



Enhancing Access to and Sharing of Data

**RECONCILING RISKS AND BENEFITS
FOR DATA RE-USE ACROSS SOCIETIES**



Enhancing Access to and Sharing of Data

RECONCILING RISKS AND BENEFITS FOR DATA
RE-USE ACROSS SOCIETIES

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Please cite this publication as:

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

ISBN 978-92-64-76402-6 (print)

ISBN 978-92-64-66065-6 (pdf)

Photo credits: Cover © Adobe Stock.

Corrigenda to publications may be found on line at: www.oecd.org/about/publishing/corrigenda.htm.

© OECD 2019

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

Foreword

This report examines the opportunities of enhancing access to and sharing of data (EASD) in the context of the growing importance of artificial intelligence and the Internet of Things. It discusses how EASD can maximise the social and economic value of data re-use and how the related risks and challenges can be addressed.

In particular, the report examines the different approaches and strategies available to policy makers and business leaders when aiming at establishing data-governance frameworks for data access and sharing. To this end, the report highlights the factors that need to be taken into account including data typologies, key data-access mechanisms and the main types of actors and their roles. The report also presents the available evidence of the direct and indirect economic and social benefits of data access and sharing. It then describes the major challenges facing policy makers, and presents current trends in public policies aimed at addressing these challenges, with examples of EASD approaches and policy initiatives in OECD countries and partner economies.

The initial draft was developed as a background report for the OECD Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use”, held on 2-3 October 2017 in Copenhagen (Denmark) (Copenhagen Expert Workshop, www.oecd.org/internet/ieconomy/expert-workshop-enhanced-access-to-data-reconciling-risks-and-benefits-of-data-re-use.htm).

The report was drafted by Christian Reimsbach-Kounatze and Elettra Ronchi, with contributions by Suguru Iwaya. It benefitted from the input of the informal Joint Steering Group established to support this work. The work was made possible by the generous contributions of Denmark, Finland, Italy, Japan, Norway, and Sweden.

A third revision of the report was discussed and approved by the OECD Working Party on Security and Privacy in the Digital Economy (SPDE) at its meeting in November 2018, and declassified by the Committee on Digital Economy Policy (CDEP) by written procedure on 6 March 2019.

The report contributes to the horizontal project on EASD under three partner committees: the CDEP, the Committee for Scientific and Technological Policy (CSTP), and the Public Governance Committee (PGC). The report complements other reports produced under the last two committees, namely the forthcoming synthesis report *Enhanced Access to Public Data for Science, Technology and Innovation*, which is based on insights stemming from analytical work performed by the CSTP between 2017 and 2018, and the 2018 OECD *Open Government Data Report*, which was written by the PGC Working Party of Digital Government Officials (E-Leader) and provides an overview of the state of open data policies across OECD countries and partner economies.

Table of contents

Foreword	3
List of acronyms, abbreviations and units of measure.....	9
Executive summary	11
1. Introduction	15
Barriers to data access, sharing and re-use	17
Objectives and structure.....	18
Implications for public policies and business strategies on data access and sharing.....	19
References.....	20
Notes	21
2. Understanding enhanced access to and sharing of data	23
Different types of data and access control mechanisms	25
Personal data and the degrees of identifiability: Reflecting the risk of harm.....	26
The overlapping domains of data: Reflecting the various stakeholder interests	27
The manner data originates: Reflecting the contribution to data creation.....	29
Data-access control mechanisms: Protecting the interests of data holders	32
Main types of actors and their roles	34
Data holders and controllers.....	35
Data users	35
Data intermediaries	36
Approaches to access and sharing and their degree of openness	39
Contractual agreements and data markets	39
Open data	41
Data portability.....	43
Other restricted data-sharing arrangements.....	44
References.....	49
Notes	54
3. Economic and social benefits of data access and sharing	59
Impact assessment studies on the economic and social benefits.....	60
Enhancing access to public-sector data	60
Enhancing access to and sharing of public and private-sector data	62
Main categories of economic and social benefits	64
Transparency, accountability and empowerment of users	64
Business opportunities including for data intermediaries and start-ups.....	65
Co-operation and competition across sectors and countries.....	66
Crowdsourcing new insights and user-driven innovation	69
Increased efficiency across society through data linkage and integration.....	70
References.....	71
Notes	74

4. Risks and challenges of data access and sharing	77
Need for balancing the benefits of data “openness” with other legitimate interests, policy objectives and risks	78
Digital security risks and confidentiality breaches in particular	80
The violation of privacy, intellectual property rights and other interests.....	81
The difficulty of applying a risk management approach.....	85
Barriers to cross-border data access and sharing.....	87
Trust and empowerment for the effective re-use of data across society	89
Supporting and engaging communities of stakeholders.....	89
Capacity building: Fostering data-related infrastructures and skills	90
Lack of common standards for data sharing and re-use	92
Data quality	94
Misaligned incentives, and limitations of current business models and markets	95
Externalities of data sharing and re-use and the misaligned incentives	95
Limitations of current business models and data markets.....	96
The risks of mandatory access to data.....	97
Uncertainties about “data ownership”	98
References.....	103
Notes	110
5. Policy initiatives enhancing data access and sharing	115
Governments leading by example in enhancing access to and sharing of public-sector data.....	117
Access to open government data and public-sector information.....	117
Facilitating data sharing within the public sector.....	119
Geospatial and transportation data: A highly valued public-sector data	120
Facilitating or regulating data access and sharing within the private sector.....	121
Voluntary and collaborative approaches	121
Data portability.....	125
Data of public interest	126
Increasing data analytic capacities across society.....	128
Supporting the development of data-related skills and infrastructures	128
Establishing and collaborating with data analytic support centres.....	129
Supporting innovation and R&D in data analytics and related technologies	130
Achieving greater policy coherence through national and sectoral data strategies	130
References.....	132
Notes	135

Tables

Table 2.1. Data categories based on origin.....	30
Table 3.1. The economic benefits and savings of open data by TfL	62
Table 3.2. Selected use cases in logistics of the Industrial Data Space.....	67

Figures

Figure 1.1. Trends in the acquisition of big data and analytics firms.....	16
Figure 1.2. European individuals restricting the use of their personal information over the Internet, 2016.....	18
Figure 2.1. The degrees of data openness and access.....	25

Figure 2.2. The personal, private and public domains of data.....	29
Figure 2.3. Data products and the different manners data originates	31
Figure 2.4. Flowminder.org’s data processing flow.....	34
Figure 2.5. High Definition 3D Map Data Platform.....	39
Figure 4.1. Data specialists in selected OECD countries	91
Figure 4.2. Enterprises using cloud computing services, by firm size, 2016	92
Figure 5.1. Number of government policy initiatives enhancing data access and sharing.....	116
Figure 5.2. Certification system for data-sharing platforms.....	125

Boxes

Box 2.1. Definitions of EASD adopted in this report.....	24
Box 2.2. Twitter’s data control over its application programming interface.....	33
Box 2.3. The diversity of revenue models in the data ecosystem.....	40
Box 2.4. Open data in the private sector: The case of Thomson Reuters.....	42
Box 2.5. Rationale for promoting data sharing on digital security risk.....	45
Box 2.6. Data-sharing communities in science: The case of the National Cancer Institute’s Genomic Data Commons	46
Box 2.7. Data philanthropy: The case of MasterCard	48
Box 3.1. Data portability as business facilitator? The case of Uber and Braintree	67
Box 3.2. Data portability, competition and co-operation: The case of two platforms	68
Box 4.1. Balancing the benefits with the risks: Australia’s data-sharing and release legislation	79
Box 4.2. Data ethics: Government initiatives in Denmark.....	82
Box 4.3. Data ethics: Government initiatives in the United Kingdom.....	83
Box 4.4. Managing the risk of disclosure: The Five Safes Framework	87
Box 4.5. The Data Transfer Project: A private-sector initiative for data portability	93
Box 4.6. The “essential facility” concept	98
Box 4.7. Data ownership controversies in data-driven agriculture	99
Box 5.1. The Revision of the EU Directive Public-sector Information.....	118
Box 5.2. The Core Principles of Ag Data.....	123
Box 5.3. Proposed EC Principles for Contractual Agreements on Non-Personal Data.....	124
Box 5.4. Fostering interoperability through enhanced access and sharing: The case of Finland’s (2018) Act on Transport Services.....	127
Box 5.5. Colombia’s Excellence and Appropriation Centre in Big Data Analytics.....	128

List of acronyms, abbreviations and units of measure

ABS	Australian Bureau of Statistics
AFBF	American Farm Bureau Federation
AI	Artificial intelligence
API	Application programming interface
ATP	Agriculture technology provider
AUD	Australian dollar
B2B	Business-to-business
B2C	Business-to-consumer
CDR	Consumer Data Right (Australia)
CEA	Excellence and Appropriation Centre (Colombia)
CMS	Centers for Medicare and Medicaid (United States)
DAC	Data Access Committees (National Cancer Institute)
DDI	Data-driven innovation
DEO	<i>Digital Economy Outlook</i>
DIPA	Data Integration Partnership for Australia
DoS	Denial of service
DS&R	Data sharing and release legislation (Australia)
DTP	Data Transfer Project (Google)
EASD	Enhancing access to and sharing of data
EC	European Commission
EFD	Essential facilities doctrine
ERB	Ethics review body
EU	European Union
EUR	Euro
G20	Group of Twenty
G7	Group of Seven
GBP	British pound
GDC	Genomic Data Commons (National Cancer Institute)
GDP	Gross domestic product
GDPR	General Data Protection Regulation (European Union)

GDS	Genomic Data Sharing (National Institutes of Health policy)
GPHIN	Global Public Health Intelligence Network
HCCI	Health Care Cost Institute (United States)
ICO	Information Commissioner's Office (United Kingdom)
ICT	Information and communication technology
IDS	Industrial Data Space
ILS	New Israeli shequel
IoT	Internet of Things
IPR	Intellectual property right
IT	Information technology
JPY	Japanese yen
M&A	Merger and acquisition
MNP	Mobile number portability
NCI	National Cancer Institute (United States)
NDAC	National Data Advisory Council (Australia)
NDC	National Data Commissioner (Australia)
NIH	National Institutes of Health (United States)
NZD	New Zealand dollar
OECD	Organisation for Economic Co-operation and Development
OWL	Web Ontology Language
PB	Petabyte (= 1 million gigabytes)
PDS	Personal data store
PIMS	Personal information management system
PPP	Public-private partnership
PSI	Public sector information
PSIH	Public-sector information holder
R&D	Research and Development
SMEs	Small and medium-sized enterprises
TB	Terabyte (= 1 000 gigabytes)
TfL	Transport for London
US	United States
USD	United States dollar
VRDC	Virtual Research Data Center (Centers for Medicare and Medicaid)

Executive summary

Current developments related to data-driven innovation, including in the context of the Internet of Things and artificial intelligence, have made data access and sharing more crucial than ever. Evidence presented in this report shows how enhancing access to and sharing of data (EASD) can help maximise the social and economic value of data re-use. It can increase the value of data to the data holder, and even more so to secondary data users, with additional positive spill-over benefits for country economies and society at large.

Overall, data access and sharing is estimated to generate social and economic benefits worth between 0.1% and 1.5% of gross domestic product (GDP) in the case of public-sector data, and between 1% and 2.5% of GDP (in a few studies up to 4% of GDP) when also including private-sector data. The estimated magnitude of the effects depends on the scope of the data and the degree of data openness, according to the studies discussed in this report.

However, despite a growing need for data and evidence of the economic and social benefits, data access and sharing has not achieved its potential. Individuals, businesses, and governments often face barriers to data access, which may be compounded by reluctance to share. To facilitate, encourage and enhance data access and sharing for the benefit of all, the following three major challenges need to be addressed:

1. **Balancing the benefits of enhancing data “openness” with the risks, while considering legitimate private, national and public interests.** These risks may include, but are not limited to, the risk of (personal) data breaches that could threaten individuals’ privacy and the violation of commercial and non-commercial interests. A risk management approach may help mitigate risks, balance trade-offs and promote data access and sharing, including across borders. However, a risk management approach remains challenging to implement for most organisations, including small and medium-sized enterprises.
2. **Reinforcing trust and empowering users through pro-active stakeholder engagement and community building.** Communities of data users and data holders can facilitate data sharing and help maximise the value of data re-use, although attention is required to address the risk of anti-competitive effects that could result from data-sharing partnerships among (potential) competitors. These communities can be heterogeneous and may include stakeholders with conflicting or opposing interests. Their establishment and management may involve significant costs, including for the development of data-related skills, infrastructures, and standards as well as for maintaining engagement.
3. **Encouraging the provision of data through coherent incentive mechanisms and sustainable business models while acknowledging the limitations of (data) markets.** This may require i) aligning current incentives structures and ii) promoting where possible the use of business models for the sustainable provision and commercialisation of data, while iii) considering the detrimental effects of mandatory access to data. It may also require iv) reducing uncertainties about “data ownership” by acknowledging the role of intellectual property rights, other ownership-like rights,

and the importance of *freedom of contract* in business-to-business (B2B) data agreements and the role of data commons for the governance of shared resources of common interests.

When addressing these multidimensional challenges, policy makers need to avoid the “data policy pitfall” by seeking one-size-fits all (“silver-bullet”) solutions. This report shows that there is no single optimal level of data “openness”; the value of data access and sharing depends on the type of the data and the context in which that data are re-used, including the social, economic, and cultural environment in which activities take place.

In promoting data access and sharing policy makers will thus need to take account of: i) the sensitivity of the data and the degree by which personal data could be re-identified; ii) the overlapping rights and interests of all relevant stakeholders; and iii) the manner by which data are generated, in order to better take into account the contributions of the various stakeholders in the creation of that data.

The report further examines more than 200 policy initiatives across 37 countries, which reveal the following policy trends:

- **Most policy initiatives still focus on public-sector data** (almost 65% of all initiatives). Most of these initiatives focus on open access to government data (open government data). There is also a noticeable trend towards facilitating data sharing within the public sector (almost 15% of all initiatives on public-sector data) and enhancing access to and sharing of geospatial data (e.g. maps) and transportation data.
- **Few countries have policy initiatives to facilitate data sharing within the private sector** (almost 15% of all initiatives), although sharing and re-use of private-sector data is the emerging challenge most frequently cited. Most of the policy initiatives (around 55%) are based on voluntary schemes. Among those that are mandatory, most focus on regulating access to i) “data of public interests”; and ii) data of network industries (e.g. transportation and energy) for ensuring the interoperability of “smart” services. In a few countries, data portability with a focus on consumer data is emerging as another policy means for promoting data access and sharing in the private sector.
- **Increasing data analytic capacities, either in the public or private sector**, is addressed by only 12% of all policy initiatives, despite the need for complementary investment in data-related skills and infrastructures. A quarter of these initiatives focus on the establishment of data analytic (technology) centres that provide support and or guidance in the re-use and analysis of data for public and/or private-sector entities. Fewer initiatives focus on supporting investments in data-related innovation and research and development.

Recognising that data openness is a continuum (rather than a binary concept) and based on the good practice and use cases examined, this report discusses several approaches that policy makers may wish to consider when encouraging, facilitating and enhancing data access and sharing. These approaches represent different strategies along the data “openness” continuum that can be leveraged to address major challenges discussed in the report, among which the most prominent are:

- **Contractual agreements**, which remain crucial as a market-based approach to enhance data access and sharing in a B2B context, in particular if leveraged through *data markets*.

- **Open data**, the most extreme approach to data openness and the most frequently used by governments, which is not always appropriate for EASD. Legitimate private and public interests may justify more restricted approaches to data access and sharing.
- **Data portability**, which provides restricted access to those involved in the creation and collection of data, such as data subjects in respect to their personal data. It promises to empower users by giving them more control over their data, but it may also expose them to new risks.
- **Restricted data-sharing arrangements**, which are used where data are considered too confidential to be shared openly with the public (as open data) or where there are legitimate (commercial and non-commercial) interests in conflict with such open sharing. These arrangements are found typically in science and health care research, but they also exist across the economy as data partnerships and “data for social good” initiatives.

The analysis uncovers the growing need for data-governance frameworks that incorporate whole-of-government approaches and are coherent across areas, sectors and ideally countries. A few countries are in the process of establishing national data strategies to meet this need. Further analytical work is needed to examine the commonalities among these strategies and the challenges countries face in their development and implementation.

1. Introduction

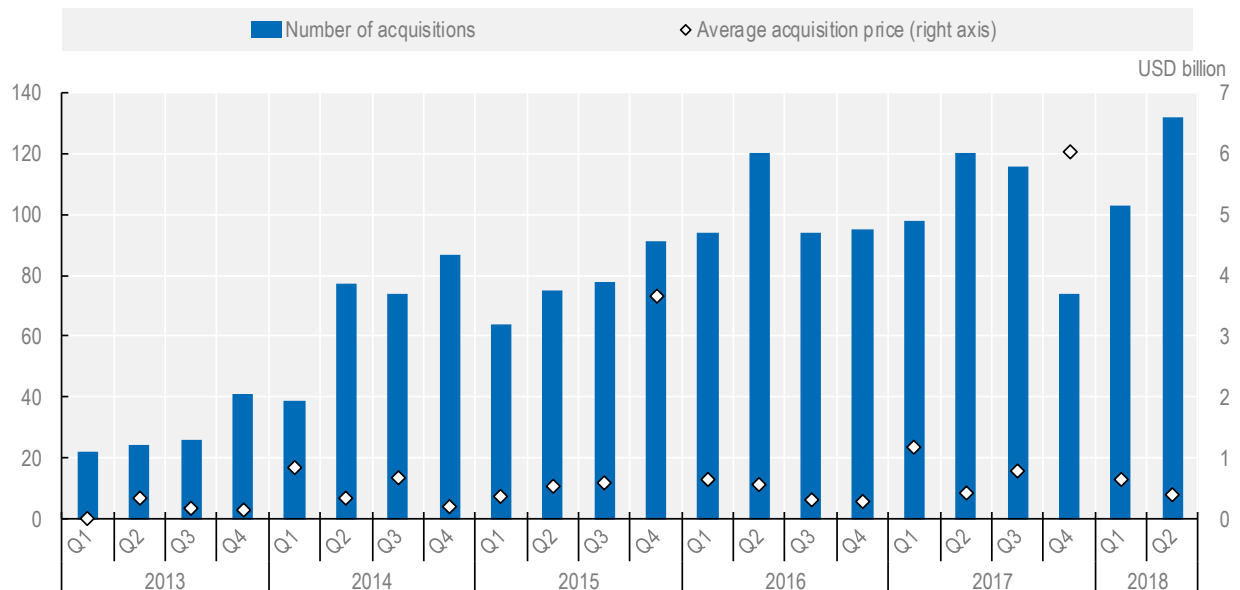
This introductory chapter explains the importance of data access and sharing in the context of current technological developments. It points to barriers to data access, sharing and re-use and some of the issues further discussed in the following chapters. It concludes by presenting the objective of the report and an overview of its structure.

The effective use of data,¹ in combination with data analytics (software), generates information of social and economic value. It can help boost productivity and improve or foster new products, processes, organisational methods and markets. This is referred to as “data-driven innovation” (DDI) (OECD, 2015^[1]). With the increasing use of artificial intelligence (AI) and the Internet of Things (IoT) the supply of, and demand for, data will increase even in traditionally less data-intensive fields, and this to a level that very few organisations will be able to meet alone. A single self-driving car, for example, can generate between 1 terabyte (TB) and 5 TB of data per hour according to some estimates (Grzywaczewski, 2017^[2]; Nelson, 2016^[3]).² Yet, even more third-party data are required for these systems to operate securely irrespective of weather conditions, visibility or road-surface quality.³

Access to data is thus crucial for competition and innovation in the digital economy. Not only for businesses, but also for governments and individuals, including researchers. Access to data can, for instance, help enhance public-service delivery and facilitate the identification of emerging governmental and societal needs. It can help improve forecasting and the reliability of infrastructures (such as in transportation and utilities) and increase their efficiency. In science and research access to data can help review and replicate scientific results, and foster new instruments and methods of data-intensive exploration and scientific experimentation.

The economic importance of data access is reflected in the growing number of mergers and acquisitions (M&As) of data-intensive firms. These M&As are meant to assure access to business-critical data. Some of the largest M&As motivated by access to big data in the last five years include: Monsanto’s acquisition of the Climate Corporation, an agriculture analytic firm, for USD 1.1 billion in 2013; IBM’s acquisition of a majority share of the Weather Company, a weather forecasting and analytic company, for over USD 2 billion in 2015 (Waters, 2015^[4]); and Alibaba’s total investment of USD 4 billion between 2016 and 2018 to acquire Lazada, a leading e-commerce platform founded in 2012 in Singapore. Start-ups specialised in big data are also increasingly the target of acquisitions. The annual number of these acquisitions increased from more than 100 acquisitions in 2013 to more than 400 acquisitions in 2017, with the average price paid exceeding USD 1 billion in some quarters (Figure 1.1).

Figure 1.1. Trends in the acquisition of big data and analytics firms



Source: OECD based on Crunchbase data.

These developments underline the growing social and economic value of data, and the need for access to and sharing of data. The intangible and non-rivalrous nature of data allows a wide range of means of access and sharing, including the commercialisation of data under non-discriminatory conditions and open data. In contrast to rivalrous goods such as oil, which is depleted once extracted, transformed and consumed during production processes, the use of data does not exhaust the supply of data and (therefore) *in principle* its potential to meet the demands of others (OECD, 2015^[1]). Where the value of (secondary) data re-use for society (i.e. social and market value) is larger than the value (of primary data use) for the individual member of society (private value), access to and sharing of data can maximise the re-use and therefore the value of data across organisations, sectors and economies.⁴

Yet, despite the growing need for data and the evidence presented in this report of the economic and social benefits of data re-use and sharing, data access and sharing remains below its potential.

Barriers to data access, sharing and re-use

There are still significant barriers to data sharing and re-use, even within public organisations. The social and economic risks associated with the possible revelation of confidential information (e.g. personal data and trade secrets)⁵ are often the main rationale for individuals and organisations not sharing their data. Identifying which data to share and defining the scope and conditions for access and re-use is perceived as a major challenge, in particular for individuals and small- and medium-sized enterprises. This remains true even in cases where commercial and other private interests would not oppose data sharing and re-use (AIG, 2016^[5]).

