMILE)

The European research project SMILE represents a showpiece for the implementation of integrated mobility. It was a joint effort between more than 10 partners that worked together from March 2012 to May 2015. Wiener Stadtwerke and Österreichische Bundesbahnen (ÖBB) was responsible for the overall project management.

The mobility platform brought together several Austrian mobility partners and operators, from public-transport companies to bike- and car-sharing providers, taxis and parking garages. In addition to intermodal route planning, SMILE provided booking and cashless payment for the whole journey in one app. The travel period of the trip and the potential carbon footprint for each choice were shown too.

Combined datasets of different mobility stakeholders were used to show diverse options of intermodality. The app indicated different options for the routing: modalities/combinations of transport means for the journey, a ticket for the use of different means of transport (e.g. from car sharing to national trains), the travel duration of the trip and the carbon footprint of each choice.

The implementation of the project required complex negotiations and the creation of a PPPs. The overall goal of the project lead (the national Austrian rail company and the local public undertaking Wiener Stadtwerke) was to foster sustainable public services by optimising intermodal route planning. Both public and private partners worked together on creating public services.

The evaluation of the project, carried out by the University of Vienna, showed that the integrated mobility and service approach and combined information management led the user to use more sustainable forms of public transport.

USING VESSEL TRAFFIC DATA FOR THE PRODUCTION OF OFFICIAL STATISTICS

At EU level, Eurostat is responsible for the development, production and dissemination of European statistics. For the maritime domain, Eurostat collects maritime transport data and publishes statistics regarding transportation of goods and passengers and vessel traffic within the EU, as well as Norway and Turkey.

In 2018, Statistics Netherlands (CBS), which is part of the European Statistical System (ESSnet) big data project ([155]), and MarineTraffic signed a memorandum of understanding aimed at sharing vessel-tracking data and knowledge to improve the quality of existing statistics and foster the production of new statistical products useful to policymakers.

The data was shared directly between the parties involved. It was anonymised and aggregated data of vessel activities, which had been collected through the automatic identification system ([156]) terrestrial networks of receivers MarineTraffic owns and operates.

The data was provided at no cost under the provision of limiting its use to research and non-commercial activities purposes only. Furthermore, the data shared was fused with datasets supplied by other providers that the project consortium had access to.

A number of tools were developed in the context of the project, including algorithms for cleaning and reducing noise in the data and various statistical products such as one to improve the calculation of ships' travelled distances between ports and travel distances within ports.

The idea behind this pilot was to investigate the possibilities for using vessel-tracking data as a fast economic indicator and a proxy for international trade.

([155]) For more information about the EESnet Big Data, please visit https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page1

([156]) International Telecommunication Union, *M.1371: Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band*, 2014, (https://www.itu.int/rec/R-REC-M.1371/en).

SMART PRAGUE STRATEGY FOR SMART MOBILITY

The Czech capital faces new challenges. How to approach the public-space development issue in a sustainable manner, which will bring an improved comfort and new services to its inhabitants? How to respond to the growing number of the capital's population and thus to greater demands on the city's infrastructure? The Smart Prague 2030 concept can answer some of these questions. The concept is based on the use of the latest technologies for transforming the capital to a better living space. The strategy of the city is ambitious: it fundamentally and structurally changes the city by means of time-proven innovative technologies.

In the sector of smart mobility, the city-owned company Operator ICT has developed a mobile application, called *My Prague* that offers to residents and visitors either static city information or dynamic data such as data from traffic cameras. The cooperation with ride-hailing and bike-sharing companies allows extending provided information.

The city plans to develop a multimodal planner that will be based not only on simple data visualisation on the map but also on data integration. According to users' preferences, the multimodal planner will combine the conventional means of transport with the ride-hailing and bike-sharing services. In the final phase, the user will be able to plan, order and pay for different mobility services in the capital and its surroundings through only one system.

The service providers get raw data from individual vehicles. The data that is shared with Operator ICT is provided in the form of aggregated data that shows the location of all parked vehicles. Providers currently do not share the location of vehicles during the whole operation time, but only when they are prepared to be used by a new client. Operator ICT can use data exclusively for the application content. Therefore, it is not permitted to share the data with third parties nor to store the data in the long term and to further analyse them. The data is transmitted through a standard representational state transfer (REST) interface. The city data platform *Golemio* then processes it while the app *My Prague* accesses the data directly from the platform.

Thanks to this collaboration, users and travellers have easier access to the location data of competing services and can easily decide which shared vehicle they will use. Service providers can take the opportunity to reach a wider target group of residents and visitors thanks to the additional communication channel. The city of Prague benefits from the increasing level of public awareness of car-sharing/bike-sharing services and from the potential reduction of citizen-owned vehicles. Finally, Operator ICT can better plan and develop new projects related to city mobility (multimodal planner) thanks to relevant experience with data sharing.