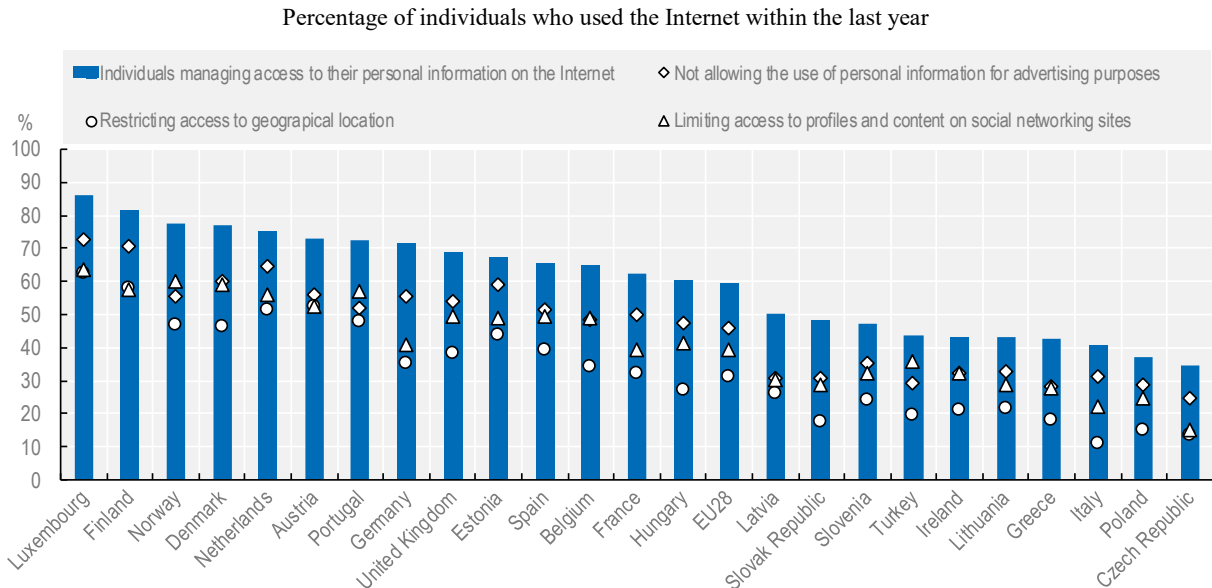
Furthermore, a survey by the Economist Intelligence Unit (2012^[6]) reports that almost 60% of companies stated that “organisational silos” are the biggest impediment to using “big data” for effective decision-making. Individuals are also increasingly wary of the re-use of their personal data. In Europe, for example, individuals i) limit the use of their personal data⁶ for advertising purposes (40% of all the surveyed population in 2016); ii) limit access to their social networking profiles (35%); and iii) restrict access to their geographic location (30%) (Figure 1.2).

As the provision of high-quality data can require significant up-front and follow-up investments, incentives to share data are often too low, in particular when individuals and organisations cannot sufficiently appropriate the returns on their investments. This is in particular true as complementary resources (e.g. additional metadata, data models and algorithms for data storage and processing and even secured information technology infrastructures) have to be made available before the data can be re-used effectively.⁷ These concerns are sometimes exacerbated by the legal complexities and uncertainties related to privacy regulation (e.g. consent), but also cross-border data access and sharing and the question of data “ownership” (see subsection “Uncertainties about ‘data ownership’” in Chapter 4).

For example, some individuals may object to the re-use of their health-related data for research purposes because of confidentiality concerns, though they may be aware of the social benefits that can be derived for themselves and for society. And though there is real market demand, some organisations may be reluctant to share or even sell or license their proprietary data because they cannot perceive their market value or because the cost of making that data available appear higher than the expected benefits. In some cases, organisations may be willing to share the data only under the condition that other organisations do the same or that there are clear benefits for them. According to an AIG-commissioned survey of 400 employees and 250 business executives across nine countries,⁸ more than two-thirds of respondents said they would “engage in the safe sharing of data if they received some

benefits from doing so” (AIG, 2016^[5]). Lack of reciprocity and concerns of “free riding” may thus require co-ordination across industries (i.e. a collective action problem) and strong leadership to establish a culture of trust for data sharing across society.

Figure 1.2. European individuals restricting the use of their personal information over the Internet, 2016



Source: OECD (2017^[7]), *OECD Digital Economy Outlook*, <https://dx.doi.org/10.1787/9789264276284-en>, based on Eurostat Digital Economy and Society Statistics, <http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/comprehensive-database> (data accessed March 2017).

In addition, the lack of dedicated funding for data-sharing infrastructures and the limited pathways for their sustainment even in critical areas like science and health care research, combined with the misalignment of incentives to invest in, curate and share data have increased the risk of data erosion over time. According to Vines et al. (2014^[8]), for instance, the probability of finding the data associated with most scientific papers declines by 17% every year.

All this may lead to significant (social and economic) opportunity costs. As Rufus Pollock, Founder and President of Open Knowledge International, stated at the OECD Technology Foresight Forum in October 2012: “The best thing to do with your data will be thought of by someone else.” This is particularly the case where the spill-overs of data cannot be easily observed or quantified (e.g. socialisation and behavioural change, cultural and scientific exchange, or greater levels of trust induced by transparency). As a result, countries’ capacity to innovate may risk being undermined if less data can be used as input to innovation, in particular in the current age of AI.

Objectives and structure

This report examines the opportunities and challenges of enhancing access to and sharing of data (EASD). It discusses in particular how EASD can be an effective means for maximising the social and economic benefits of data and data re-use, while, at the same time, addressing related risks and challenges and protecting the private interests of individuals and organisations. It provides examples of some approaches to EASD that can enable the free flow of data across nations, sectors and organisations, and at the same time address the legitimate concerns

of individuals and organisations, including governments, while assuring DDI, growth and well-being across societies.

The report builds on the findings of the OECD Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use”, held in Copenhagen (Denmark) on 2-3 October 2017 (Copenhagen Expert Workshop) and the findings of the Joint CSTP-GSF Workshop “Towards New Principles for Enhanced Access to Public Data for Science, Technology and Innovation”, held at the OECD on 13 March 2018 (CSTP-GSF Workshop) and the Open Government Workshop held in Stockholm (Sweden) on 16 March 2018 and organised by OECD Directorate for Public Governance (GOV) (Stockholm Open Government Workshop).

The rest of the report is structured as follows: Chapter 2 presents data typologies, key data access mechanisms, and the main types of actors and their roles, and then examines the different policy approaches and degrees of openness of EASD. Chapter 3 highlights the benefits of data access and sharing. Chapter 4 discusses the risks and challenges. Chapter 5 presents recent government policy initiatives that promote data access and sharing.

Implications for public policies and business strategies on data access and sharing

The policy issues identified in this report (Chapter 4) require differentiated approaches to data access and sharing. Open data, the most extreme approach to data access and sharing and the most commonly used by policy makers, remains highly relevant, in particular for public-sector and research data. But other approaches and strategies are available to policy makers and business leaders, and they are being adopted across application areas (Chapter 2).

Overall, this report shows that there is no one-size-fits-all optimal level of data “openness”. The optimal level depends on the context of data access, sharing and re-use, including the social, economic, and cultural environment in which these activities take place. At the same time, the report recognises the need for data-governance frameworks that incorporate a whole-of-government approach and are coherent across application areas, sectors and ideally countries. It highlights a number of differentiating factors that data-governance frameworks need to reflect to enhance access to and sharing of data for the benefits of all.

The findings of this report are not only relevant for public policies and business strategies related to data access and sharing. They also have implications on the legal instruments related to data access and sharing developed by the OECD. To this date, the OECD has developed seven Council Recommendations that address data access and sharing directly or indirectly. These include:

- the OECD (2006^[9]) *Recommendation of the Council concerning Access to Research Data from Public Funding*
- the OECD (2008^[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*
- the OECD (2009^[11]) *Recommendation of the Council on Human Biobanks and Genetic Research Databases*
- the OECD (2013^[12]) *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*
- the OECD (2014^[13]) *Recommendation of the Council on Digital Government Strategies*
- the OECD (2016^[14]) *Recommendation of the Council on Health Data Governance*.

Member countries are currently reviewing these legal instruments to ensure their coherence and continued relevance, in light of the findings of this report.

References

- AIG (2016), *The Data Sharing Economy: Quantifying Tradeoffs that Power New Business Models*, <http://www.aig.com/content/dam/aig/america-canada/us/documents/brochure/the-data-sharing-economy-report.pdf>. [5]
- Economist Intelligence Unit (2012), *The deciding factor: big data & decision making*, <http://www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/>. [6]
- Grzywaczewski, A. (2017), *Training AI for Self-Driving Vehicles: the Challenge of Scale*, <https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/>. [2]
- Nelson, P. (2016), “Just one autonomous car will use 4,000 GB of data/day”, *NetworkWorld*, <http://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html>. [3]
- OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/eedfee77-en>. [15]
- OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264276284-en>. [7]
- OECD (2016), “Health Data Governance Recommendation”, in *Recommendation of the Council on Health Data Governance*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433>. [14]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [1]
- OECD (2014), *Recommendation of the Council on Digital Government Strategies*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0406>. [13]
- OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>. [12]
- OECD (2009), *Recommendation of the Council on Human Biobanks and Genetic Research Databases*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0375>. [11]
- OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/122/122.en.pdf>. [10]
- OECD (2006), *Recommendation of the Council Concerning Access to Research Data from Public Funding*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>. [9]

- Simonite, T. (2018), *Some Startups Use Fake Data to Train AI*, [16]
<http://www.wired.com/story/some-startups-use-fake-data-to-train-ai/>.
- Vines, T. et al. (2014), “The Availability of Research Data Declines Rapidly with Article Age”, [8]
Current Biology, Vol. 24/1, pp. 94-97, <http://dx.doi.org/10.1016/j.cub.2013.11.014>.
- Waters (2015), “IBM’s latest deal is a new test case for the big data economy”, *Financial Times*, [4]
<http://www.ft.com/content/0fe3ac2e-7e22-11e5-a1fe-567b37f80b64>.

Notes

¹ The term “data” can have multiple meanings. Depending on the context and jurisdiction, for example, “data” can be used to refer to: raw or unprocessed data, whether in analogue or electronic format; personal information; information in electronic form (including reports, maps, and photographs, which are sometimes broadly referred to as “digital content”); or all recorded information. “Data” can refer to “factual records” including single, or a large collection of, items such as numerical scores, textual records, images and sounds used as primary sources for scientific research (see e.g. “research data” in OECD (2006_[9])). In other words, in some contexts, “data” is used interchangeably with “information”, and in other contexts is distinguished from “information”, where the latter is understood as “the meaning resulting from the interpretation of data” (OECD, 2015_[1]). For the purposes of this report, the term “data” covers only electronic versions of data (digital data) and distinguishes between raw data and information as defined in OECD (2015_[1]). Subcategories [which create important distinctions for policy makers] including personal data are further defined and explained in Chapter 2. These and other distinctions are provided in more detail also in endnotes 3 and 7 of this chapter, and endnotes 4, 10, 12, 24, 30 and 31 in Chapter 2.

² As a comparison, an average person is estimated to generate up to 1.5 gigabytes per day by 2020 (Nelson, 2016_[3]).

³ Although “synthetic data” (i.e. data generated via computer simulations) can be used to increase the volume of training data, many real-life problems such as driving are too complex to be simulated realistically and therefore still necessitate access to real life data (OECD, 2019_[15]; Simonite, 2018_[16]). As Intel chief executive officer Bria Krzanich explained, besides the technical (sensor) data, there will also be a need for “societal data, also called crowd-sourced data”, which include anonymised data from online platforms such as e.g. Waze, as well as personal data on individual driving patterns (Nelson, 2016_[3]).

⁴ In this report, the term “sharing” refers to the joint use of a resource such as data, be it in exchange of other resources (money or other goods or services) or for free. It thus includes the re-use of data based on both, data commercialisation, and open access to data (free of costs) and other non-commercial provisions and uses of data.

⁵ This includes “undisclosed information” according to Art. 39 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which provides some conditions for the information to be considered a trade secret: i) The information must be “secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question”. ii) It must have “commercial value because it is a secret”. And iii) it must have been “subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep

it secret”. Art. 39 (3) TRIPS also includes provisions for undisclosed test or other data submitted to obtain regulatory approval for the marketing of pharmaceutical or of agricultural chemical products.

⁶ Personal data is defined by the OECD (2013^[12]) *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* as “any information relating to an identified or identifiable individual (data subject)”.

⁷ For example, data from the distributed array telescope may create large data sets, which however require additional meta-data on the direction of the telescopes to be interpreted correctly.

⁸ The countries included Australia, France, Germany, Italy, Japan, United Kingdom, United States, China and Singapore.

2. Understanding enhanced access to and sharing of data

This chapter introduces the concept of enhanced access to and sharing of data and its various degrees of openness. The chapter then examines the different approaches and strategies available to policy makers and business leaders when aiming at establishing data-governance frameworks that do enough justice to important specificities but are comprehensive enough to be coherently applicable across application areas. To this end, the chapter highlights the factors that need to be taken into account including data typologies, key data-access mechanisms and the main types of actors and their roles.

Although definitions may vary, enhancing access to and sharing of data (EASD) can be described broadly as being any practical and lawful means through which digital data (data) can be effectively accessed by, and shared with, an entity (individual or organisation) other than the data holder, for the purpose of fostering data re-use by the entity, or a third-party chosen by the entity. At the same time, the private interests of individuals and organisations concerned must be taken into account, as well as national security and public interests (see Box 2.1 for further details). EASD may be based on voluntary and mutually agreed commercial or non-commercial terms – such as in the case of contractual agreements between information technology (IT) vendors and their customers – or be regulated by law and mandatory – such as in the case of data portability according to Article 20 of the European Union (2016^[1]) General Data Protection Regulation (GDPR).

Box 2.1. Definitions of EASD adopted in this report

In this report, “**data access**” is used to describe the act of retrieving and storing digital data (data) that has been provided by the data holder and which may be subject to specific technical, legal, organisational requirements. “**Open data**” is used to refer to data that can be accessed and re-used by anyone without technical, legal or organisational restrictions.

“**Data sharing**” refers to the provision of data by the data holder, on a voluntary basis. It includes the re-use of data based on commercial and non-commercial conditional data-sharing agreements, as well as open data.¹

“**Enhanced access and sharing**” refers to mechanisms and approaches aimed at maximising the social and economic benefits from the wider and more effective use of data, while, at the same time, addressing related risks and challenges. The term does not cover cases where governments access private-sector data either for law enforcement and national security purposes or for granting regulatory approval (e.g. for the marketing of pharmaceutical or agricultural chemical products).

1. Data sharing assumes common interests between the entities agreeing to share their data, including the interest and expectation that data holders can become data users and vice versa. Data sharing therefore can come with an expectation of some kind of reciprocity among the stakeholders engaged in data-sharing agreements.

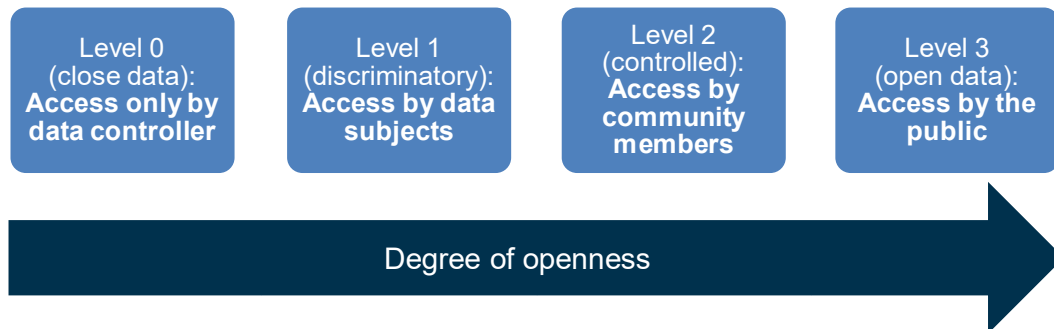
Different approaches can be employed to enhance data access, sharing and re-use along the data openness continuum. As it is the case with data openness, enhanced access and sharing should not be considered a binary concept opposing closed to open data, but rather a continuum of different degrees of data openness, ranging from internal access and re-use (only by the data holder), restricted (unilateral and multilateral) external access and sharing, to open data, the most extreme form of data openness (Figure 2.1).

The following sections highlight the different contextual factors that may affect data governance that need to be considered for a more effective management of the risks of data access and sharing. These sections focus on i) the different types of data and access control mechanisms; ii) the main types of actors and their roles; and iii) the different approaches to EASD and their relevant types of interactions.

There is no single, silver-bullet solution to the challenges raised by data access and sharing. The most appropriate approach will typically depend on the different types of data and the risks associated with their re-use, and the different actors and their roles. Data that is effectively anonymised and aggregated will, in principle, be shared more openly as it is less likely to lead to privacy violations. Application programming interfaces (APIs) and

data sandboxes can be promising mechanisms through which data access and (re-)use can be controlled through time (see subsection “Data-access control mechanisms: Protecting the interests of data holders”).

Figure 2.1. The degrees of data openness and access



Source: OECD (2015^[2]), *Data-Driven Innovation: Big Data for Growth and Well-Being*, <http://dx.doi.org/10.1787/9789264229358-en>.

Different types of data and access control mechanisms

A few misconceptions, and lack of clarity on key concepts and terms continues to clutter the policy debate around data and data sharing and re-use. Data are often treated as a monolithic entity, although evidence shows that they are heterogeneous goods whose value depends on the context of their use, with different implications for individuals, businesses and policy makers (OECD, 2015^[2]; OECD, 2013^[3]).¹ From a privacy perspective, personal data, for instance, typically requires more restrictive access regimes than non-personal public-sector data. On the other hand, industrial data will in most cases be proprietary data and therefore data access and sharing may have to be more restrictive compared to public-sector data, which in many cases can be shared through open data.

Different taxonomies have been developed to help identify and address data-governance implications raised by the different types of data and to address the lack of definitions, which remains a source of confusion. As Taylor (2013^[4]) highlights in reference to personal data: “A new taxonomy of data is badly needed. Industry, government and citizens are too frequently in disagreement as to what exactly constitutes personal data and what does not – and without an understanding of how data get positioned in each category, or flow between them, it is impossible to have a discussion about how to govern and regulate those flows.” Taylor (2013^[4]) refers to the distinction between personal and non-personal data, which is also the most relevant categorisation given the legal and regulatory implications that comes with the collection, processing, sharing and (re-)use of personal data.

The following sections present and further discuss four major dimensions that are considered critical for a data taxonomy to be relevant for the governance of data access and sharing. These include: i) **personal data and the degrees of identifiability**, given that a higher degree of identifiability would typically be associated with higher risks and therefore would require more restrictive data-access control; ii) **the domain of the data**, which describes whether data are personal, private or public, and thus the legal and regulatory regime applicable to the data; and iii) **the manner data originates**, which reflects the level of awareness and control that data subjects can have about data collected about or from them. Finally, iv) among possible access control mechanisms, **downloads, APIs, and data sandboxes** are discussed,

the latter two of which can be considered to enhance access to data while also protecting the interests and rights of individuals and organisations.

Combined, the four dimensions further discussed below can help address data governance in a more differentiated manner. For example, in certain conditions access to highly sensitive (identified) personal data could be granted, but only within a restricted digital and/or physical environment (data sandboxes) to trusted users (see subsection “Data sandboxes for trusted access and re-use of sensitive and proprietary data” below). If sufficiently anonymised and aggregated, that data could however also be provided to the public via e.g. APIs that would in addition help reduce the level of risk of re-identification. As another example, what data should be made accessible to consumers or business customers may depend on the manner that data originated and whether the data are considered personal and/or proprietary (see subsection “The manner data originates: Reflecting the contribution to data creation” below). Where data are directly provided by a user to a service provider, e.g. when they explicitly share information about themselves or others, expectations in general are that the user should be able to access that same data. Expectation may be different and eventually diverge however, where data about the user has been created by the service provider, for instance, through data analytic inferences.

Personal data and the degrees of identifiability: Reflecting the risk of harm

Most privacy regulatory frameworks as well as the OECD (OECD, 2013^[5]) *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (hereafter the “OECD Privacy Guidelines”) include a definition of personal data. The OECD Privacy Guidelines define it as “any information relating to an identified or identifiable individual (data subject)”. Once data are classified as personal data, access to and sharing of this data become predominantly governed by the applicable privacy regulatory framework. This remains true irrespective of the sector of data collection, processing and (re-)use, even if different privacy regulatory frameworks may apply across these sectors.

However, the binary nature of this dichotomy (personal vs non-personal data) has been criticised for two reasons:

1. **The dynamic nature of personal data:** Current developments in data analytics (and artificial intelligence [AI]) have made it easier to link and relate seemingly non-personal data to an identified or identifiable individual (Narayanan and Shmatikov, 2006^[6]; Ohm, 2009^[7]). This blurs the distinction between personal and non-personal data (OECD, 2011^[8]) and challenges any regulatory approach that determines the applicability of rights, restrictions and obligations on the sole basis of a static concept of “personal data”.
2. **Personal data itself encompasses many different types of data** that in some contexts deserve to be distinguished and addressed differently, given the different level of risks associated with their collection, processing and use. This is reflected in some privacy regulatory frameworks such as the GDPR (Art. 9), which provides elevated protection for certain categories of personal data, often considered sensitive, by expressly prohibiting their processing (unless certain conditions are met).²

More detailed personal data taxonomies have been introduced to further differentiate between the different categories of personal data.³ Some have been recognised as standards, such as ISO/IEC 19941 (2017^[9]), which has been developed to achieve interoperability and portability in cloud computing services.⁴ Most importantly, ISO/IEC 19441 distinguishes between five categories or states of data identifiability, which include (in reverse order to the degree identifiability):⁵

- **Identified data:** Data that can be unambiguously associated with a specific person because personal identifiable information is observable in the information.
- **Pseudonymised data:** Data for which all identifiers are substituted by aliases for which the alias assignment is such that it cannot be reversed by reasonable efforts of anyone other than the party that performed them.
- **Unlinked pseudonymised data:** Data for which all identifiers are erased or substituted by aliases for which the assignment function is erased or irreversible, such that the linkage cannot be re-established by reasonable efforts of anyone including the party that performed them.
- **Anonymised data:** Data that is unlinked and which attributes are altered (e.g. attributes' values are randomised or generalised) in such a way that there is a reasonable level of confidence that a person cannot be identified, directly or indirectly, by the data alone or in combination with other data.
- **Aggregated data:** Statistical data that does not contain individual-level entries and is combined from information about enough different persons that individual-level attributes are not identifiable.

ISO/IEC 19441 is relevant for data access and sharing as it reflects the degree to which data can refer back to an identity (including an individual or an organisation). It can therefore help assess the level of risk to privacy, which in turn can help determine the degree to which legal and technical protection may be necessary, including the level of access control required (see subsection “Data-access control mechanisms: Protecting the interests of data holders” below). The less data that can be linked to an identity, because it may be effectively anonymised and sufficiently aggregated, the lower the risks to privacy and thus the more openly can the data be made available.⁶ The actual level of data openness will therefore depend on the potential impact that data re-use will have on privacy, or more generally on confidentiality.

The overlapping domains of data: Reflecting the various stakeholder interests

Besides the dichotomy between personal and non-personal data discussed above, the most frequently made distinction is between private sector and public-sector data. It is generally accepted and expected that public-sector data should be made available through open data, free of charge and free of any restrictions from intellectual property rights (IPRs) – where there are no conflicting national security or private interests. This is reflected in a number of open data initiatives such as data.gov (United States), data.gov.uk (United Kingdom), data.gov.fr (France) and data.go.jp (Japan) (see subsection “Open data” below).

According to the OECD (2008^[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* (OECD PSI Recommendation),⁷ public-sector information (PSI) is defined as information generated, created, collected, processed, preserved, maintained, disseminated or funded by or for the government or public institutions.⁸ In analogy, private-sector data can be defined as data that is generated, created, collected, processed, preserved, maintained, disseminated or funded by or for private sector, which comprises “private corporations, households and non-profit institutions serving households” according to the OECD *Glossary of Statistical Terms* (OECD, 2001^[11]).

The private-public data distinction as defined above raises three major issues, which are rarely fully acknowledged:

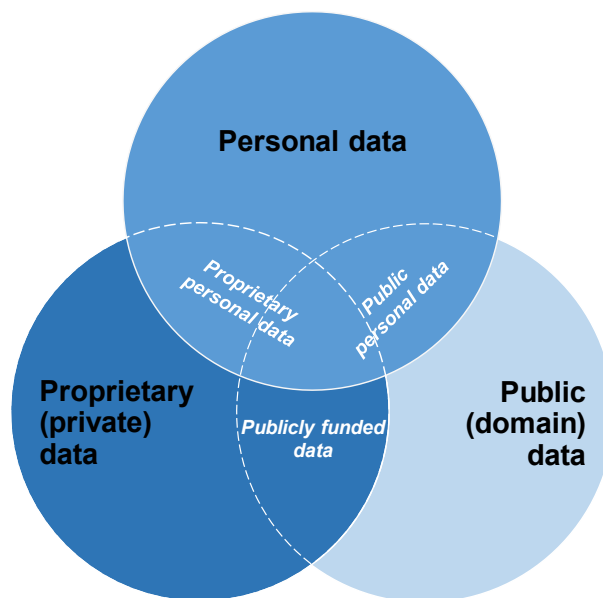
1. **Public-sector and private-sector data cannot always be distinguished:** Data can often qualify as being both public sector *and* private-sector data, and this irrespective of any joint activities between public and private-sector entities (e.g. public-private partnerships [PPPs]). For instance, data generated, created, collected, processed, preserved, maintained or disseminated by the private sector, that are funded by or for the public sector would classify as public sector *and* private-sector data.⁹ As a result, some of the existing presumptions about public-sector data (e.g. that they should always be made available through open data, free of charge and free of any IPR restrictions) have to be questioned.¹⁰
2. **Public-sector data and private-sector data are often mistakenly used as synonyms for public (domain)¹¹ data and private (proprietary) data, respectively:** However, data produced and controlled by the public sector is usually proprietary data at first, before being put in the public (domain) thanks to open data initiatives (see subsection “Open data”). Similarly, even if most data produced and controlled by the private sector can be considered proprietary (private) data, some data in the private sector may remain in the public domain, for instance if they are open data.
3. **The distinction (private sector vs. public-sector data) does not fully reflect the data of households, most of which is personal data.** Even though the private sector includes households (besides private corporations and non-profit institutions), private-sector data are too often assumed to only include data of private sector institutions. Most importantly, household data are to a large extent personal data. However, traditionally the distinction between private sector and public-sector data is rarely put in relation to personal data (and the different degrees of identifiability).

For data-governance frameworks to be applicable across society, it seems thus crucial to also distinguish between the following three domains of data (Figure 2.2):

- the **personal domain**, which covers all personal data “relating to an identified or identifiable individual” for which data subjects have privacy interests,
- the **private domain**, which covers all proprietary data that are typically protected by IPRs (including copyright and trade secrets) or by other access and control rights (provided by e.g. contract and cyber-criminal law), and for which there is typically an economic interest to exclude others, and
- the **public domain**, which covers all data that are not protected by IPRs or any other rights with similar effects, and therefore lie in the “public domain” (understood more broadly than to be free from copyright protection), thus free to access and re-use.
- These three domains not only overlap as illustrated in Figure 2.2, but they are also typically subject to different data-governance frameworks that can affect each domain differently. For instance, privacy regulatory frameworks typically govern the personal domain, while the private domain is typically governed by frameworks governing property rights, and most prominently IPRs. These overlaps may partly explain why data governance is often perceived as complex. They may also explain the potential conflicting views and interests of some stakeholder groups, as reflected for instance in issues related to “data ownership” (Chapter 4) and “data of public interests” (Chapter 5).
- Furthermore, depending on the jurisdiction, some domains may be prioritised over others. This is, for example, reflected in current data portability rights, which vary significantly across government initiatives (see subsections “Data portability” in

this chapter and in Chapter 5). These initiatives can be seen as attempts to address the conflicting interests of individuals and organisations over their “proprietary personal data” (Figure 2.2). By enabling individuals to re-use their personal data, individuals gain more control rights over their personal data. However, what type of data falls within the scope of data portability varies significantly across initiatives, reflecting the (implicit) priorities of personal domain vs. the proprietary domain (see follow-up subsection).

Figure 2.2. The personal, private and public domains of data



The manner data originates: Reflecting the contribution to data creation

At the Copenhagen Expert Workshop, some experts highlighted that multiple stakeholders were often involved in the contribution, collection and control of data, including the data subject in the case of personal data. However, the data categories discussed above – in particular the distinction between the personal domain and the proprietary domain – do not help differentiate how different stakeholders contribute to data co-creation. Data categories that differentiate according to the way data are collected or created (Schneier, 2009^[12]; WEF, 2014^[13]; Abrams, 2014^[14]; OECD, 2014^[15]) can provide further clarity in this respect. These categorisations are motivated by the recognition that the way data are generated and collected determines the level of awareness that data subjects can have about the privacy impact of the data collection and process.¹²

Abrams (2014^[14]), for instance, notes that “legacy privacy governance regimes are based on a presumption that data are primarily being collected from the individual with some level of their awareness”.¹³ However, increasingly, data are collected from individuals without their awareness of the actions that may lead to data origination. Abrams distinguishes between four categories of data: i) provided; ii) observed; iii) derived; and iv) inferred data (Table 2.1).

This distinction already plays an important role in policy making and regulation;¹⁴ for instance, in the interpretation of the right to data portability under the GDPR (Art. 20), which specifically regulates personal data provided by the data subject to a data controller.¹⁵ According to the

Article 29 Working Party (Art. 29 WP) the right to data portability under the GDPR would include volunteered as well as observed data. It would however exclude data derived (inferred) from additional processing that are often considered proprietary.

Table 2.1. Data categories based on origin

Category	Sub-Category	Examples	Level of individual awareness
Provided	Initiated	Applications Registrations Public records (e.g. filings and licenses) Credit card purchases	High
	Transactional	Bills paid Inquiries and surveys responded to Public records (e.g. health, schools and courts)	High
	Posted	Speeches in public settings Social network postings Photo and video services	High
Observed	Engaged	Cookies on a website Loyalty card	Medium
	Not anticipated	Activated location sensors on personal devices Data from sensor technology in my car Time paused over a pixel on the screen	Low
	Passive	Facial images from CCTV Obscured web technologies Wi-Fi readers in buildings that establish location	Low
Derived	Computational	Credit ratios Average purchase per visit	Medium to Low
	Notational	Classification based on common attributes	Medium to Low
Inferred	Statistical	Credit and fraud scores Response score	Low
	Advanced analytical	Risk of developing a disease-based multifactor analysis College success score based on multivariable analysis	Low

Source: OECD based on Abrams (2014^[14]), *The Origins of Personal Data and its Implications for Governance*, <http://dx.doi.org/10.2139/ssrn.2510927>.

The Australian Productivity Commission (2017^[16]) distinguishes between i) data posted online by the consumer;¹⁶ ii) data created from online transactions; iii) data purchased; and iv) “other data associated with transactions or activity that are held in digital form”. This categorisation was used for Australia’s Consumer Data Right (CDR), which gives consumers the right to safely access certain data about them held by businesses. Depending on the industry, the type of data made available to consumers may vary quite significantly. In the case of the banking sector, for instance, not only volunteered data will be made available, but also data on financial products including credit and debit card, deposit and transaction accounts, and data on mortgages.

This report combines the various categories used by Abrams (2014^[14]) and the Productivity Commission (2017^[16]) as follows:

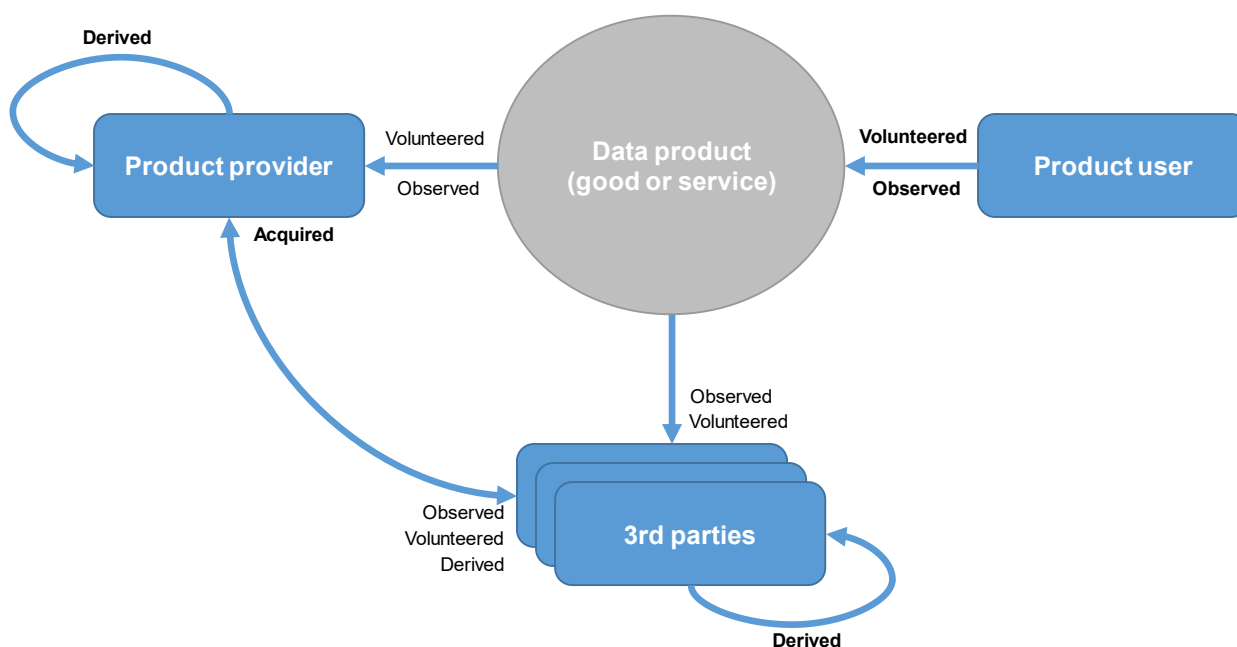
- **Volunteered (or surrendered or contributed or provided) data** are data provided by individuals when they explicitly share information about themselves or others. Examples include creating a social network profile and entering credit card information for online purchases.
- **Observed data** are created where activities are captured and recorded. In contrast to volunteered data where the data subject is actively and purposefully sharing its data, the role of the data subject in the case of observed data is passive. The data

controller plays the active role. Examples of observed data include location data of cellular mobile phones and data on web usage behaviour.

- **Derived (or inferred or imputed) data** are created based on data analytics, including data “created in a fairly ‘mechanical’ fashion using simple reasoning and basic mathematics to detect patterns” (OECD, 2014^[15]). In this case, it is the data processor that plays the active role. The data subject typically has little awareness over what is inferred about her or him. Examples of derived data include credit scores calculated based on an individual’s financial history. It is interesting to note that personal information can be derived from several pieces of seemingly anonymous or non-personal data (Narayanan and Shmatikov, 2006^[6]).
- **Acquired (purchased or licenced) data** are obtained from third parties based on commercial licensing contracts (e.g. when data are acquired from data brokers) or other non-commercial means (e.g. when data are acquired via open government initiatives). As a result, contractual and other legal obligations may affect the re-use and sharing of data.

This categorisation can help assess the extent to which different stakeholders are involved in the creation of data, including cases where users (consumers and businesses) interact with a data product (good or service) such as an e-government service, a social networking service, or a portable smart-health device. These products typically i) observe the activities of their users, in which case observed data are created; and/or ii) are used to input data volunteered by their users (volunteered data). The data can then be accessed for further processing (and the creation of derived data) by the product provider and by any third parties who may have been granted direct or indirect access to the original (volunteered and observed) data – or in a less identifiable form. The creation of derived data can also be enriched when combined with acquired data from (other) third parties (Figure 2.3).

Figure 2.3. Data products and the different manners data originates



Notes: Arrows represent potential data flows between the different actors and a data product (good or service). The type of data is highlighted in bold to indicate the moment at which the data are created.

Data-access control mechanisms: Protecting the interests of data holders

There are a wide range of different mechanisms for accessing data. The most commonly used is data access via i) (ad-hoc) downloads; and via ii) application programming interfaces (APIs). Additionally, iii) data sandboxes are increasingly recognised as means to access sensitive and proprietary data, while assuring the privacy rights and IPRs of right holders. These access control mechanisms are discussed further below.¹⁷

(Ad-hoc) downloads

In the case of data access via downloads, the data are stored, ideally in a commonly used format and made available online (e.g. via a web site). Many open data portals are still predominantly based on data downloads as highlighted during the Stockholm Open Government Workshop (see also Ubaldi (2013_[17])).

Data access via downloads however raises several issues. Interoperability is a major issue for data re-use across applications (data portability). Even when commonly used machine-readable formats are employed, interoperability is not guaranteed. These formats may enable data syntactic portability, i.e. the transfer of “data from a source system to a target system using data formats that can be decoded on the target system”. But they do not guarantee data semantic interoperability, “defined as transferring data to a target such that the meaning of the data model is understood within the context of a subject area by the target”. In addition to common machine-readable data format, data semantic portability requires mutually understood ontologies and metadata such as the W3C Web Ontology Language (OWL)¹⁸ and the Dublin Core Schema¹⁹ to assure a common meaning of the data.

Furthermore, data access via (ad-hoc) downloads can increase digital security and privacy risks since data once downloaded are outside the information system of the data holder and thus out of his/her control. Data holders would thus lose their capabilities to enforce any data policies including those meant to protect the privacy of data subjects and the IPRs of right holders.

Application programming interfaces

As applications increasingly rely on data, accessing data without human intervention becomes essential. APIs enable service providers to make their digital resources (e.g. data and software) available over the Internet. APIs thus enable the smooth interoperability of the different actors and their technologies and services, particularly through the use of cloud computing.

In the case of Transport for London (TfL), a local government body responsible for the transport system in Greater London (United Kingdom), a unified API was used as common gateway for TfL data: TfL powered their own website with the same data and the same API was used to give third-party developers access to TfL data. The infrastructure was built to allow new data sets to be easily included from different systems and sources and to be updated efficiently. By using the cloud, the infrastructure was also scalable.

A key advantage of an API (compared to an ad-hoc data download) is that an API enables a software application (or app) to directly use the data it needs. Data holders can also implement several restrictions via APIs to better control the use of their data including means to assure data syntactic and synthetic portability. Furthermore, they can control the identity of the API user, the scale and scope of the data used (including over time), and even the extent to which the information derived from the data could reveal sensitive/personal information. Box 2.2 briefly describes how Twitter uses APIs to control the re-use of its data by third-party app developers.

Box 2.2. Twitter's data control over its application programming interface

Twitter's API allows outside developers to build apps that can pull in information directly from Twitter to display in their own apps. The availability and openness of proprietary APIs have been instrumental for the rapid expansion of apps and the growth of platforms such as Twitter.

Twitter has been pursuing a vertical integration strategy by acquiring and building a portfolio of apps. The company purchased apps such as TweetDeck (2011), Tweetie (2010) and Summize (2008) intending to later transform them into brand extensions that serve different platforms and services, e.g. search engines.

Since then, Twitter has been discouraging developers from using their APIs to make apps that compete directly with their platform by rejecting apps that rely on tweet feeds via its API and by revoking API access. In 2012, Twitter restricted the number of individual app tokens that could access their APIs to 100 000. This essentially means that app developers were limited to 100 000 app installs without special permission from Twitter to increase the number.

Source: OECD (2015^[2]), *Data-Driven Innovation: Big Data for Growth and Well-Being*, <http://dx.doi.org/10.1787/9789264229358-en>.

Data sandboxes for trusted access and re-use of sensitive and proprietary data

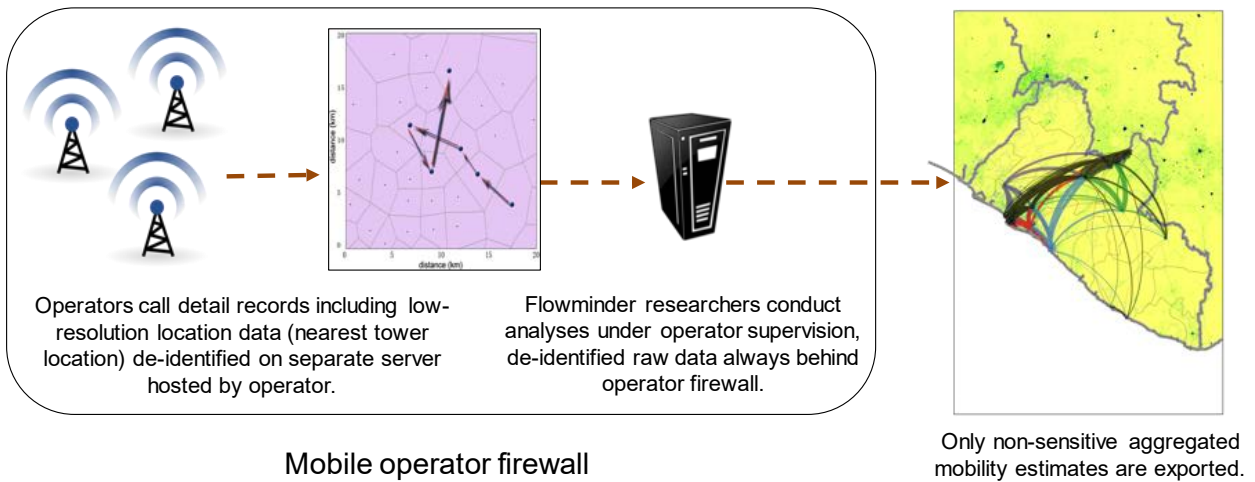
The term “data sandbox” is used in this paper to describe any isolated environment, through which data are accessed and analysed, and analytic results are only exported, if at all, when they are non-sensitive. These sandboxes can be realised through technical means (e.g. isolated virtual machines that cannot be connected to an external network) and/or through physical on-site presence within the facilities of the data holder (where the data are located). Data sandboxes would typically require that the analytical code is executed at the same physical location as where the data are stored.

Compared to the other data access mechanisms presented above, data sandboxes offer the strongest level of control. Data sandboxes are therefore promising for providing access to very sensitive/personal and proprietary data. Examples include:

- The Centers for Medicare and Medicaid (CMS) Virtual Research Data Center (VRDC), a virtual research environment that provides timely access to Medicare and Medicaid programme data (such as beneficiary-level protected-health information) (ResDAC, n.d.^[18]). Researchers working in the CMS VRDC have direct access to approved data files and can conduct their analysis within the CMS secure environment. They can download aggregated reports and results to their own personal workstation. Researchers can in addition upload external data files into their workspace to link and analyse their data with the approved CMS data files. Access is provided over a virtual private network (VPN) and a virtual desktop to satisfy all CMS privacy and security requirements.
- Flowminder.org, an initiative combining new types of data to support people in low- and middle-income countries. It relies on a secured access to personal data of mobile operators' call detail records including de-identified low-resolution location data (on nearest tower location) (Figure 2.4). To assure the privacy of their users mobile operators host separate dedicated servers behind their firewalls, on top of which Flowminder.org researchers conduct their analyses. The data always resides within the operators' servers; only non-sensitive aggregated estimates are exported

outside the servers. This configuration allows sensitive data to remain under the operators' control, minimising privacy, security and commercial concerns.

Figure 2.4. Flowminder.org's data processing flow



Source: Slides presented at the Copenhagen Expert Workshop by Erik Wetter (Professor, Stockholm School of Economics; Chairman, Flowminder.org).