ANNEX III: MEMBERS OF THE HIGH-LEVEL EXPERT GROUP

- Mr Alberto Alemanno, Jean Monnet Professor of EU law at Haute Ecole de Commerce (HEC) Paris
- Ms Linda Austere, Chief policy officer at the Finance Latvia Association
- Mr Richard Benjamins, Data and AI ambassador at Telefónica
- Mr Laurent Cytermann, Maître de Requêtes (judge) at the Conseil d'Etat in France
- Mr Michel Isnard, Inspecteur Général at the French Statistical Office
- Ms Dörthe Koerner, Legal adviser at the Brønnøysund Register Centre
- Ms Petra Konecna, Senior consultant at Deloitte Advisory and project manager and adviser for a start-up, Obce v datatech
- Ms Helena Koning, Data protection officer and senior legal counsel at Mastercard
- Mr Juan Murillo Arias, Senior manager on data strategy and data science innovation at the Spanish bank BBVA
- Mr Paulo Nogueira, Director of the directorate of analysis and information at the Portuguese directorate-general of health
- Ms Nuria Oliver, Chief data scientist at Data-Pop Alliance and chief scientific adviser at Vodafone Institute for Society and Communications
- Mr Milan Petkovic, Vice-President of the Big Data Value Association (BDVA), part-time professor and head of data science at Philips
- Mr Dominik Rozkrut, President of the Polish statistical office
- Ms Greta Schoeters, Environmental health manager at VITO
- Ms Eva Schweitzer, Deputy Head of unit and researcher for urban research on digital cities, risk prevention and transport at the German federal institute for research on building, urban affairs and spatial development
- Ms Virpi Johanna Stenman, Project manager at the Finnish forest centre
- Ms Ioana Raluca Stoenescu, Government affairs manager at Roche Holding — Roche Romania

- Mr Per-Olof Svensk, Senior advisor and project manager at the Swedish transport administration
- Mr Joost Vantomme, Director of smart mobility at ACEA (European Automobile Manufacturer's Association)
- Mr Stefaan Verhulst, co-founder of and Chief of R & D at The GovLab, New York University
- Ms Susan Wegner, Vice-president on data, artificial intelligence and governance and chief data officer at Deutsche Telekom
- Mr Erik Wetter, Assistant professor at the Stockholm School for Economics and co-founder and Chair of the Flowminder Foundation
- Mr Dimitris Zissis, Associate professor at the University of the Aegean in Greece and Head of R & D at MarineTraffic

ANNEX IV: MANDATE OF THE HIGH-LEVEL EXPERT GROUP

As envisaged in the 2017 mid-term review of the digital single market strategy, the Commission intends to support the creation of a common European data space: a seamless digital area with the scale to enable the development of new products and services based on data. Data should be available for reuse as much as possible, as a key source of innovation and growth. The measures announced in the 2018 Communication, *Towards a common European data space* to help create the common European data space cover different types of data and therefore have different focuses:

- a proposal for a review of the directive on the reuse of public-sector information (PSI directive);
- an update of the 2012 recommendation on access to and preservation of scientific information; and
- guidance on sharing private-sector data among companies and with public-sector bodies for public-interest purposes.

The public consultation on the review of the directive on the reuse of public-sector information showed support for improving access to private-sector data by public authorities for public-interest purposes in general ('business-to-government' or 'B2G' data sharing). However, private-sector data holders expressed concerns over data confidentiality, perceived risks to companies' commercial interests and absence of clearly-defined compensation models allowing companies to recover the investments made into sharing the data.

To address these concerns, and in order to help the public sector to embrace the opportunities for using business data for the public interest, the Commission outlined governing principles for B2G data-sharing collaborations in its 2018 communication. These principles are accompanied by *Guidance on private-sector data sharing*, which details some practical and legal considerations of B2G data collaborations. The Commission also committed to setting up an expert group on access to and reuse of private-sector data for public-interest purposes.

ABOUT THE HIGH-LEVEL EXPERT GROUP ON B2G DATA SHARING

On 24 August 2018, the Directorate-General for Communications Networks, Content and Technology (DG CNECT) published a call for applications for the selection of up to 24 members for an expert group on access to and reuse of private-sector data for public-interest purposes.

Composition

High-Level Expert Group on B2G Data Sharing is composed of 23 experts with knowledge on sharing private-sector data with public organisations and reusing it for public-interest purposes. They act independently, in their personal capacity and in the public interest (see Annex III).

The group reflects a balanced representation of high-level experience in different areas (e.g. statistics, health, transport, law, ICT) and sectors (i.e. with experience in the public and/or private sector), as well as of geography and gender.

In addition, the expert group includes three international organisations (the Organisation for Economic Cooperation and Development (OECD), the World Bank and the UN global pulse), who were granted the status of observers. They took part in the discussions of the group and provided expertise. Furthermore, the group's secretariat ensured interaction with other Commission directorates-general (DGs) and expert groups whenever necessary or relevant for the work of the group.

Mandate

The mandate of the High-Level Expert Group is to assist the Commission by:

1. identifying good practices on private-sector data sharing in B2G contexts in order to contribute to more efficient and better public-service delivery and/or more reliable and evidence-based policymaking in fields such as statistics, traffic management, protection of the environment, crisis management and natural disasters, food safety, protection of human and animal health, development and cooperation and management of smart cities;
2. assessing the legal, economic and technical obstacles preventing B2G data sharing, and advise on actions to promote B2G data-sharing for public-interest purposes;
3. giving recommendations to the Commission on how to further develop its policy on the reuse of private-sector data for public-interest purposes in the European Union.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

Publications Office
of the European Union