Main types of actors and their roles

The complexity of many ecosystems in which data are shared and re-used is determined to a significant extent by different communities of actors. The citizen science²⁰ project Galaxy Zoo, a community²¹ of over 850 000 people who have taken part in more than 20 citizen science projects over the years, provides a good illustration. Madison (2014_[19]) observes that the most important reasons for the effectiveness of Galaxy Zoo “have less to do with the character of its information resources (scientific data) and rules regarding their usage, and more to do with the expanded community constructed from hundreds of thousands of Galaxy Zoo volunteers”.

These communities can be very heterogeneous, encompassing actors with various, partly opposing, interests and (market) power. The community of Galaxy Zoo, for example, includes three types of actors, each level having different rights with different implications on data governance.²² Madison (2014_[19]) shows that this complex and heterogeneous membership structure was key to the success of the Galaxy Zoo project, where the more closed leadership team of expert astronomers guided the open community of volunteers, thereby assuring the alignment of incentives and quality control.

The following subsections discuss in more detail the role of i) data holders and controllers as potential data providers; ii) data users; and iii) data intermediaries, including data repositories and data brokers, data marketplaces, personal information management systems (PIMSs) or personal data stores (PDSs), and trusted third parties, some of which fulfil multiple intermediary roles.

Data subjects should also be recognised as a key actor where personal data are shared and re-used, in which case they act as data provider. As highlighted further below in the case of data portability, the role of data subjects can be leveraged, where enhanced access and sharing is used to give the data subjects greater control over their data and to help them achieve “informational self-determination”. In this case, data subjects can act as both data users and data holders at the same time.

Data holders and controllers

In alignment with the definition of “data controller” provided by the OECD Privacy Guidelines (OECD, 2013^[5]), this reports defines a “data holder” as a party who, according to domestic law, is competent to decide about the contents and use of (personal and non-personal) data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf.²³ They are sometimes considered “data owners”, even though they may not have any legal ownership rights over the data they control. For this reason, they are sometimes called “data stewards”.

Data holders are not limited to the private sector. The public sector is one of the economy’s most data-intensive sectors (OECD, 2015^[2]). In the United States, for example, public-sector agencies stored on average 1.3 petabytes (PB) of data in 2011, making them the country’s fifth most data-intensive sector (OECD, 2015^[2]). This makes the public sector a major potential source of data for the rest of the economy and has motivated open data initiatives such as data.gov (United States), data.gov.uk (United Kingdom), data.gov.fr (France) and data.go.jp (Japan) (see subsection “Open data” below).²⁴

Data holders are also not limited to organisations. An individual (data subject) in control of his or her personal data, and with the means to share the data with other stakeholders, should also be considered a data holder. Increasingly, data intermediaries, including PIMS and PDS, are providing individuals with the necessary means to *actively* contribute their data to the data ecosystem (see subsection “Data intermediaries” below). Data portability initiatives have significantly contributed to this trend (see subsection “Data portability” below).

Data holders are among the most critical actors for data sharing and re-using because without their active contributions there would be no data available. Therefore, properly aligned incentive mechanisms that target data holders without discouraging their data-related investments are crucial for a well-functioning data-sharing ecosystem. The effectiveness of incentive mechanisms will depend on the extent to which data holders can benefit from data sharing and be protected from risks. The availability of sustainable business models, IPR and privacy protection, and mediation through trusted intermediaries are among the most crucial factors for incentivising and facilitating data sharing across society, but challenges remain (Chapter 4).

Data users

Data users are responsible for generating the social and economic value of data sharing. As is the case with data holders, data users may be very diverse. They may include i) consumers, who directly access data about them that are controlled by businesses; ii) citizens, who access public-sector data made available by governments via their open data initiatives; iii) researchers that access scientific data made available via open science project; and iv) businesses that access data provided through e.g. data partnerships, open data or data portability initiatives. In some cases, types of data users may overlap. For instance, public-sector data are made available via open data to citizens for overseeing public-sector activities (transparency), scientists for research purposes and businesses for the creation of new commercial opportunities (Chapter 3).

Engaging with data users can be in the genuine interest of data holders willing to share their data. Making data machine-readable and downloadable is necessary, but not sufficient, to engage the community of users, including developers. As experts at the Copenhagen Workshop demonstrated, it is critical to understand and communicate how data are creating value to better engage with the community of users. In the case of TfL, for example, data-sharing projects are not viewed as technology projects but as ways for engaging customers in the very early phases of project design and for keeping them involved.

Data intermediaries

Data intermediaries enable data holders to share their data, so it can be re-used by potential data users. They may also provide additional added-value services such as data processing services, payment and clearing services and legal services, including the provision of standard-lifecycle schemes.

Anecdotal evidence presented at the Copenhagen Expert Workshop suggested that many end users, consumers and businesses included, often do not use the data available to them. They rather rely on intermediaries that access the data to provide the embedded information in more user-friendly ways, sometimes enhanced in terms of data quality and enriched through additional, often inferred, data. These intermediaries typically provide added-value services including advanced data analytic services. The Copenhagen Expert Workshop also revealed that while businesses tend to use data brokers as main data intermediaries, consumers often have access to added-value information services via apps. Overall, this has led to new demand for added-value services and thus to new business opportunities for new and old intermediaries including data brokers, app developers, but also for some incumbents in information and communication technology (ICT) and non-ICT industries (e.g. telecommunication and financial services firms).

There are a wide variety of types of data intermediaries. The most popular types are data repositories and data brokers. While data brokers are primarily active in the commercial space and are profit-based (OECD, 2015^[2]), data repositories are primarily serving public good objectives, for instance by acting as (part of) a national library, and/or by serving the scientific community as data storage, processing and/or sharing infrastructure (OECD, 2017^[20]). Recent years have also seen the emergence of “data marketplaces” – online platforms that host data from various publishers and offer the (possibly enriched) data to interested parties (Dumbill, 2012^[21]), and PIMS/PDS, which serve consumers in managing the sharing of their personal data. The following subsection discusses these in more detail, while the subsequent ones focus on data intermediaries that also act primarily as trusted third parties to facilitate data sharing.

(Research) data repositories

Data repositories, which sometimes are also referred to as data libraries or data archives, preserve data as a resource of knowledge for society. They thus fulfil the same function as traditional libraries and archives. Data repositories are particularly relevant for science and research.²⁵ The OECD (2017^[20]) shows that research data repositories have become more important as data preservation and open data policies are increasingly widespread and influential, especially in the context of open science. Citing Beagrie and Houghton (2012^[22]; 2014^[23]; 2013^[24]; 2013^[25]; 2016^[26]), which look, respectively, at the social and economic impacts of the Economic and Social Data Service, the Archaeology Data Service, the British Atmospheric Data Centre, and the European Bioinformatics Institute, the report highlights two major types of benefits enabled by data repositories:

First, there are substantial and positive efficiency impacts, not only reducing the cost of conducting research, but also enabling more research to be done, to the benefit of researchers, research organisations, their funders, and society more widely. Second, there is substantial additional re-use of the stored data, with between 44% and 58% of surveyed users across the studies saying they could neither have created the data for themselves nor obtained it elsewhere (OECD, 2017, p. 16^[20]).

Data brokers and data marketplaces

The core business objective of data brokers is to collect and aggregate data, including personal data (Federal Trade Commission (US), 2014^[27]). Data brokers such as Bloomberg, Nielsen, STATS (sports data) and World Weather Online tap into a variety of data sources that are used for data-related services. These include, for example, data that are disclosed or provided by individual firms and citizens; data from firms that install sensors; data “crawled” from the Internet; and data from non-profit and public-sector agencies (e.g. earth observation data and demographic, health and other statistics). Some data brokers also analyse their data sets to provide information and intelligence services to their clients in a wide range of domains for a variety of purposes, including verifying an individual’s identity, product marketing, and fraud detection.

One important distinguishing factor between data brokers and data market providers is that data brokers are actively engaged in the collection of additional data and their aggregation, while data market providers are passive intermediaries through which data controllers, including brokers, can offer their data sets. The best-established data markets are provided by Infochimp, DataMarket, and Factual, although there are several more (OECD, 2015^[2]). Some data markets try to offer all the data they can, such as Infochimp. Others focus on specific kinds of data, such as Factual, which originally started with location data and has since branched out to offer other specific verticals, including services to i) analyse customer intent and movements; ii) build highly customisable, scalable and accurate audience segments; iii) measure the impact of marketing campaigns; and iv) develop software engines for mobile marketers and app developers. Another type of specialisation is to choose a specific target group, such as Figshare, a data market for researchers.

However, despite the growth of data intermediaries, there is no single data marketplace where organisations and individuals can sell or exchange data directly with each other. Thus a variety of business models may be viable. A few platforms provide services tailored to specific, tightly integrated value chains that are heavily dependent on each other. However, many fail to scale. Microsoft’s DataMarket and Data Services solutions, for instance, were integrated in Microsoft’s cloud computing platform (Microsoft Azure). The uptake of both services has been, however, not as expected, forcing Microsoft to discontinue both services as of March 2017. The inability to scale up Microsoft’s DataMarket and Data Services could be due to Microsoft’s business model decision to bundle both (DataMarket and Data Services) to its cloud service offering.²⁶

Personal information management systems and personal data stores

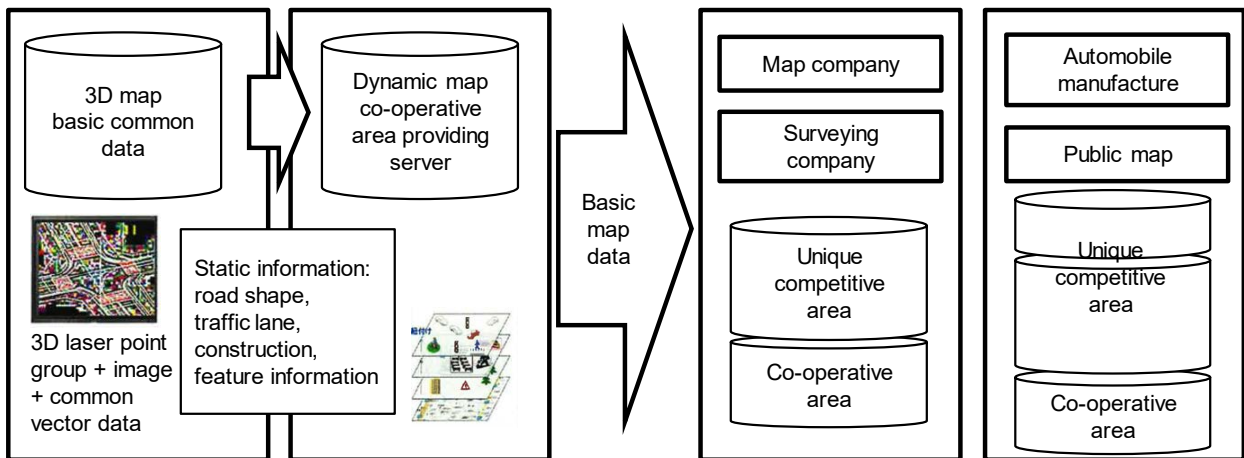
While most data intermediaries target business-to-business (B2B) data sharing and re-use, PIMS/PDS are emerging as promising platforms to give data subjects (consumers) more control over their personal data and thus to restore user agency, including in the context of the Internet of Things (Urquhart, Sailaja and Mcauley, 2017^[28]). The concept of PIMS/PDS therefore is receiving a lot of attention as a driver for data portability, as they can function as a centralised data infrastructure allowing individuals to manage their personal data. By assessing and confirming the reliability and trustworthiness of data users, PIMS/PDS can increase trust in data re-use and function as an “Information Trust Bank”. An example of a PIMS/PDS application in tourism is Omotenashi, an app in Japan that can collect existing personal information from social network services, which could be shared with local businesses (provided user consent is given). The app gives recommendations for places and businesses to visit based on the data the user consents to include.

Trusted third parties

Trusted third parties can act as a facilitator of data sharing and re-use among all stakeholders. Several models of trusted third parties or platforms were presented at the Copenhagen Expert Workshop.

- **Data intermediary acting as a third-party certification authority:** In some cases, data intermediaries can act as a certification authority as in the case of the Industrial Data Space (IDS). The certification authorities of the IDS certifies all participants based on standards defined by the IDS regarding, for example, security, privacy, and terms of use. Data owners define terms of use and the fees of data use, which data brokers use to match with other data owners and users.
- **Private sector designation of a trusted data-sharing platform:** In many, if not most, cases presented at the Copenhagen Expert Workshop, major data providers would come together and either designate an existing trusted organisation or create a new trusted organisation and platform. Examples include the *Ship Big Data Platform*, which was established by the Japanese Sea Association and combines and provides different data sets from weather data to ship data, including information of ship ownership, operators and trajectory. Another example is the High Definition 3D Map Data Platform. Founded by 15 Japanese companies, it provides a platform for co-operative research and development in AI-enabled driving technologies (Figure 2.5). Alternative models involve existing trusted organisations that engage with potential data providers, who in turn agree to collaborate by sharing their data. An example is the Health Care Cost Institute (HCCI), a non-profit organisation that provides information about health care utilisation and costs in the United States based on data contributed by health care and health insurance companies (e.g. Aetna, Humana, Kaiser Permanente, and United Healthcare). The data provided to the HCCI are shared with selected researcher institutions after the information about the data providers (the providing health care and health insurance companies) has been removed.
- **Public-sector designation of a trusted data-sharing platform:** In some cases, governments can act as or create a trusted third party. The Australian government initiated the Data Integration Partnership for Australia, “an investment to maximise the use and value of the Government’s data assets” (Department of the Prime Minister and Cabinet (Australia), 2017_[29]). While agencies in social services, health, education, finance and other government agencies would provide data for linking and integration, “sectoral hubs of expertise, independent entities that are funded by the Commonwealth” and denominated Accredited Integrating Authorities, would enable the integration of longitudinal data assets – “housed in a secure environment, using privacy preserving linking methods and best practice statistics to link social policy and business data” (Productivity Commission, 2017_[16])²⁷ (see subsection “Voluntary and collaborative approaches” in Chapter 5 for more examples).
- **Independent ethics review bodies (ERBs)** have been highlighted in some cases as critical trusted-third parties, in particular where access to and sharing and re-use of personal data are concerned. In some cases, in the scientific community, the evaluation of applications for access to publicly funded personal data for research purposes can depend on the existence of ERBs. A Global Science Forum Expert Group recently concluded that ERBs can increase trust between parties with an interest in the use of personal data for research purposes, particularly in situations where consent for research use is impractical or impossible (OECD, 2016_[30]).²⁸

Figure 2.5. High Definition 3D Map Data Platform



Source: Dynamic Map Platform Co., Ltd, slide presented at the Copenhagen Expert Workshop by Naoto Ikegai (Interfaculty Initiative in Information Studies, University of Tokyo).

Approaches to access and sharing and their degree of openness

Three approaches to enhanced access and sharing have been most prominently discussed in the literature and by policy makers: open data, and more recently data markets and data portability. Besides these three, a wide range of other approaches exist, with different degrees of data openness, responding to the various interests of stakeholders and the risks they face in data sharing, such as bilateral or multilateral data partnerships. Many of these approaches are based on voluntary and mutually agreed terms between organisations. Others are mandatory, such as the European Union's GDPR right to data portability (Art. 20) (European Union, 2016^[1]), or Australia's recently passed CDR (see Chapter 5 for more examples).

The following subsections describe in more detail the different approaches to data access and sharing. The first subsection discusses the crucial role of bilateral contractual agreements and how this basic approach to data monetarisation can be leveraged through data markets. The two successive subsections then focus on open data as the most extreme form of data openness, and data portability as the most restrictive form of data access, as it typically restricts access to those that were involved in the creation and collection of the data – this is for instance the case with data subjects and their personal data. Finally other, less restrictive data-sharing arrangements such as data partnerships and the so-called “data for social good” initiatives are discussed.

Contractual agreements and data markets

Increasingly, businesses are recognising the opportunities of commercialising their proprietary data (OECD, 2015^[2]). While some organisations offer their data for free (via open access), especially non-governmental organisations and governments as highlighted below, many businesses have already engaged bilateral arrangements to sell or licence their data. For example, the French mobile Internet service provider Orange acts as a data provider by using its Floating Mobile Data technology to collect mobile telephone traffic data, which determine speeds and traffic density at a given point in the road network. The anonymised mobile telephone traffic data are sold to third parties to identify “hot spots” for public interventions or to provide traffic information services.

Box 2.3. The diversity of revenue models in the data ecosystem

Businesses in the data ecosystem use a diversity of revenue models. The most common are:

- **Freemium:** The term “freemium” is a portmanteau of “free” and “premium”. This revenue model is one of the most frequently used: products are provided free of charge, but money is charged for additional, often proprietary features (i.e. premium). This model is often combined with the advertising-based revenue model, where the free product is offered with advertisement while the premium is advertisement-free.
- **Advertisement:** Advertisement is most frequently used for business-to-consumer (B2C) offers. Products are offered free of charge or with a discount to users in exchange for required viewing of paid-for advertisements. Advertisement-based revenue models are often used in multi-sided markets, where a service is provided for free or at a low price on one side of the market and subsidised with revenues from other sides of the market.
- **Subscription:** Subscription-based revenue models are by far the most frequently used models in the data ecosystem, and for B2B offers in particular. Examples of subscription-based models include regular (daily, monthly or annual) payments for access to the Internet and digital content, including data. Subscription-based revenue models are often combined with the freemium revenue model.
- **Usage fees:** Usage fees are the second most frequently used revenue model used and a prominent revenue model for B2B offers. Usage fees are typically charged to customers for use of online services – or offers that are provided “Everything-as-a-Service” (XaaS), such as cloud computing. These services are offered through a pay-as-you-go (PAYG) model, where usage fees are charged for the actual use of the service.
- **Selling of goods** (including digital content): Asset sale is still used in the data ecosystem, including by service platform providers that sell sensor-equipped smart devices as a source for generating data and delivering added-value services. This includes pay-per-download revenue models, where users pay per item of download, including data and digital content such as music and videos.
- **Selling of services:** This revenue model includes the provision of B2B services including services provided by intermediaries, such as web hosting and payment processing. It thus overlaps with the revenue models that are based on subscriptions and usage fees often used for IT service contracts.
- **Licensing:** This revenue model is often used to generate revenues from intangible assets that are protected through IPRs, such as patents and copyrights. Licensing is thus often used to monetise data, software, and software components, including algorithms, libraries and APIs.
- **Commission fees:** This is mainly used in B2C markets by intermediaries that use data analytics to better match supply and demand. Payments are often calculated based on a percentage of the price of products supplied, and will only be obtained when successfully matching supply and demand.

Source: OECD (2015^[2]), *Data-Driven Innovation: Big Data for Growth and Well-Being*, <http://dx.doi.org/10.1787/9789264229358-en>.

Other well-known examples are Facebook and Google, whose vast collections of personal data are a valuable resource for advertisers and other third parties, including researchers. In some cases, these firms collaborate with third parties such as analytic service providers that help exploit the data; examples are Gnip, a social media API aggregation company and Datasift, a start-up that “enable organisations to identify and extract valuable insights from all types of human-generated data in real-time”. In other cases, third parties are provided access to the data so they can use and/or commercialise it. An example is the now-defunct Cambridge Analytica, a data mining, analysis and brokerage company that became known for having accessed personal data on more than 50 million individuals, although only about 270 000 individuals “had consented to having their data harvested” (Granville, 2018^[31]; Cadwalladr and Graham-Harrison, 2018^[32]).

However, data commercialisation remains below its potential, even among data-intensive firms, despite the increasing interest of organisations to commercialise their data and meet the growing demand for data. According to a recent survey by Forester Research of almost 1 300 data and analytics businesses across the globe, only one-third of the respondents reported they are commercialising their data. High-tech, utilities and financial services rank among the top industries commercialising their data, while pharmaceuticals, government and health care are in the bottom of the list (Belissent, 2017^[33]).

At the Copenhagen Expert Workshop some experts expressed doubts about the potential of data markets, arguing that there was something fundamentally new and unique about data, which made pricing a challenge and required looking beyond current market-based models. The success of certain data intermediaries (including the aforementioned data brokers and data aggregators) and their business models suggests however otherwise. These businesses are using and combining different revenue models (Box 2.3) and leveraging multi-sided markets²⁹ to cross-subsidise various activities and assure the collection and commercialisation of data. That said, new business models that sufficiently take into account the risks and interests of all relevant stakeholders may be needed.

The emergence of data intermediaries who provide potential sellers and buyers with services such as standard-licence schemes, and a payment-and-data exchange infrastructure, could make the commercialisation of data more mainstream. Even less data-savvy firms may find it easier to commercialise their data.

Open data

Open data are the most prominent approach used to enhance access to data (OECD, 2015^[2]). In the public sector, open government data has been promoted for many years by initiatives, including data.gov (United States), data.gov.uk (United Kingdom), data.gov.fr (France), or data.go.jp (Japan) (Ubaldi, 2013^[17]). In science, the term “openness” generally means “access [to research data]³⁰ on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination”, as specified by the OECD (2006^[34]) *Recommendation of the Council concerning Access to Research Data from Public Funding* (Recommendation concerning Access to Research Data) (see also OECD (2005^[35])).

Most definitions for open data, including those cited in the aforementioned OECD Council Recommendation, point to a number of criteria or “principles” for open data. The most prominent are non-discriminatory access and costs of access. Some definitions also put emphasis on redistribution. The International Open Data Charter, for example, defines open data as “digital data that is made available with the technical and legal characteristics necessary for it to be freely used, re-used and redistributed by anyone, anytime, anywhere” (International Open Data Charter, n.d.^[36]).³¹

Because open data should be accessed on “equal or non-discriminatory terms” (OECD, 2006^[34]), the conditions under which data can be provided via open access are very limited. In most cases, for instance, confidential data, such as personal data, cannot be shared via open access. Furthermore, as highlighted above, open data is expected to be provided for free or at no more than the marginal cost of production and dissemination. Therefore, businesses that want to commercialise their data, either directly by selling the data or indirectly by providing added-value services, may find open data less attractive.

That said, organisations in the public and private sector are increasingly recognising that non-discriminatory access is crucial for maximising the value of data, as it creates new business opportunities and economic and social benefits. However, assessing the resulting economic and social benefits of moving towards open data remains challenging. As highlighted by Dan Meisner, Thomson Reuters’ Head of Capability for Open Data, there are indirect benefits and network effects at play that “don’t really fit very well into an Excel model for calculating your internal rate of return” (Box 2.4).

Box 2.4. Open data in the private sector: The case of Thomson Reuters

Thomson Reuters is a multinational organisation that provides news and information to industry professionals. Formed through Thomson Corporation’s merger with Reuters Group in 2008, the company has since continued to grow, in part through acquisition. Because of this growth path, there have been challenges in integrating legacy data resources and systems.

In order to tackle this challenge, Thomson Reuters began by establishing central data repositories to ensure that key entity data, such as organisations and people, were only created and stored once. Each of these entities was assigned a Permanent Identifier (PermID, see <https://permid.org/>) as its unique reference point.

Thomson Reuters decided to expose PermID to its customers. Past experience has taught the industry that releasing proprietary identifiers with restrictive licensing conditions can create significant problems. One key issue is the inability to expose these proprietary identifiers to customers’ clients, and even, in some cases, to other departments within the same business. Thomson Reuters realised that the only way customers would embrace PermID was if it took an open access approach.

Being a commercial information business meant that Thomson Reuters had to justify this decision internally. This can be challenging because “customers see an awful lot of value in this but commercially it’s not easy to put a value on it”. The issue is that these indirect benefits and network effects “don’t really fit very well into an Excel model for calculating your internal rate of return”.

Ultimately, the financial case was made, based on the recognition that “the incremental cost [...] to expose [the data] externally is not that great”. As such, Thomson Reuters decided to publish a subset of its data, including associated PermIDs under a Creative Commons licence (CC-BY 4.0). An extended set of fields has been released under a Creative Commons non-commercial licence (CC-NC 3.0). They launched this service as Open PermID in 2015, obtaining an Open Data Institute (ODI) Open Data Certificate in the process of release.

Source: Open Data Institute (2016^[37]), *Open Enterprise: How Three Big Businesses Create Value with Open Innovation*, <https://theodi.org/article/open-enterprise-how-three-big-businesses-create-value-with-open-innovation/>.

Data portability

Data portability is often regarded as a promising means for promoting cross-sectoral re-use of data while strengthening the control rights of individuals over their personal data and businesses (in particular small and medium-sized enterprises [SMEs]) over their business data (Productivity Commission, 2017^[16]). Data portability provides restricted access through which data holders can provide customer data in a commonly used, machine-readable structured format, either to the customer or to a third party chosen by the customer. Prominent data portability initiatives include the US government's My Data series, launched in 2010. The Green Button energy initiative is an example (US Department of Energy, n.d.^[38]). Other examples are the Midata data portability initiative of the United Kingdom in 2011 (Department for Business Innovation and Skills (UK), 2011^[39]), the right to data portability (Art. 20) of the European Union (2016^[1]) GDPR, and Australia's CDR legislation (2019).

However, data portability initiatives may vary significantly in terms of their nature and scope across jurisdictions. The GDPR right to data portability, for instance, states that "the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance". It differs in important ways from the data portability concept explored in the voluntary based Midata initiative of the United Kingdom government.

The following points highlight in greater detail the various rationales for data portability. They are based in particular on the Midata initiative, a voluntary data portability initiative in the United Kingdom, and the European Union's GDPR and Australia's CDR. The latter two are mandatory and are among the most recent and comprehensive data portability frameworks to date.

- Data portability as a means to achieve "informational self-determination":** The EU data protection regime is underpinned, in part, by the objective of providing individuals with greater control over their personal data (Kokott and Sobotta, 2013^[40]). Only personal data are within the scope of the GDPR. The right does not apply to any data that are anonymous or do not concern the data subject. However, pseudonymous data that can be clearly linked to a data subject (e.g. by him or her providing the respective identifier), are within the scope. Individuals are, for example, provided with a right to access their personal data and a right to delete or amend this data in certain circumstances to facilitate control over their data. This objective is reflected in some EU member state legal systems³² and the GDPR now explicitly recognises this objective of data protection law.³³ Moreover, it could be argued that this control over personal data is one of the objectives of the right to data protection, a right recognised in Article 8 of the EU Charter of Fundamental Rights, which differentiates that right from the right to privacy (Lynskey, 2015^[41]).
- Data portability as a means to increase competition and choice:** Data portability is expected to increase competition between providers of digital goods and services (e.g. social networking service providers) and in other analogue markets (e.g. utilities markets) (Chapter 3). In the case of the GDPR, some EU member states objected to the inclusion of the right to data portability, which they viewed as a tool to enhance competition, thus falling outside the scope of data protection law (Council of the European Union, 2014^[42]; Graef, 2015^[43]). Data portability may enhance competition by i) reducing information asymmetries between individuals and the providers of goods and services; ii) limiting switching costs for individuals; and iii) potentially reducing barriers to market entry. The importance of data portability for fostering

competition has led to discussions about the extent to which businesses should be granted data portability rights in some OECD countries.

- **Data portability as a means to encourage markets in new products:** In addition to fostering competition on the markets to which the consumer or transaction data relate (for instance, the electricity or social network services market), data portability may also act as a stimulus for innovation and the creation of new products and services, or the expansion of existing markets. According to the Midata impact assessment, the programme was rolled out in the anticipation that the release of transaction data would stimulate innovation and the expansion of third-party choice engines such as price comparison websites (Department for Business Innovation and Skills (UK), 2012^[44]). The Midata consultation also highlighted other potential spin-off services. For example, a leading Finnish grocery retailer has coupled with a third party to inform customers of the nutritional content of their shopping basket based on data aggregated through loyalty cards. This, according to the government, provides a “real-time weight and diet management tool for individuals and families” (Department for Business Innovation and Skills (UK), 2012^[45]).
- **Data portability as a means to facilitate data flows:** In 2017, the European Commission proposed an initiative on the “free flow of non-personal data in the European Union” and published a framework proposal specifically on this matter (European Commission, 2017^[46]).³⁴ While this initiative encompasses all varieties of non-personal data, the right to data portability is expected to contribute to the attainment of this objective. The “Estonian Vision Paper on the Free Movement of Data” (European Union, 2017^[47]), which presents data flows as “the Fifth Freedom of the European Union”, goes a step further by proposing a “framework for data access and portability of personal and non-personal data in the private sector”.³⁵

To what extent data portability may effectively empower individuals and foster competition and innovation remains to be seen. Estimates on the costs and benefits of data portability are still rare. Although not specific to data, other portability studies suggest that data portability may have overall positive economic effects, specifically by reducing switching costs. For example, a study on current limitations to move mobile apps across platforms shows that switching cost can be a barrier for moving from one platform to another. Enabling app portability, changing from Apple’s iOS to another smartphone operating system, for instance, would help reduce switching cost, which are estimated to be between USD 122 and USD 301 per device (OECD, 2013^[48]; iClarified, 2012^[49]). Studies on mobile number portability (MNP) show that MNP reduces average prices by 6% to up to 12% and encourages switching when the switching process is rapid (e.g. less than 5 days) (Lyons, 2006^[50]).³⁶

Other restricted data-sharing arrangements

In cases where data are considered too confidential to be shared openly with the public or where there are legitimate commercial and non-commercial interests opposing such sharing, restricted data-sharing arrangements can be more appropriate. This is the case when there may be privacy, intellectual property (e.g. copyright and trade secrets), and organisational or national security concerns legitimately preventing open sharing. In these cases, however, there can still be a strong economic and/or social rationale for sharing data between data users within a restricted community,³⁷ under voluntary and mutually agreed terms.

It is, for example, common to find restricted data-sharing agreements in areas such as digital security (Box 2.5), science and research (Box 2.6), and as part of business arrangements for shared resources (e.g., within joint ventures). These voluntary data-sharing arrangements

can be based on commercial or non-commercial terms depending on the context. Two cases are highlighted in the following subsections in more detail: i) data partnerships, which are based on the recognition that data sharing provides not only significant economic benefit to data users, but also to data holders; and ii) data for societal objectives initiatives, where data are shared to support societal objectives.

Box 2.5. Rationale for promoting data sharing on digital security risk

Data sharing is a cost-effective way to undermine the economics of hacking. When more companies share information, they can start leveraging that information to better defend their own systems and prevent hackers from using the same method for multiple breaches, driving up the cost of successful attacks. For example, if the first targeted firm shares the identifying characteristics of the attack with all its partners, who in turn share with their partners, even if the first attack was successful, the rest of the network will have the knowledge needed for greater resilience. In this model, the adversary must craft a unique attack method for each target and will experience significantly higher costs that may be unsustainable for all but the most sophisticated attackers.

Data sharing creates positive externalities within a sharing network. In general, externalities arise when one organisation's behaviour has side effects that affect the net risk borne by others. This is related to strategic complementarities, where agents' decisions mutually reinforce one another and an agent's marginal return increases when the other agents' increase their action. Access to better digital security information can enrich existing information, making it more actionable and enabling all firms to make better risk management decisions.

Data sharing enables benchmarking within the sharing community (e.g. comparing to peers in digital security readiness and digital security risk management practice) and promotes good practice. The promise of access to data supporting peer-to-peer comparisons could incentivise a wide variety of organisations to participate in an information sharing effort. For example, the opportunity to better understand what digital security risk management investments other companies are making – even at an anonymised, aggregate level – and what success these companies are having in risk reduction and mitigation would likely incentivise participation.

Greater data sharing can encourage the growth of a digital security product market as well as a digital insurance market by allowing better quantification and more accurate assessment of risk and the effectiveness of security products. The limited availability of data on digital security incidents, the rapid pace of change in the nature of digital security risk and uncertainty about the effectiveness of different digital security products and risk management practices have a negative impact on the supply of insurance coverage for digital security risk and lead to challenges in underwriting coverage. The lack of historical data also effects the availability of reinsurance coverage for digital security risk. This may be an additional impediment to the capacity of primary insurers to provide coverage.

Source: OECD (2017^[51]), “Summary of OECD Expert Workshop on Improving the Measurement of Digital Security Incidents and Risk Management”.

As membership to the community is critical, the main emphasis is on mechanisms for governing membership. In the case of the Genomic Data Commons (GDC) of the US National Cancer Institute (NCI) (presented in Box 2.6), for example, access is granted by programme-specific Data Access Committees (DACs).

Box 2.6. Data-sharing communities in science: The case of the National Cancer Institute's Genomic Data Commons

The GDC of the NCI “provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine”. The GDC Data Portal enables researchers to i) search and query genomic data; ii) download data directly from the web browser; and iii) analyse cancer data including clinical information, genomic characterisation data, and high-level sequence analysis of tumour genomes.

Some data in the GDC (including high-level genomic data that are not individually identifiable, as well as most clinical and all bio specimen data) are provided as open data and thus require no authentication or authorisation to access it. Any user accessing GDC open data must, however, adhere to the US National Institutes of Health (NIH) Genomic Data Sharing (GDS) Policy, which indicates that investigators who download unrestricted-access data from NIH-designated data repositories should: i) “not attempt to identify individual human research participants from whom the data were obtained”; and ii) “acknowledge in all oral or written presentations, disclosures, or publications the specific data set(s) or applicable accession number(s) and the NIH-designated data repositories through which the investigator accessed any data”.

Other data, including individually identifiable data such as low-level genomic sequencing data, germline variants, SNP6 genotype data, and certain clinical data, are provided with controlled access and required authorisation and authentication for access. In this case, access is granted by programme-specific DACs. The DACs review, approve or disapprove all requests from the research community for data access. Decisions to grant access are made based on whether the request conforms to the specifications within the NIH GDS Policy and programme-specific requirements or procedures (if any). All uses proposed for controlled-access data must be consistent with the data use limitations for the data set as indicated by the submitting institution and identified on the public website for database of Genotypes and Phenotypes (dbGaP). DACs also review and approve or disapprove all requests for access to dbGaP data for programmatic oversight by NIH employees.

By June 2017, the GDC had collected and harmonised more than 4.5 PB of cancer genomics data of over 30 000 cancer patients. As of October 2015, almost 7.1 million user downloads or 3 PB of controlled-access data had been approved compared to almost 64 million downloads or 122 terabytes (TB) for open access data. Most user downloads are from locations outside of the United States – 3.2 PB) – compared to 3.6 TB from within the United States.

Source: National Cancer Institute (n.d.^[52]), *The NCI's Genomic Data Commons*, <https://gdc.cancer.gov>.

Data-sharing partnerships (including data public-private partnerships)

In data partnerships organisations agree to share and mutually enrich their data sets, including through cross-licensing agreements. One big advantage is the facilitation of joint production or co-operation with suppliers, customers (consumers) or even potential competitors. This also enables the data holder to create additional value and insights that a single organisation would not be able to create. This provides opportunities “to join forces *without* merging” (Konsynski and McFarlan, 1990^[53]). Examples include:

- The co-operation on air-mileage credit points between airline and credit card companies, based on sharing data on their joint customer base. Airline companies can increase the

loyalty of their customers, while the credit card companies gain access to a new and highly credit-worthy customer base for cross-marketing (Konsynski and McFarlan, 1990^[53]).

- The pooling of data between Take Nectar, a UK-based programme for loyalty cards that collaborates with firms such as Sainsbury (groceries), BP (gasoline) and Hertz (car rentals). “Sharing aggregated data allows the three companies to gain a broader, more complete perspective on consumer behaviour, while safeguarding their competitive positions” (Chui, Manyika and Kuiken, 2014^[54]).
- The joint venture between DuPont Pioneer and John Deere. This data partnership, which was initiated in 2014, aimed at the development of a joint agricultural data tool, which “links Pioneer’s Field360 services, a suite of precision agronomy software, with John Deere Wireless Data Transfer architecture, JDLink and MyJohnDeere” (Banham, 2014^[55]).
- The collaboration of Telefónica with organisations such as Facebook, Microsoft, and the United Nations Children’s Fund (UNICEF) to exchange data of common customers (based on the customers’ consent) for Telefónica’s personalised AI-enabled service Aura. Thanks to this collaboration, customers will be able to talk to Aura through Telefónica’s own channels and third-party platforms like Facebook Messenger, and, in the future, through Google Assistant and Microsoft Cortana.

Similar partnerships also exist in the form of data PPPs:

- For TfL, the provision of data (including through open data) enabled new strategic partnerships with major data, software and Internet services providers such as Google, Waze, Twitter and Apple. In some cases, this enabled TfL to gain access to new data sources and crowdsource new traffic data (“bringing new data back”), to undertake new analysis and thus to improve TfL’s business operation. In doing so, TfL could gain access to updated navigation information on road works and traffic incidents, and enhance the efficiency of its planning and operation (Telefónica, n.d.^[56]).
- The Norwegian tax authority has partnered with the financial sector to implement automatic exchange of loan-application data. Instead of having to repeatedly ask users for information they have already provided to public administration, such data can be re-used (based on users’ consent) during loan applications. The data are stored by Altinn, a digital infrastructure that links data from public agencies, municipalities and registers of more than 4 million inhabitants and 1 million enterprises in Norway. The benefit for financial institutions and their customers is a more efficient credit-risk assessment process, leading to loan grants being given in the span of minutes rather than days. Conversely, tax authorities may get automatic access to certain account data in their pursuance of tax fraud. This requires, however, an appropriate legal basis for such access.

However, data partnerships, including data PPPs, raise several challenges. For instance, ensuring a fair data-sharing agreement between partners can sometimes be difficult, in particular where partners have different levels of market power. Privacy and IPR considerations may also limit the potential of data partnerships by making it harder to sustain data sharing in some cases (see for comparison barriers to knowledge sharing during pre-competitive drug discovery). Where data partnerships involve competing businesses, data sharing may increase the risk of implicit collusion including the formation of cartels and fixing of price. In the case of data PPPs, there may also be some challenges due to the double role of governments, namely as an authority and service data provider. In this case, questions have been raised about what types of rules should apply for this type of data sharing, and what should the private sector exchange in return for the data.

Data for social-good initiatives

Data-sharing arrangements can also be found where private-sector data are provided (donated) to support societal objectives, ranging from science and health care research to policy making. In an era of declining responses to national surveys, the re-use of public and private-sector data can significantly improve the power and quality of statistics, not just in OECD countries, but also in developing economies (Reimsbach-Kounatze, 2015^[57]). The re-use of private-sector data also provides new opportunities to better inform public policy making, for instance, when close to real-time evidence is made available to “nowcast” policy-relevant trends (Reimsbach-Kounatze, 2015^[57]). Examples include trends goods and services consumption, flu epidemics, employment/unemployment, and the monitoring of information systems and networks to identify malware and cyberattack patterns (Choi and Varian, 2009^[58]; Harris, 2011^[59]; Carrière-Swallow and Labbé, 2013^[60]). Some of these arrangements have been classified as “data philanthropy” to highlight the gains from the charitable sharing of private-sector data for public benefit (United Nations Global Pulse, 2012^[61]).³⁸ Box 2.7 presents MasterCard’s Data Philanthropy Programs as examples.

More recently, a number of governments have started to mandate access to “data of public interest” needed to achieve societal objectives. These include France’s (2016^[62]) Law for a Digital Republic (*Loi pour une République numérique*), which defines “data of general interest” (*données d’intérêt général*). This and similar initiatives are discussed further in Chapter 5.

Box 2.7. Data philanthropy: The case of MasterCard

MasterCard is a digital technology company in the payments space. It connects buyers and sellers in over 210 countries and territories across the globe. In 2013, the company established the MasterCard Center for Inclusive Growth, an independent subsidiary focussed on promoting equitable and sustainable economic growth and financial inclusion around the world. Through its data philanthropy initiatives, the Center has been working to leverage insights for social good while maintaining the highest privacy standards.

Central to the mission of the Center is the notion of “data philanthropy” implemented through two flagship programmes:

- The **Data Grant Recipients programme**, where governments, universities, non-profits and other institutions are granted access to proprietary insights in support of research initiatives advancing social good. Data are accessed through MasterCard’s secure research centre in a way that fully protects consumer privacy. Resulting data insights are provided via a secure file-transfer mechanism approved by MasterCard Corporate Security. Data Grant Recipients have to check in with MasterCard on a quarterly basis and MasterCard reserves the right to review the research results before publication.
- The **Data Fellows programme**, where individual academic researchers are selected for longer-term research collaborations. Fellows come from a variety of disciplines and work on-site with “MasterCard’s data scientists to identify patterns, develop research papers and glean insights”. Projects, data sets, and any combinations of data are reviewed by the MasterCard’s privacy counsel.

Source: Presentation at the Copenhagen Expert Workshop by JoAnn Stonier (Chief Data Officer, MasterCard).

References

- Abrams, M. (2014), *The Origins of Personal Data and its Implications for Governance*, [14]
<http://dx.doi.org/10.2139/ssrn.2510927>.
- Banham, R. (2014), *Who Owns Farmers' Big Data?*, [55]
<https://www.forbes.com/sites/emc/2014/07/08/who-owns-farmers-big-data/>.
- Beagrie, N. and J. Houghton (2016), *The Value and Impact of the European Bioinformatics Institute*, [26]
<https://beagrie.com/static/resource/EBI-impact-report.pdf>.
- Beagrie, N. and J. Houghton (2014), *The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres*, JISC, Bristol and London, [23]
<http://repository.jisc.ac.uk/5568/>.
- Beagrie, N. and J. Houghton (2013), *The Value and Impact of the Archaeology Data Services*, [24]
 Joint Information Systems Committee, Bristol and London.
- Beagrie, N. and J. Houghton (2013), *The Value and Impact of the British Atmospheric Data Centre*, [25]
 Joint Information Systems Committee and the Natural Environment Research Council UK, Bristol and London,
http://www.jisc.ac.uk/whatwedo/programmes/di_directions/strategicdirections/badc.aspx.
- Beagrie, N. and J. Houghton (2012), *Economic Evaluation of Research Data Infrastructure (ESDS)*, [22]
 Economic and Social Research Council, London,
<https://esrc.ukri.org/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/>.
- Belissent, J. (2017), “Insights Services Drive Data Commercialization”, *Insights*, [33]
https://go.forrester.com/blogs/17-03-08-insights_services_drive_data_commercialization/.
- Cadwalladr, C. and E. Graham-Harrison (2018), “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach”, *The Guardian*, [32]
<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
- Carrière-Swallow, Y. and F. Labbé (2013), “Nowcasting with Google trends in an emerging market”, *Journal of Forecasting*, Vol. 32/No. 4, pp. 289-298. [60]
- Choi, H. and H. Varian (2009), “Predicting the present with Google trends”, *Google Research Blog*, [58]
<http://dx.doi.org/10.2139/ssrn.1659302>.
- Chui, M., J. Manyika and S. Kuiken (2014), *What executives should know about open data*, [54]
<http://www.mckinsey.com/industries/high-tech/our-insights/what-executives-should-know-about-open-data> (accessed on 11 February 2019).
- Council of Europe (2015), *Criminal justice access to data in the cloud: challenges*, [73]
<https://rm.coe.int/1680304b59>.

- Council of the European Union (2014), *Interinstitutional File: 2012/0011 (COD)*, 5879/14, [42]
https://eur-lex.europa.eu/procedure/EN/2012_11.
- Creative Commons (2013), *Public domain*, [65]
https://wiki.creativecommons.org/wiki/public_domain (accessed on 5 February 2019).
- DCMI Usage Board (2012), *DCMI Metadata Terms*, <http://dublincore.org/documents/dcmi-terms/>. [74]
- Department for Business Innovation and Skills (UK) (2012), *Midata: Government response to 2012 consultation*, BIS/12/1283, [45]
http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/34700/12-1283-midata-government-response-to-2012-consultation.pdf.
- Department for Business Innovation and Skills (UK) (2012), *Midata: Impact assessment for midata*, [44]
http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32689/12-944-midata-impact-assessment.pdf.
- Department for Business Innovation and Skills (UK) (2011), “Better Choices: Better Deals – Consumers Powering Growth”, <http://www.gov.uk>. [39]
- Department of the Prime Minister and Cabinet (Australia) (2017), *Information about the Data Integration Partnership for Australia*, [29]
<http://www.pmc.gov.au/sites/default/files/publications/DIPA-information.pdf>.
- Dumbill, E. (2012), *Microsoft’s plan for Hadoop and big data*, [21]
<http://radar.oreilly.com/2012/01/microsoft-big-data.html>.
- Enterprivacy Consulting Group (2017), *Categories of Personal Information*, [72]
<https://enterprivacy.com/2017/03/01/categories-of-personal-information/>.
- European Commission (2017), *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Building a European Data Economy”*, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2017%3A9%3AFIN>. [46]
- European Union (2017), *Estonian Vision Paper on the Free Movement of Data - the Fifth Freedom of the European Union*, https://www.eu2017.ee/sites/default/files/inline-files/EU2017_FMD_visionpaper.pdf. [47]
- European Union (2016), *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*, <http://data.europa.eu/eli/reg/2016/679/oj>. [1]
- Federal Trade Commission (US) (2014), *Data brokers: A call for transparency and accountability*, <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. [27]

- Frischmann, B., M. Madison and K. Strandburg (eds.) (2014), *Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo*, Oxford University Press. [19]
- German Constitutional Court (1983), *1983 Population Census Decision (judgment of 15 December 1983, 1 BvR 209/83, BVerfGE 65)*. [70]
- Government of France (2016), *Loi pour une République numérique*, <http://www.senat.fr/leg/pjl15-744.html>. [62]
- Graef, I. (2015), “Mandating Portability and Interoperability in Online Social Networks: Regulatory and Competition Law Issues in the European Union”, *Telecommunications Policy*, Vol. 39/No. 6, pp. 502-514, <http://dx.doi.org/10.2139/ssrn.2296906>. [43]
- Granville, K. (2018), “Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens”, *The New York Times*, <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>. [31]
- Harris, D. (2011), “Hadoop kills zombies too! Is there anything it can’t solve?”, *Gigaom*, <http://gigaom.com/cloud/hadoop-kills-zombies-too-is-there-anything-it-cant-solve/>. [59]
- iClarified (2012), *Goldman Sachs Values iPhone/iPad Customer Base at \$295 Billion*, <https://www.iclarified.com/22914/goldman-sachs-values-iphoneipad-customer-base-at-295-billion>. [49]
- International Open Data Charter (n.d.), *Principles*, <https://opendatacharter.net/principles/> (accessed on 11 February 2019). [36]
- ISO/IEC (2018), *Privacy enhancing data de-identification terminology and classification of techniques*, <http://www.iso.org/standard/69373.html>. [64]
- ISO/IEC (2017), *Information technology -- Cloud computing -- Interoperability and portability*, <http://www.iso.org/standard/66639.html>. [9]
- Kokott, J. and C. Sobotta (2013), “The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR”, *International Data Privacy Law*, Vol. 3/222. [40]
- Konsynski, B. and F. McFarlan (1990), *Information Partnerships – Shared Data, Shared Scale*, <https://hbr.org/1990/09/information-partnerships-shared-data-shared-scale>. [53]
- Lynskey, O. (2015), *The Foundations of EU Data Protection Law*, Oxford University Press. [41]
- Lyons, S. (2006), “Measuring the Benefits of Mobile Number Portability”, *Trinity College Dublin*, http://www.tcd.ie/Economics/TEP/2006_papers/TEP9.pdf. [50]
- Narayanan, A. and V. Shmatikov (2006), “How To Break Anonymity of the Netflix Prize Dataset”, *CoRR abs/cs/0610105*, <http://arxiv.org/abs/cs/0610105>. [6]

- National Board of Trade [Sweden] (2014), “No Transfer, No Trade: the Importance of Cross-Border Data Transfers for Companies Based in Sweden”, *Kommerskollegium* 1, http://www.kommers.se/Documents/dokumentarkiv/publikationer/2014/No_Transfer_No_Trade_webb.pdf. [63]
- National Cancer Institute (n.d.), *The NCI’s Genomic Data Commons*, <https://gdc.cancer.gov> (accessed on 11 February 2019). [52]
- OECD (2017), “Business models for sustainable research data repositories”, *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, <http://dx.doi.org/10.1787/302b12bb-en>. [20]
- OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264276284-en>. [69]
- OECD (2017), “Summary of OECD Expert Workshop on Improving the Measurement of Digital Security Incidents and Risk Management” (internal document). [51]
- OECD (2016), “Research ethics and new forms of data for social and economic research”, *OECD Science, Technology and Industry Policy Papers*, No. 34, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jln7vnpxs32-en>. [30]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [2]
- OECD (2014), *Public sector*, OECD, Paris, <http://stats.oecd.org/glossary/detail.asp?ID=2199>. [78]
- OECD (2014), *Summary of OECD Expert Roundtable: “Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking”*, OECD, Paris, <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg%282014%293&doclanguage=en>. [15]
- OECD (2013), “Exploring the economics of personal data: A survey of methodologies for measuring monetary value”, *OECD Digital Economy Papers*, No. 220, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k486qtxldmq-en>. [3]
- OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>. [5]
- OECD (2013), “The app economy”, *OECD Digital Economy Papers*, No. 230, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k3ttftlv95k-en>. [48]
- OECD (2011), *Thirty Years after the OECD Privacy Guidelines*, OECD, Paris, <http://www.oecd.org/sti/ieconomy/49710223.pdf>. [8]
- OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/122/122.en.pdf>. [10]

- OECD (2006), *Recommendation of the Council Concerning Access to Research Data from Public Funding*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>. [34]
- OECD (2005), *Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, <http://www.oecd.org/sti/sci-tech/38500813.pdf>. [35]
- OECD (2001), *Private sector*, OECD, Paris, <https://stats.oecd.org/glossary/detail.asp?ID=2130>. [11]
- OECD (forthcoming), *Promoting comparability in personal data breach notification reporting*, OECD Publishing, Paris. [79]
- Ohm, P. (2009), “The rise and fall of invasive ISP surveillance”, *University of Illinois Law Review* No. 08-22, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1261344. [7]
- Open Data Institute (2017), *What is ‘open data’ and why should we care?*, <https://theodi.org/article/what-is-open-data-and-why-should-we-care/>. [77]
- Open Data Institute (2016), *Open enterprise: how three big businesses create value with open innovation*, <https://theodi.org/article/open-enterprise-how-three-big-businesses-create-value-with-open-innovation/>. [37]
- Open Knowledge International (n.d.), *Open Data*, <http://opendatahandbook.org/glossary/en/terms/open-data/> (accessed on 5 February 2019). [68]
- Productivity Commission (2017), *Productivity Commission Inquiry Report: Data Availability and Use*, Productivity Commission, <https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf> (accessed on 19 March 2018). [16]
- re3data.org (2018), *2,000 Data Repositories and Science Europe’s Framework for Discipline-specific Research Data Management*, <https://blog.datacite.org/re3data-science-europe/>. [71]
- Reimsbach-Kounatze, C. (2015), “The proliferation of “big data” and implications for official statistics and statistical agencies: A preliminary analysis”, *OECD Digital Economy Papers*, No. 245, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5js7t9wqzvg8-en>. [57]
- ResDAC (n.d.), *CMS Virtual Research Data Center (VRDC)*, <https://www.resdac.org/cms-virtual-research-data-center-vrdc> (accessed on 5 February 2019). [18]
- Schneier, B. (2009), *A Taxonomy of Social Networking Data*, https://www.schneier.com/blog/archives/2009/11/a_taxonomy_of_s.html (accessed on 1 September 2018). [12]
- Symantec (2015), *Underground black market: Thriving trade in stolen data, malware, and attack services*, <http://www.symantec.com/connect/blogs/underground-black-market-thriving-trade-stolen-data-malware-and-attack-services>. [67]
- Taylor, L. (2013), *Hacking a Path through the Personal Data Ecosystem*, <https://linnettaylor.wordpress.com/2013/12/12/hacking-a-path-through-the-personal-data-ecosystem/> (accessed on 3 September 2018). [4]

- Telefónica (n.d.), *Aura*, <https://aura.telefonica.com/> (accessed on 29 September 2019). [56]
- Ubaldi, B. (2013), “Open government data: Towards empirical analysis of open government data initiatives”, *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k46bj4f03s7-en>. [17]
- UNESCO (2003), *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cybe*, http://portal.unesco.org/en/ev.php-URL_ID=17717&URL_DO=DO_TOPIC&URL_SECTION=201.html. [66]
- United Nations Global Pulse (2012), *Big data for development: Opportunities & challenges*, <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>. [61]
- Urquhart, L., N. Sailaja and D. Mcauley (2017), “Realising the right to data portability for the domestic Internet of things”, *Personal and Ubiquitous Computing*, <http://dx.doi.org/10.1007/s00779-017-1069-2>. [28]
- US Copyright Office (n.d.), *Definitions*, <http://www.copyright.gov/help/faq-definitions.html> (accessed on 2 February 2019). [75]
- US Department of Energy (n.d.), *Green Button: Open Energy Data*, <https://www.energy.gov/data/green-button> (accessed on 6 September 2018). [38]
- W3C (2012), *Web Ontology Language (OWL)*, <http://www.w3.org/OWL/>. [76]
- WEF (2014), *Rethinking Personal Data: A New Lens for Strengthening Trust*, http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf [13]

Notes

¹ Referring to most privacy frameworks, WEF (2014_[13]), for instance, notes that “many existing privacy regulatory frameworks do not take this [the heterogeneous nature of data] into account. The effect is that they indiscriminately apply the same rules to different types of data, resulting in an inefficient and less than trustworthy ecosystem.”

² Art. 9(1) GDPR (European Union, 2016_[1]) states that “[p]rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation shall be prohibited”.

³ The OECD (forthcoming^[79]) report “Promoting comparability in personal data breach notification reporting”, for example, presents a number of taxonomies that have been used by some privacy enforcement authorities (PEAs) to monitor and regulate personal data breaches. Sweden’s National Board of Trade (2014^[63]), as another example, discusses a data taxonomy developed to guide the governance of cross border (personal) data flows in the context of trade. And Enterprivacy Consulting Group (2017^[72]) differentiates personal data into five (partly overlapping) subcategories: i) internal (e.g. on what a person knows or believes and his/her preferences); ii) external (e.g. on what uniquely or semi-uniquely identifies a specific individual including his/her ethnicity, sexual preferences, behavioural tendencies, demographic, health and physical characteristics); iii) historical (about an individual’s personal history); iv) financial (e.g. on an individual’s financial account, an individual’s purchasing, spending or income, his/her ownership rights and credit reputation); and v) social (e.g. on an individual’s educational or professional career, criminal records, family and social networks, and communication pattern).

⁴ ISO/IEC 19941 (2017^[9]) establishes common terminology and concepts and differentiates between: i) customer data, that is mainly contributed data from a user of a cloud service provider (e.g. credentials, personal health data and medical records, and financial details); ii) derived data, that is observed and/or inferred data about user; iii) cloud service provider data including mainly operations data and access and authentication data; and iv) account data, including mainly account or administration contact information and payment instrument data.

⁵ These are aligned with the privacy enhancing de-identification techniques included under the ISO/IEC 20889 (2018^[64]) standard.

⁶ In most cases, personal data that is effectively anonymised and/or aggregated would fall out of the scope of privacy regulation frameworks. However, pseudonymous data that can be linked back to a data subject (e.g. by him or her providing the respective identifier) can easily fall back within the scope. Unlinked pseudonymised data (data for which all identifiers are irreversible erased or replaced) is a special case. Such data will in most cases be considered non-personal; but where risks of re-identification are unneglectable, the privacy regulation frameworks will remain pertinent.

⁷ The OECD (2008^[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* (OECD PSI Recommendation) defines public sector (government) data as a subset of public-sector information (PSI), which includes not only data but also digital content, such as text documents and multimedia files. The terms “public-sector data” and “government data” are used as synonyms. The often used term “open government data” refers to public-sector data made available as open data. These data are: i) dynamic and continuously generated; ii) often directly produced by the public sector or associated with the functioning of the public sector (e.g. meteorological data, geo-spatial data, business statistics); and iii) often readily useable in commercial applications with relatively little transformation, as well as being the basis of extensive elaboration.

⁸ According to the *OECD Glossary of Statistical Terms*, the public sector is defined as covering the (central and local) government administration as well as all “public corporations including the central bank” that are engaged more or less in commercial and/or public service delivery (OECD, 2014^[78]).

⁹ See as another example data funded by the private sector, but collected, processed, preserved, and maintained by the public sector.

¹⁰ Both the OECD (2008^[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* (OECD PSI Recommendation) and the OECD (2006^[34]) *Recommendation on Principles and Guidelines for Access to Research Data from Public Funding* (hereafter the “OECD Recommendation on Research Data”), call for open access to *publicly funded data*, irrespective of whether the data are controlled by an entity in the public or private sector, acknowledges however the need to protect the commercial interests of the private sector.

¹¹ The concept of public domain used in this report goes beyond that used in the context of copyright. The United States Copyright Office (n.d.^[75]) defines public domain as follows: “The public domain is not a place. A work of authorship is in the ‘public domain’ if it is no longer under copyright protection or if it failed to meet the requirements for copyright protection. Works in the public domain

may be used freely without the permission of the former copyright owner.” Creative Commons (2013^[65]) defines work in the public domain as work that “is free for use by anyone for any purpose without restriction under copyright law. Public domain is the purest form of open/free, since no one owns or controls the material in any way”. UNESCO (2003^[66]) defines the term as “Public domain information is publicly accessible information, the use of which does not infringe any legal right, or any obligation of confidentiality. It thus refers on the one hand to the realm of all works or objects of related rights, which can be exploited by everybody without any authorisation, for instance because protection is not granted under national or international law, or because of the expiration of the term of protection. It refers on the other hand to public data and official information produced and voluntarily made available by governments or international organisations.”

¹² For example, Schneier (2009^[12]) has developed a taxonomy of personal data, using social networking sites as an example, and differentiated six types: i) **service data**, which is provided to open an account (e.g., name, address, credit card information, etc.); ii) **disclosed data**, which is entered voluntarily by the user; iii) **entrusted data**, taking as an example the comments made on other people’s entries; iv) **incidental data**, which is about a specific user, but uploaded by someone else; and v) **behavioural data**, which contains information about the actions users are undertaking; as well vi) **inferred data**, which is deduced from someone’s disclosed data, profile or activities.

¹³ See summary of the discussion with Abrams (2014^[14]) on personal data taxonomies at the 2014 OECD Expert Roundtable Discussion on “Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking” held on 21 March 2014 (OECD, 2014^[15]).

¹⁴ A special case of the taxonomy is also used for data access in the context of law enforcement, where the distinction between subscriber data, content data, and traffic (meta-)data is made (Council of Europe, 2015^[73]).

¹⁵ The OECD Privacy Guidelines (OECD, 2013^[5]) define “data controller” as “a party who, according to domestic law, is competent to decide about the contents and use of (personal and non-personal) data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf”. This term has become a very specific term of art in the privacy field, which necessitated the introduction of a different term. Where a party is in control of personal data and his privacy protection obligations are to be emphasised in the report, the term “data controller” will be used. Otherwise this report will use the more general term “data holder”.

¹⁶ The Australian Productivity Commission (2017^[16]) uses the term “consumer” for both individual consumers and SMEs.

¹⁷ Increasingly, blockchain technology (i.e. decentralised infrastructure for the storage and management of data) have been proposed as a solution to address some of the challenges related to data sharing as well. Instead of relying on a centralised operator, a blockchain operates on top of a peer-to-peer network, relying on a distributed network of peers to maintain and secure a decentralised database. What is significant for trust in data access and sharing are the following properties of blockchain technology: a blockchain is highly resilient and tamper resistant (i.e. once data has been recorded on the decentralised data store, it cannot be subsequently deleted or modified by any single party), thanks to the use of cryptography and game theoretical incentives (OECD, 2017^[69]).

¹⁸ OWL is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things (W3C, 2012^[76]).

¹⁹ The Dublin Core Metadata Terms were endorsed in the Internet Engineering Task Force (IETF) RFC 5013 and the International Organization for Standardization (ISO) Standard 15836-2009 (DCMI Usage Board, 2012^[74]).

²⁰ In citizen science projects, open data is used to involve citizens and “amateur researchers” at different stages of scientific processes, from data collection to solving more complex scientific problems.

²¹ The term “community” as used in this report does not imply that access to data is free, nor that access is unregulated. The term rather describes social groups (of individuals or organisations) that have something in common, such as norms, values, identity or interests.

²² These include: i) the leadership team of expert astronomers, which is relatively closed given that scientific and institutional credentials are required to become an active member of that group; and ii) the community of volunteers (“Zooite”), which is fully open to anyone with access to an Internet connection; but which includes a group of iii) more active participants in forum discussions that also participate in drafting scientific publications together with expert astronomers and thus have more rights.

²³ See Endnote 15.

²⁴ Some public sector initiatives also focus on the larger concept of public-sector information (PSI), which also includes public sector digital content such as multimedia content (see the OECD (2008_[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* [OECD PSI Recommendation]).

²⁵ See for a list of research data repositories the Registry of Research Data Repositories (re3data.org, 2018_[71]).

²⁶ Data and in particular personal data are also sold via illegal (cyber-crime) markets. These markets are run as online forums (mostly within channels on Internet Relay Chat servers), where criminals buy and sell software, information and services as diverse as malware code, botnets, denial of service attacks, scam page hosting, and last but not least personal data such as government-issued identification numbers, credit card numbers, user accounts, email address lists, and bank accounts (Symantec, 2015_[67]).

²⁷ The Australian Productivity Commission (2017_[16]) Data Availability and Use Inquiry Report recommends to the i) “government [to establish] a new Office of the National Data Custodian”; ii) “(predominantly existing) public bodies [to] be accredited as sector-based, national data release authorities”; and that iii) “a small number of nationally beneficial data sets be designated as National Interest Data sets”.

²⁸ The Expert Group specifically recommended that: “Ethics review bodies should, where consent for research use of personal data is not deemed possible or would impact severely upon potential research findings, evaluate the potential risks and benefits of the proposed research. If the proposed project is deemed ethically and legally justified without obtaining consent, ethics review bodies should ensure that information is made publically [sic] available about the research and the reasons why consent is not deemed practicable and should impose conditions that minimise the risk of disclosure of identities” (OECD, 2016_[30]).

²⁹ Multi- (two-) sided markets are online service platforms with distinct user groups, where activities of one group on one side of the market generate benefits (externalities or spill-overs) to other groups. As highlighted in OECD (2015_[2]), these benefits are enabled thanks to data collected on one side of the market and exploited and used on the other sides. These markets are also taking advantage of network effects emerging on at least one side. The revenue model of data generating platforms therefore relies heavily on the combination of network effects that typically affect all sides of the market of the service platform provider.

³⁰ In the OECD (2006_[34]) Recommendation on Access to Research Data, “research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.”

³¹ Other prominent definitions have been provided by the Open Data Institute (Open Data Institute, 2017_[77]) and the Open Data Handbook (Open Knowledge International, n.d._[68]). The former defines open data as “data that anyone can access, use or share” while the latter defines it as data that “can be freely accessed, used, modified and shared by anyone for any purpose – subject only, at most, to requirements to provide attribution and/or share-alike”.

³² See for instance in Germany where the concept of “informational self-determination” was recognised by the German Constitutional Court (1983^[70]).

³³ For instance, recital 7 of the GDPR explicitly states that “[n]atural persons shall have control of their own personal data”.

³⁴ See also Beagrie and Houghton (2016^[26]).

³⁵ The Estonian Vision Paper suggests for instance to extend the scope of data portability to non-personal data in both B2C and B2B through “the obligation to provide data portability [of] raw non-personal data where the consumer buys (or leases) a device that generates data to enhance competition and consumer choice, stimulate data sharing and avoid vendor lock-in” (European Union, 2017^[47]).

³⁶ Using international time-series cross-section data, the author shows that the estimated short run effect of implementing MNP was a fall in real average prices of about 6.6% for those countries with MNP delivery times of five days or less. The estimated long run reduction was significantly higher, at 12%.

³⁷ See endnote 21.

³⁸ In this context two ideas are debated: i) “data commons”, where some data are shared publicly after adequate anonymisation and aggregation; and ii) “digital smoke signals”, where sensitive data are analysed by companies, but the results are shared with governments.

3. Economic and social benefits of data access and sharing

This chapter presents the available evidence of the direct and indirect economic and social benefits of data access and sharing. It then analyses the different types of benefits in more details. These include greater transparency and empowerment of users, new business opportunities, competition and co-operation within and across sectors and nations, crowdsourcing and user-driven innovation, and increasing efficiency.

The use, and in particular the re-use, of data across the economy underline the importance of data as a new form of capital for 21st-century knowledge economies. Data cannot be depleted as it can be re-used for a theoretically unlimited range of purposes (OECD, 2015^[1]).¹ This can create beneficial spill-overs, where data can be re-used to open up significant growth opportunities, or to generate benefits across society in ways that could not be foreseen when the data were first created.² For instance, the spill-over benefits of public-sector data across the economy has motivated a range of government initiatives that have made public-sector data more openly accessible and free of costs to users. These initiatives have not only contributed to enhancing trust in governments but also have enabled data-driven innovation across the economy. Where data linkages are possible, data access and sharing can also boost spill-over benefits by enabling “super-additive” insights that may be greater than the sum of insights from isolated parts (data silos), leading to increasing returns to scope (OECD, 2015^[1]).

The re-use of data as a public, private, or public-private platform to support a range of upstream social and economic activities has led some experts to consider data as an infrastructural resource under certain conditions. Examples could include health care quality data that is re-used to assess health care system efficiencies and performance and to support health-related research activities. That said, not all data can be considered an infrastructure and certainly not a public infrastructure, as experts also made clear during the Copenhagen Expert Workshop.³

Evidence shows that data access and sharing can generate positive social and economic benefits for data providers (direct impact), their suppliers and data users (indirect impact), and the wider economy (induced impact). However, quantifying the overall benefits of data access and sharing is difficult.⁴ Recent available studies by sector (public vs. private sector) further discussed below provide a rough estimate of the magnitude of the relative effects of data access and sharing. They suggest that data access and sharing can increase the value of data to holders (direct impact), but it can help create 10 to 20 times more value for data users (indirect impact), and 20 to 50 times more value for the wider economy (induced impact). In some cases, however, data access and sharing may also reduce the producer surplus of data holders. Overall, these studies suggests that data access and sharing can help generate social and economic benefits worth between 0.1% and 1.5% of gross domestic product (GDP) in the case of public-sector data, and between 1% and 2.5% of GDP (in few studies up to 4% of GDP) when also including private-sector data.

Impact assessment studies on the economic and social benefits

Two main groups of studies are discussed in dedicated subsections, within which studies can be compared: i) studies focussing on the impact of public-sector data; and ii) those focussing on the impact of data from both the public and private sector.

Enhancing access to public-sector data

The former UK Office of Fair Trading (2006^[2]) surveyed more than 400 UK public-sector information holders (PSIHs) and 300 UK businesses buying or licensing data from PSIHs. The data-sharing arrangements employed by the PSIHs were based on open data and/or paid-for licences, which accounted for most of the cases surveyed. The study estimates that the direct impact of public sector information (PSI) (i.e. the producer surplus generated by the PSIHs) in the United Kingdom was around GBP 66 million (USD 86 million) per annum,⁵ and the indirect impact (including the consumer surplus of PSI re-use) was around GBP 518 million (USD 674 million). The study shows that the high price of paid-for licences is a major barrier to data access. It suggests that reducing costs to the level of cost recovery

would increase overall surplus. It also identified the distortion of downstream competition in the private sector through restricted access to raw data that PSIHs used themselves to provide added-value services (crowding out). The study estimates that if these issues were addressed, the producer surplus would vanish in favour for an increase of the indirect impact by GBP 585 million (USD 761 million), leading to an overall economic value of GBP 1.1 billion (USD 1.43 billion) or 0.1% of the GDP.

A report by Deloitte for the UK Department for Business, Innovation and Skills (Deloitte, 2013^[3]) focusses particularly on Trading Funds such as the HM Land Registry, the Registers of Scotland, the Companies House, the Ordnance Survey, the UK Hydrographic Office, the Environment Agency, the Met Office, and the Office of National Statistics. It estimates that the direct economic impact (as revenues of PSIHs) is around GBP 0.1 billion (USD 0.13 billion), while the indirect impact on data users and suppliers of data PSIHs is between GBP 1.2 billion (USD 1.6 billion) and GBP 1.8 billion (USD 2.4 billion) per year.⁶ The wider indirect and induced impact of PSI was conservatively estimated to be around GBP 5 billion (USD 6.5 billion) per year. This included time saved as a result of access to real-time travel data, which is valued at between GBP 15 million (USD 19.5 million) and GBP 58 million (USD 75 million). This led to an overall estimate of GBP 6 billion (USD 8 billion) to GBP 7 billion (USD 9 billion), or around 0.5% of GDP.

In two studies, ACIL Tasman (Tasman, 2008^[4]; Tasman, 2009^[5]) estimate the economic contribution of geospatial data to the economy in Australia and New Zealand respectively. The studies conclude that geospatial data provided by the public sector at costs above the cost of production and distribution and issues around data formats and licensing schemes have impeded access, mainly by smaller firms. The aggregated turnover of the geospatial data service industry in Australia was estimated to be around AUD 1.4 billion (USD 1 billion) per year, with a gross value added of AUD 682 million (USD 484 million) thanks to geospatial data from the public sector (Tasman, 2008^[4]). The indirect impact of geospatial data is estimated to have increased productivity across the economy,⁷ leading to an aggregated induced impact between AUD 6.4 billion (USD 4.5 billion) and AUD 12.6 billion (USD 9 billion) in 2008, which corresponds to 0.6% and 1.2% of GDP.⁸ (Tasman, 2008^[4]) concludes that further 5% to 15% productivity gains could be unleashed if barriers to geospatial data would be removed, with the largest impacts most likely to occur in sectors such as agriculture, transport, asset management and property.

Recent studies on PSI re-use in 27 EU countries estimate that the value of the market for PSI was of the order of EUR 28 billion (USD 33 billion) in 2008 and EUR 32 billion (USD 38 billion) in 2010 (Vickery, 2011^[6]; Vickery, 2012^[7]). The aggregate indirect and induced economic impact from PSI across the whole EU27 economy is estimated to be of the order of EUR 140 billion (USD 165 billion) annually – roughly 1.5% of GDP. A similar study focussing on the OECD PSI market estimates the value of PSI to be around USD 97 billion in 2008 and USD 111 billion by 2010 (OECD, 2015^[8]). The aggregate indirect impact is estimated to be around USD 500 billion in 2008.⁹

Estimates by Deloitte (2017^[9]) based on open data provided by Transport for London (TfL) strongly confirm the positive net benefits of open data. The study shows that the re-use of TfL's open data was generating annual economic benefits and savings of up to GBP 130 million a year for TfL customers, road users, London, and TfL itself (Table 3.1). This includes a gross value added of GBP 12 million to GBP 15 million per year for businesses which also directly created more than 500 jobs. But it does not account for the significant contribution TfL's open data has made to improving societal outcomes, facilitating innovation and improving the wider environment (e.g. air quality and lower emissions).

Table 3.1. The economic benefits and savings of open data by TfL

TfL passengers and other road users	London	TfL
Saved time for network passengers - Passengers are able to plan their journeys better with apps that use TfL's open data to provide them real-time information and advice on how to adjust routes. - This provides greater certainty on when the next bus/tube will arrive and saves time (the equivalent of an estimated GBP 70 million to GBP 90 million per annum).	Gross value added - A number of companies use and re-use TfL data commercially, generating revenue, many of which are based in London. - We estimate that the total Gross Value Add from using TfL data by these companies directly and across the supply chain and wider economy is between GBP 12 million and GBP 15 million gross value added per annum.	Savings from not having to produce apps in-house - With over 13 000 registered developers currently, TfL is allowing the market to develop innovative new transport apps and services. - This creates potential cost savings for TfL of not having to build apps itself or through co-developing with third-party developers.
Saved time for other road users - The availability of data on road works and traffic incidents can feed into Sat Navs, driving software and apps that can allow private and commercial drivers to adjust their routes to avoid congestion. - This saves time and can reduce emissions as less time is spent waiting in traffic queues and journeys are shorter.	High-value job creation - TfL open data is estimated to directly support around 500 jobs that would not have existed otherwise. - Many of these jobs are in sectors associated with high productivity.	Savings from not having to invest in campaigns and systems - The publication of open data gives passengers information directly, reducing the pressure on the Contact Centre. - Undertaking an equivalent campaign to make available this information could cost GBP 1 million – open data allows TfL to make available the same data at a much reduced cost, expanding customer reach and improving transparency. - The cost for TfL of publishing open data is estimated at around GBP 1 million annually, suggesting a significant return on investment.
Savings made from moving from SMS alerts - Passengers are able to switch to using free apps or free web services for real-time data that use TfL's open data. - This creates a cost saving for those who previously subscribed to fee-based SMS alerts, estimated to be worth up to GBP 2 million per annum. The use value of new use-value alert services is estimated to be up to GBP 3 million per annum.	Wider job creation in the supply chain A further 230 indirect jobs in the supply chain and wider economy have also been created.	Leveraging value and savings from partnerships - Through partnerships with major data and software organisations, TfL receives back significant data on areas in which it does not itself collect data (e.g. crowdsourced traffic data). - This allows TfL to undertake new analyses and improve its operations.
Better information to plan journeys, travel more easily and take more journeys - Passengers are now able to better plan journeys, enabling them to use TfL services more regularly and access other services. - This can result in more journeys on the network. Conservatively, the value of these journeys is estimated at up to GBP 20 million per annum.		
Plus improved customer satisfaction from having accurate and reliable information available instantly	Plus supporting the wider UK digital economy in London and other cities	Plus new commercial opportunities arising from open data

Source: Deloitte (2017^[9]), *Assessing the Value of TfL's Open Data and Digital Partnerships*, <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>.

Enhancing access to and sharing of public and private-sector data

A study by the McKinsey Global Institute (2013^[10]) looks at the benefits of the re-use of public and private-sector data in seven areas of the global economy (education, transportation, consumer products, electricity, oil and gas, health care and consumer finance). It estimates that the re-use of data across these seven areas could help create value worth USD 3 trillion per year worldwide.¹⁰ By scaling the results of the McKinsey Global Institute (2013^[10]) to the Group of Twenty (G20) economies, Lateral Economics (2014^[11]) estimates that open

data could increase G20 output by around USD 13 trillion over the next five years. “This would boost cumulative G20 GDP by around 1.1 percentage points of the 2% growth target over five years” (Lateral Economics, 2014^[11]). Similar scaling for Australia suggest that “more vigorous open data policies could add around AUD 16 billion per annum to the Australian economy” (this would represent almost 1% of GDP or USD 13 billion).

A study conducted by Mitsubishi Research Institute (2017^[12]) focusses on the economic roles and impacts of data platforms in Japan, in particular, their role as integrator of various forms of data and thus facilitator of data re-use across organisations. The study is based on case studies assessing the business model of major data platforms in Japan, including public-sector data such as geospatial data as well as private-sector data such as Internet of Things (IoT) data, traffic data (ranging from pedestrians to cars), consumer behaviours, and language translation data.¹¹ Data platforms are estimated to improve data re-use of firms, which face a barrier to data access. Based on the results of Nomura Research Institute (2014^[13]), the study estimates the contribution of data platforms with the assumption that the proportion of the respondents having barriers to data use is lowered by data platforms and that their contribution is proportionate to the gross value added created from data re-use. The overall gross value added attributed to data platforms in Japan was estimated to be between JPY 604 billion (USD 5 billion) and JPY 1.45 trillion (USD 13 billion) in fiscal year 2014.

The IDC and Lisbon Council (2018^[14]) study assesses the data market size and the GDP impact of the data economy in EU28 countries, focussing on the value added created from data re-use, including the provision of data and its exploitation in the private sector. The data market is defined as the marketplace where digital data are exchanged as “products” and “services” as a result of the (re-)processing of raw data. The impact on the data economy is defined more broadly as the overall effects of the data market on the economy, involving generation, collection, storage, processing, distribution, analysis elaboration, delivery, and exploitation of data enabled by digital technologies. Therefore, the overall impact is estimated by summing up the direct, indirect, and induced impact.¹² The direct impact is estimated by the volume of the data market as a proxy (i.e. revenues of data suppliers and adjusted through including imports and excluding exports). According to the study, the data market volume in EU28 countries is estimated to be EUR 59 billion in 2016 and EUR 65 billion in 2017 (an increase of roughly 20% year-on-year). The indirect impact (i.e. the impact on data suppliers and the impact on data users through innovation and efficiency gains) was above 50% of total impact in 2017. Overall, the study suggests an overall impact of the data economy impact on GDP of 2.2% (EUR 306 billion) in 2016 and 2.4% (EUR 336 billion) in 2017.

The impact on GDP in 2025 is forecasted based on three different scenarios related to the concentration of power in data access, control and exploitation. These scenarios affect the composition of direct, indirect and induced impacts:

- The Baseline scenario, which “is characterised by a healthy growth of data innovation, a moderate concentration of power by dominant data owners with a data-governance model protecting personal data rights, and an uneven but rather wide distribution of data innovation benefits in the society”, is forecasted to lead to an overall impact of 4.2% of GDP (EUR 669 billion) in 2025.
- The High Growth scenario (Data-driven Reality), “characterised by a high level of data innovation, low data power concentration, an open and transparent data-governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society”, is forecasted at 6% of GDP (EUR 1 trillion).
- The Challenge scenario (Digital Maze), which “is characterised by a low level of data innovation, a moderate level of data power concentration due to digital markets

fragmentation, and an uneven distribution of data innovation benefits in the society”, is forecasted at 3.0% of GDP (EUR 470 billion).

Main categories of economic and social benefits

The different types of benefits and cost savings highlighted in previous sections can be clustered around the following five categories: i) greater transparency, accountability and empowerment of users, for instance, when open data are used for cross-subsidising the production of public and social goods; ii) new business opportunities, including the creation of start-ups and in particular for data intermediaries and mobile app developers; iii) competition and co-operation within and across sectors and nations, including the integration of value chains; iv) crowdsourcing and user-driven innovation; and v) increased efficiency thanks to linkage and integration of data across multiple sources.

These benefits suggest that data access and sharing is a major enabling condition for open innovation, a concept that according to the OECD Innovation Strategy (OECD, 2010^[15]; OECD, 2015^[16]) describes the “use of purposive inflows and outflows of knowledge to accelerate internal innovation and expand the markets for external use of innovation”. This includes proprietary-based business models that make active use of licensing, collaborations, joint ventures, etc. “Here ‘open’ is understood to ‘denote the arms’ length flow of innovation knowledge across the boundaries of individual organisations” (OECD, 2013^[17]).

The following subsections discuss these five categories of economic and social benefits in greater detail.

Transparency, accountability and empowerment of users

Enhanced access and sharing is a key means for improving transparency and empowering users, including small and medium-sized enterprises and consumers. In the private sector, open data initiatives like the Open Banking initiative demonstrate how data can be used to help people transact, save, borrow, lend and invest their money. By increasing transparency in the financial market, the initiative can empower consumers so they become able to better compare existing offerings. This in turn can contribute to a higher level of competition in the market. It is estimated that in the United Kingdom alone, consumers could save up to GBP 70 (USD 90) per year by switching to a bank account that would better fit their needs (Staff, 2017^[18]).

In the case of data portability, customers are empowered to retrieve their data and move more easily to an alternative supplier, which puts competitive pressure on suppliers to keep prices low and compete on features, including privacy-enhancing features. As noted in the subsection “Data portability” in Chapter 2, this can: i) reduce information asymmetries between individuals and the providers of goods and services; ii) limit switching costs for individuals and lower lock-in effects; and iii) potentially reduce barriers to market entry. In doing so, data portability can contribute to more vigorous competition among vendors and greater consumer choice. For these reasons, data portability is considered not only a means to strengthening the control rights of individuals over their personal data, but also as a way to increase competition among providers of data-driven products (OECD, 2015^[1]).

In science, enhanced access and sharing, typically through open data, is critical for transparency and for scrutinising and replicating scientific results. This remains challenging, for instance, when test results of drug interventions need to be validated by the scientific community. Evidence suggests that the quality of scientific research depends on the extent to which the underlying data can be accessed by other scientists, which is not always the case. At the

Copenhagen Expert Workshop, it was emphasised that the poor availability of data required to scrutinise and replicate research results was one of the main causes for the significant share of false scientific results, and for the risk of the erosion of trust in science.

The CSTP-GSF Workshop concluded that researchers need to share the data, software, workflows and details of the computational environment in open repositories to facilitate reproducibility of their research results (OECD, 2018^[19]). Participants stressed that published articles would need to include persistent links to the underlying digital artefacts, including data and software code, to enable discoverability, and that scientific journals should include “reproducibility checks” as part of their publication process.

Citizens’ use of open data as provided by governments through their open data initiatives can also help increase openness, transparency and accountability of government activities and thus boost public trust in governments. According to the McKinsey Global Institute (2011^[20]), full use of data analytics in Europe’s 23 largest governments might reduce administrative costs by 15% to 20%, creating the equivalent of EUR 150 billion to EUR 300 billion in new value, and accelerating annual productivity growth by 0.5 percentage points over the next ten years. The main benefits would be greater operational efficiency (due to greater transparency), increased tax collection (due to customised services, for example) and fewer frauds and errors (due to automated data analytics).

All these benefits show that by empowering data users enhanced access and sharing can be used to cross-subsidise economic activities, including the production of public and social goods (such as science and research) that would otherwise require picking winners (users or applications). Governments can support the production of public goods i) by directly producing these goods; or ii) by supporting private firms’ production of public and social goods through research grants, procurement programmes, contracted research and tax incentives. But all these strategies raise several issues, including difficulties in picking winners and losers, and the fact that resources are limited.

Enhanced access and sharing can be a more efficient and politically attractive “indirect intervention” to support economic activities relying on data re-use.¹³ While this has been regarded as an important feature, some authors, such as Johnson et al. (2017^[21]) or Onsrud (2007^[22]), have considered this as a risk and hidden cost to society.¹⁴ Especially where data are provided largely for private-sector consumption without significant spill-over benefits for citizens, or where there are insufficient complementary investments in skills and infrastructures needed for the effective re-use of data (see section “Trust and empowerment for the effective re-use of data across society” in Chapter 4). Johnson et al. (2017^[21]) therefore warn that open data initiatives could be (mis-)used as “a kind of ‘smoke and mirrors’ that obscures a government’s actual commitment to citizen participation, transparency and accountability”.

Business opportunities including for data intermediaries and start-ups

Enhanced access can also create new business opportunities for smaller and larger firms. Better access to open government data, for instance, can allow entrepreneurs to develop innovative commercial and social goods and services. An example is RowdMap, an analytics company using open data to help health care plans, physician groups and hospital systems identify, quantify, and reduce low-value care. In July 2017, the company was acquired for USD 70 million by Cotiviti, a provider of analytics-driven payment accuracy solutions.

Enhanced access and sharing enables many business opportunities for data intermediaries, including data brokers, mobile apps and personal information management systems. This is because most end users, consumers and businesses included, often do not directly use raw

data. They rather rely on data intermediaries that access raw data to extract and present the embedded information in more user-friendly ways, sometimes enriched through additional, inferred, data. These intermediaries typically provide added-value services including advanced data analytic services. While businesses tend to use data brokers as main data intermediaries, consumers often access added-value information services via apps (the Copenhagen Expert Workshop). Overall, this has led to new demand for added-value services and thus to new business opportunities for new and old intermediaries, including data brokers and app developers, but also for some incumbents in information and communication technology (ICT) and non-ICT industries (e.g. telecommunication and financial services firms).

For example, a major part of the benefits of open data by TfL were realised thanks to the development of apps that used TfL open data to provide real-time traffic information for more accurate navigation systems (Table 3.1). More than 80 data feeds were made available for developers through a free unified application programming interface (API),¹⁵ which ensured accurate real-time data for over 13 000 registered developers and more than 600 apps. This generated a gross value added of GBP 12 million to GBP 15 million per year for businesses and led to the direct creation of more than 500 jobs and more than 230 indirect jobs across the supply chains and the wider London economy.

Co-operation and competition across sectors and countries

Enhanced access and sharing can facilitate joint production or co-operation with suppliers, customers or even competitors. This is not a new phenomenon. Joint research ventures or patent pools are well-known examples, where firms share common resources under non-discriminatory access regimes. This is “because independent research efforts are inhibited by complexity, expense, strategic concerns, transaction costs, or other impediments” (Frischmann, 2012^[23]). Sharing agreements are very often an important part of these collaboration efforts. In the case of data, access does not need to be open to the public, but may be limited to the partners who share their data to “overcome collective action problems, sometimes mere co-ordination problems and sometimes more difficult prisoner’s dilemma problems” (see subsection “Other restricted data-sharing arrangements” in Chapter 2).

At the Copenhagen Expert Workshop, experts presented a cases confirming that the re-use of data enabled the integration of value chains across sectors and even across national borders. The data provided by TfL through open data, for example, enabled the integration of transport and navigation services across different means of transport (i.e. multimodal transport and navigation information), a condition for the deployment of smart transportation services (Alissa Walker, 2016^[24]). Thanks to access to TfL open data, services such as Google Maps could provide more accurate multimodal navigation information including for the first and last miles.

Another example of the integration of value chains across sectors through enhanced access and sharing is the Industrial Data Space (IDS), a platform for the commercialisation of data in a business-to-business (B2B) context. The development of the IDS was motivated by the recognition that the value of data was growing when combined to deliver added-value services. The main benefits and strengths of the IDS was its ability to link and integrate data from multiple sources and of different types (e.g. product data and environment data of production) to enable the creation of “smart services”. Table 3.2 lists a few IDS use cases in logistics, where data needed to be shared across the supply chain. Another IDS use case is in mobility services. Here, different types of data (such as geolocation data of the means of transport, data on traffic flows, and maintenance data) need to be combined for innovative smart mobility services and for added-value services such as new insurance models (e.g. pay as you drive) and just-in-time maintenance services.

Table 3.2. Selected use cases in logistics of the Industrial Data Space

Application partner	Use case
AUDI AG	Transparency in supply network
DB Mobility Logistics AG/ DB Schenker	Transparency in supply chain - reinforced structured/automatised exchange of information between all involved parties along the supply chain
KOMSA AG	From shipment to customer and consumer behaviour
REWE Systems	Autonomous transparency in the logistics chain
Robert Bosch GmbH	High performance supply chain – accumulation and exchange of relevant events along the supply chain
Robert Bosch GmbH	Luggage control – support from travelling salesmen
SICK AG	Coaster – assistance system for workers
ThyssenKrupp AG	Transport logistics – optimisation of efficiency and observability of truck transport processes
ThyssenKrupp AG	Energy supply for flexible manufacturing plants
Wacker Chemie AG	Tracing of consignment of goods and alerting in case of deviation

Source: Presentation at the Copenhagen Expert Workshop by Jakob Rehof (Director, Fraunhofer Institute for Software and Systems Engineering, ISST, Dortmund, Germany).

In science and research, for instance, data-sharing platforms (research data repositories) can reduce the cost of conducting research by enabling collaboration among researchers across disciplines. As OECD (2017^[25]) shows, citing Beagrie and Houghton (2012^[26]; 2014^[27]; 2013^[28]; 2013^[29]; 2016^[30]), “there is substantial additional re-use of the stored data, with between 44% and 58% of surveyed users across the studies saying they could neither have created the data for themselves nor obtained them elsewhere”. In areas where co-operation across countries is needed to tackle global challenges, such as infectious diseases, and improve early detection and warning of emerging threats and events, data sharing is often crucial.

Box 3.1. Data portability as business facilitator? The case of Uber and Braintree

Uber allows users to request and pay for a ride with just one click via an app. The app’s simplicity has been one of Uber’s major success factors. As Uber expanded internationally, it needed a payment gateway to simplify the complexities of international mobile payment. During its initial expansion into Paris, for example, Uber had to charge passengers in US dollars and display euros on-screen. This was a large source of confusion and customer complaints.

Uber wanted a payment gateway that was specifically created for smartphones, so it would be consistently fast. But most importantly, the gateway needed to offer 100% data portability if Uber ever decided to switch providers. With many providers, merchants are not able to quickly retrieve their data, creating significant switching costs for merchants, like Uber, who rely on ease of payment to satisfy and retain customers.

Uber switched to Braintree (a PayPal company) in February 2011 for its entire international and US-based payments. After a cost-intensive process to extract its users’ data from their previous payment provider, Uber was able to quickly and easily integrate Braintree’s technology into its existing service with no visible effects on the customer end. This enabled Uber to expand into other international cities, while using local currencies for local rides.

Source: Braintree (n.d.^[31]), *Case Study: Uber*, www.braintreepayments.com/en-fr/learn/braintree-merchants/uber.

Examples include the Program for Monitoring Emerging Diseases (ProMED), established in 1993, which has demonstrated the power of data sharing through networks and the feasibility of designing effective, low-cost global reporting systems. ProMED has also encouraged the development of additional electronic-surveillance data-sharing networks – such as the

Global Public Health Intelligence Network (GPHIN)¹⁶ and HealthMap.¹⁷ Influenza surveillance is another well-developed global surveillance and monitoring systems enabled through data sharing. Established by the World Health Organization in 1948, it has developed over the years into a highly successful global partnership now including 110 collaborating laboratories in 82 countries that constantly monitor locally isolated influenza viruses and provide real-time streams of data on the emergence and spread of different strains.

Besides co-operation, enhancing access to and sharing of data is also seen as a major enabler and even driver of competition. This is particularly true for personal data portability, which, as was previously noted, is expected to increase competition between providers of digital goods and services, such as social networking service providers, and in analogue markets, such as utilities markets (see subsection “Data portability” in Chapter 2). Depending on their relative market power, some firms may therefore view data portability in some cases as beneficial (see Box 3.1 on Uber) and in other cases as contrary to their business interests (see Box 3.2 on Google and Facebook).

Box 3.2. Data portability, competition and co-operation: The case of two platforms

Online platforms such as Facebook and Google have made it possible for users to extract their personal data in machine-readable formats, even prior to the General Data Protection Regulation’s right to data portability (Art. 20). The situation remains challenging, however, when considering the connections, or “friends,” that a user may have on these platforms. As an illustration, porting a user’s entire Facebook identity “construed as both the data that she/he has uploaded and the friends that she/he has acquired” (Becker, 2012^[32]) is still very complex. Looking back at the history of Facebook, Becker (2012^[32]) notes that the company may have used a lack of data portability to strengthen its market position: “In short, a lack of portability might have helped the site to obtain its current share of the market as well as help it to preserve its current market position.”

The competition challenges related to data portability were strongly apparent in a public disagreement between Facebook and Google back in 2010, which prevented the interoperability of two of the largest online platforms. Horizontal interoperability was clearly viewed by the two platforms as contrary to their respective business interests and in particular lack of reciprocity a disincentive to engage in data portability. As Becker (2012^[32]) further explains:

Facebook aided users in identifying additional friends in their social networks by accessing the Gmail contacts application programming interface (API). However, in late 2010, Google altered its terms of use to prevent Facebook from accessing a user’s Gmail contacts, ostensibly in retaliation for Facebook’s failure to reciprocate [...]. Facebook’s response was to implement a two-step workaround that required users to export their Gmail contacts to their computers and then upload them to Facebook.

More recently these platforms have reconsidered their positions. For example, in 2010 Facebook begun to allow its users to download their personal data (including profile information, photos, videos, wall posts, event information, and a list of friends) (Tsotsis, 2010^[33]). Similarly, Google featured a service called “Google Takeout” in 2011 (now called “Download your Data”) that also allowed users to download all their personal data (Willard, 2018^[34]). Most recently, Google announced the Data Transfer Project, a collaborative initiative together with Microsoft, Twitter, and Facebook to promote interoperability standards (Willard, 2018^[34]).

Policy makers and competition authorities are also looking at non-personal data portability as a possible additional remedy to address competition barriers in a B2B context. The Japan Fair Trade Commission, for instance, issued a report on “Data and Competition Policy” in June 2017, which stressed that “it may fall under the antitrust violation (abuse of superior position) if a large enterprise forced a smaller business partner to provide data which is gathered independently” (Japan Fair Trade Commission, 2017_[35]). The report concluded that some practices that unjustly denied access to data could be an antitrust violation classified as “unfair enclosure”.¹⁸

However, regulatory interventions imposing access and sharing to private-sector data for competition purposes would have to be assessed carefully on a case-by-case basis. The risk of abuse of market power would typically depend not only on access to data, but also on other factors. These include the market segment under consideration, in particular its rate of technological change;¹⁹ the data sources used; the degree of detriment to consumer welfare; the potential barriers to entry, including the level of investments required for building comparable data sets; and last but not least, other control points such as APIs and intellectual property rights used sometimes in combination with data. Furthermore, it may also depend on the available means to escape the control of the dominant actor, including in particular the availability of open standards and personal data portability.

Crowdsourcing new insights and user-driven innovation

For data providers, enhanced access and sharing can provide significant economic and social benefits, even when data are made available free of costs. It can for instance enable new strategic partnerships, where organisations agree to share, cross-licence the re-use and mutually enrich their data sets, or where a community emerges that creates additional value that a single organisation would not be able to create.

Data access and sharing, and open data in particular, can be an optimal strategy for organisations “when they recognise that users may be best positioned to create value” (Frischmann, 2012_[23]). Where users are granted access to their personal data through data portability, they can gain “better visibility into their own consumption, often revealing information that can lead to changes in behaviour” (McKinsey Global Institute, 2013_[10]). In its most extreme form, where access is granted to the public through open data, users (including individuals and businesses) are empowered to “provide input to improve the quality of goods and services” (McKinsey Global Institute, 2013_[10]).²⁰

Supporting and engaging a community of data users via enhanced access and sharing can thus be in the genuine interest of a business. This is in particular true in the era of artificial intelligence, where no single organisation can expect to meet unilaterally all application and customer data needs. For Thomson Reuters, for example, it was the following three reasons which led to the decision to engage in data-sharing partnerships via open data:

- **Encouraging a community response to scale – inclusion, specialisation:** This would include i) helping others (customers) to add value to business solutions by lowering barriers to participation and co-operation in data sharing and re-use (in the case of PermID, helping make data sharing and re-use easy and valuable for others); and ii) giving others a reason to include the business in their solutions.
- **Identifying where to collaborate, new ways to compete, how to foster an ecosystem:** This includes i) technical and commercial experimentation (inside and outside of the business); ii) learning by converting capabilities into products (e.g. Thomson Reuters Labs); and iii) using customers as signposts.

- **Building partnerships and collective action, powering user-driven innovation while maximising the option value of investments:** This includes the development of open standards and, in the case of PermID, understanding the key barriers to data sharing and re-use and the role of identity.

Enhanced access and sharing can also maximise the option value of data and related investments (i.e. the value of allowing flexibility in reaping the benefits of the investments by enabling multiple use options) when there is high uncertainty regarding sources of future market value (OECD, 2006^[36]). This is in particular the case where organisations know that users are best placed to create future value. They adopt enhanced access strategies, taking advantage of the increased value of experimentation by users, market selection of the best services and learning over time about user preferences and possible paths for continued development. The advantage for the organisation is that it “maintains flexibility and avoids premature optimisation or lock-in to a particular development path or narrow range of paths” (Frischmann, 2012^[23]).

Increased efficiency across society through data linkage and integration

Enhanced access and sharing is an enabler of increasing returns to scope where data linkage across organisations and sectors is possible. This is because data linkage enables “super-additive” insights, leading to increasing returns to scope. Linking data is a means to contextualise data and is thus a source for insights and value that are greater than the sum of isolated parts (data silos) (OECD, 2015^[1]). The benefits of data linkage within organisations have been described for instance by Newman (2013) in the case of Google: “It’s not just that Google collects data from everyone using its search engine. It also collects data on what they’re interested in writing in their Gmail accounts, what they watch on YouTube, where they are located using data from Google Maps, a whole array of other data from use of Google’s Android phones, and user information supplied from Google’s whole web of online services.”²¹ These diverse data sets enable profiling hardly possible otherwise.

Data linkage across institutions has been recognised as key for monitoring and increasing the efficiency and quality of the health care system (OECD, 2015^[37]). These include the development of health care quality and system performance indicators to measure care co-ordination and outcomes of care pathways and compliance with national health care guidelines; and to produce indicators of health care utilisation and costs, and disease prevalence, by socio-economic status. In the United Kingdom, for example, administrative hospital records have been linked (via unique patient health service number) with a number of cancer screening registries and used to improve how and when cancer is diagnosed (to increase early detection and survival) (Productivity Commission, 2017^[38]).

Data linkage and integration may also be critical for deploying smart applications across sectors, such as for smart cities. The data produced and collected in these cities are created by multiple actors. Key among them are citizens and consumers, innovators and entrepreneurs, governments and utilities, data brokers and platforms, and infrastructure and system operators. Each of these groups is in principle connected to all the others through a digital layer and in multiple possible combinations. The extent to which data can be exchanged and linked among these actors and across systems, as well as the extent to which they can easily be re-used for different purposes, determines the ability to integrate the different types of applications and to enable synergies. Integrating different applications via the IoT, for example, can multiply the systems, machines, devices and services connected via electricity grids and information systems – such as solar cells on roofs, detailed weather forecasts, home heating systems and air conditioning, and supermarket stocks.

There are, however, various reasons why linking data across different data silos and organisations might be challenging. There may be legal, cultural and technical barriers to data access and sharing as described above. Other barriers may be related to skills. As the OECD (2013^[39]) states: “Even though techniques for record linkage are now well developed, and are used by numerous organisations regularly, the capacity with which to carry out successful linkages may be in short supply.”

Some of the “barriers” to data linkage are legitimate, however, since data linkage is not only a source for great insights but may increase the risk of re-identification (see subsection “The violation of privacy, intellectual property rights, and other interests” in Chapter 4). In addition, data aggregation, when leading to one big data set, can create a single point of failure that a digital security threat can exploit to disrupt the availability, integrity, or confidentiality of the data on which economic and social activities rely (see subsection “Digital security risks and confidentiality breaches in particular” in Chapter 4). In many jurisdictions, the separation of linkage and analysis processes is therefore considered as best practice for confidentiality, meaning that those conducting the linkage (often a “trusted third party”) only have access to a set of identifiers, while those analysing the linked data only have access to de-identified data.

References

- Alissa Walker (2016), “Los Angeles Has Invented the Multimodal Navigation App of My Dreams”, <https://gizmodo.com/la-has-invented-the-multimodal-navigation-app-of-my-dre-1756497278> (accessed on 2 March 2018). [24]
- Beagrie, N. and J. Houghton (2016), *The Value and Impact of the European Bioinformatics Institute*, <https://beagrie.com/static/resource/EBI-impact-report.pdf>. [30]
- Beagrie, N. and J. Houghton (2014), *The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres*, JISC, Bristol and London, <http://repository.jisc.ac.uk/5568/>. [27]
- Beagrie, N. and J. Houghton (2013), *The Value and Impact of the Archaeology Data Services*, Joint Information Systems Committee, Bristol and London. [28]
- Beagrie, N. and J. Houghton (2013), *The Value and Impact of the British Atmospheric Data Centre*, Joint Information Systems Committee and the Natural Environment Research Council UK, Bristol and London, http://www.jisc.ac.uk/whatwedo/programmes/di_directions/strategicdirections/badc.aspx. [29]
- Beagrie, N. and J. Houghton (2012), *Economic Evaluation of Research Data Infrastructure (ESDS)*, Economic and Social Research Council, London, <https://esrc.ukri.org/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/>. [26]
- Becker, M. (2012), “Interoperability Case Study: Cloud Computing”, *The Berkman Center for Internet & Society Research Publication Series*, Research Publication No. 2012-11, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=204. [32]

- Boston Children's Hospital (n.d.), *About*, <http://www.diseasedaily.org/about> (accessed on 5 February 2019). [41]
- Braintree (n.d.), *Case Study: Uber*, <http://www.braintreepayments.com/en-fr/learn/braintree-merchants/uber> (accessed on 1 October 2018). [31]
- Corrado, C., C. Hulten and D. Sichel (2009), "Intangible capital and U.S. economic growth", *Review of Income and Wealth* Series 55, No.3, http://www.conference-board.org/pdf_free/IntangibleCapital_USEconomy.pdf. [40]
- Deloitte (2017), *Assessing the value of TfL's open data and digital partnerships*, <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> (accessed on 2 March 2018). [9]
- Deloitte (2013), *Market Assessment of Public Sector Information: A Report to the Department for Business, Innovation and Skills*. [3]
- Department of Industry, Innovation, Science, Research and Tertiary Education (Australia) (2012), *APS200 Project: The Place of Science in Policy Development in the Public Service*, <https://apo.org.au/node/31390>. [42]
- Frischmann, B. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press. [23]
- IDC and Lisbon Council (2018), *Updating the European Data Market Monitoring Tool*, <http://datalandscape.eu/study-reports/first-report-facts-and-figures-updating-european-data-market-monitoring-tool>. [14]
- Japan Fair Trade Commission (2017), "Report of Study Group on Data and Competition Policy", http://www.jftc.go.jp/en/pressreleases/yearly-2017/June/170606_files/170606-4.pdf. [35]
- Johnson, P. et al. (2017), "The Cost(s) of Geospatial Open Data", *Transactions in GIS*, Vol. 21/3, <http://dx.doi.org/10.1111/tgis.12283>. [21]
- Lateral Economics (2014), *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, http://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FN_L.pdf. [11]
- McKinsey Global Institute (2013), *Open data: Unlocking innovation and performance with liquid information*, http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information. [10]
- McKinsey Global Institute (2011), *Big data: The next frontier for innovation, competition and productivity*, http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx. [20]
- Mitsubishi Research Institute (2017), *De-ta ryutuu Purattofo-mu ni kannsuru tyo-sajigyo [Study on platforms for data sharing]*, http://www.meti.go.jp/meti_lib/report/H28FY/000467.pdf. [12]

- Nomura Research Institute (2014), *データの高度な利活用による業務・サービス改革が我が国経済及び社会に与える波及効果に係る調査研究 [Research on Spillover Effects of Evolution in Operation and Services by Data Use on the Economy and Society]*. [13]
- OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-en. [19]
- OECD (2017), “Business models for sustainable research data repositories”, *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, <http://dx.doi.org/10.1787/302b12bb-en>. [25]
- OECD (2015), “Assessing government initiatives on public sector information: A review of the OECD Council Recommendation”, *OECD Digital Economy Papers*, No. 248, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5js04dr9l47j-en>. [8]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [1]
- OECD (2015), *Health Data Governance: Privacy, Monitoring and Research*, (policy brief), OECD, Paris, <https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdf>. [37]
- OECD (2015), *The Innovation Imperative: Contributing to Productivity, Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264239814-en>. [16]
- OECD (2013), “Knowledge networks and markets”, *OECD Science, Technology and Industry Policy Papers*, No. 7, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k44wzw9q5zv-en>. [17]
- OECD (2013), *New Data for Understanding the Human Condition: International Perspectives*, OECD, Paris, <http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf>. [39]
- OECD (2010), *The OECD Innovation Strategy: Getting a Head Start on Tomorrow*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264083479-en>. [15]
- OECD (2006), “Quasi option value”, in *Cost-Benefit Analysis and the Environment: Recent Developments*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264010055-11-en>. [36]
- Office of Fair Trading (2006), *The Commercial Use of Public Information*. [2]
- Onsrud, H. (2007), *Research and theory in advancing spatial data infrastructure concepts*, Esri Press; Redlands, CA. [22]
- Productivity Commission (2017), *Productivity Commission Inquiry Report: Data Availability and Use*, Productivity Commission, <https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf> (accessed on 19 March 2018). [38]

- Staff, P. (2017), *What is the Open Banking Standard? How a global system of banking APIs can help consumers, businesses, and banks*, <https://fin.plaid.com/articles/what-is-the-open-banking-standard> (accessed on 2 March 2018). [18]
- Tasman, A. (2009), *Spatial Information in the New Zealand economy*, <http://cnig.gouv.fr/wp-content/uploads/2014/05/spatial-information-in-the-new-zealand-economy-2009.pdf>. [5]
- Tasman, A. (2008), *The Value of Spatial Information*, <http://www.crcsi.com.au/assets/Resources/7d60411d-0ab9-45bc-8d48-ef8dab5abd4a.pdf>. [4]
- Tsotsis, A. (2010), “Facebook Now Allows You To “Download Your Information””, *TechCrunch*, <https://techcrunch.com/2010/10/06/facebook-now-allows-you-to-download-your-information/>. [33]
- Vickery, G. (2012), *Review of recent studies on PSI reuse and related market developments*, http://www.scb.se/statistik/_publikationer/NR9999_2012A01_BR_X76BR1201.pdf. [7]
- Vickery, G. (2011), *Review of recent studies on PSI reuse and related market developments*, http://ec.europa.eu/information_society/policy/psi/facilitating_reuse/economic_analysis/index_en.htm. [6]
- Willard, B. (2018), “Introducing Data Transfer Project: an open source platform promoting universal data portability”, *Google Open Source*, <https://opensource.googleblog.com/2018/07/introducing-data-transfer-project.html>. [34]

Notes

¹ This property is at the source of significant spill-overs, which provide the major theoretical link to multi-factor productivity growth according to a number of scholars including Corrado, Hulten and Sichel (2009^[40]).

² The OECD (2015^[1]) suggests that data can be considered as the new “research and development” (R&D), i.e. as input to innovation for 21st-century innovation systems. Both, data and R&D share a number of common properties: both are intangible assets that can be combined with other innovation investments like training, software, organisational change, etc.; both enable the creation of knowledge with positive externalities or spill-overs across society; and both face the challenge of these externalities possibly negatively impacting on incentives to invest. Organisations may well be able to capture the private benefits of their investment in data, but do not yet always see the larger benefits that the data can bring to society.

³ See findings of the Copenhagen Workshop (www.oecd.org/internet/ieconomy/expert-workshop-enhanced-access-to-data-reconciling-risks-and-benefits-of-data-re-use.htm) for more details and for a discussion of the economic conditions under which data can be considered an infrastructural resource.

⁴ Current studies significantly differ in terms of the scope of the sectors (e.g. whether the public sector and/or the private sector was included), the types of data (e.g. whether personal, proprietary or public data was included), and the degrees of data openness (and the arrangements included such as open data) as well as the methodologies, including in particular the different level of the impact assessed (i.e. whether the effects were assessed at the organisational, sectoral or macroeconomic level).

⁵ These can be decomposed in the following categories of data: i) economic, business and legal information (GBP 24 million); ii) geographical information (GBP 25 million); iii) environmental and scientific information (GBP 16.5 million); and iv) other information.

⁶ These are based on 2011 data and include around GBP 100 million in revenues generated from sales of PSI, GBP 100 million through supply chain effects from increased jobs and related consumer spending from the production of PSI, and GBP 1.6 billion through consumer surplus from direct use and consumption PSI related products.

⁷ Tasman (2008^[4]) estimates that 4.00% to 5.14% of total factor productivity gain in fisheries are based on the re-use of geospatial data from the public sector, around 1.93% in forestry, 1.40% to 1.53% in road transport, 1.35% to 1.50% in sheep/cattle, 0.98% to 1.32% in communication and 0.93% to 1.08% in agriculture.

⁸ The induced impact of the use and re-use of geospatial data in New Zealand is estimated to generate values worth NZD 1.2 billion in 2009, which also corresponds to 0.6% of GDP (Tasman, 2009^[5]).

⁹ The OECD (2015^[8]) concludes that there could be close to USD 200 billion of additional gains to the indirect benefits if barriers to use of data were removed, skills enhanced and the data infrastructure improved.

¹⁰ Altogether, over 50% of the total potential value of open data is estimated to be generated from consumer and customer surplus (McKinsey Global Institute, 2013^[10]). The largest share of the total benefits of open data is attributed to better benchmarking, “an exercise that exposes variability and also promotes transparency within organisations” (McKinsey Global Institute, 2013^[10]). Better benchmarking would enable “fostering competitiveness by making more information available and creating opportunities to better match supply and demand” as well as “enhancing the accountability of institutions such as governments and businesses [to] raise the quality of decision [making] by giving citizens and consumers more tools to scrutinise business and government” (McKinsey Global Institute, 2013^[10]).

¹¹ The objective of the study is to assess the contribution of data platforms in the increase of data re-use, irrespective of whether public- or private-sector data is used. Different types of data platforms are also included in the study, ranging from data markets (Windows Azure Marketplace, Datamarket and EverySense), market-based data services (Sakura IoT Platform and G Space Information Center) and open data platforms (G Space Information Center).

¹² For the estimation of data market and the data economy, IDC and Lisbon Council (2018^[14]) identifies data companies which consist of data suppliers and data users. Data suppliers have, as their main activity, the production and delivery of digital data-related products, services and technologies, while data users are organisations that generate, exploit, collect and analyse digital data intensively to improve their business activities.

¹³ As Frischmann (2012^[23]) highlights such an approach “is not a direct subsidy to [...] users who produce public or social goods, but it effectively creates cross-subsidies and eliminates the need to rely on either the market or the government to ‘pick winners’ – that is, to prioritise or rank [...] users worthy of access and support”.

¹⁴ Referring to open government data, Johnson et al. (2017^[21]), for instance, note that: “Given the benefits for the private sector in using open data instead of generating or purchasing similar data from other sources [...], this raises questions as to what degree the public sector is subsidising private-sector business models by opening data. [...] The rhetoric of open data, which often refers to it as data for which taxpayers have already paid [...], may obscure the real costs to government of making data open, masking the true value of the subsidy to the private sector.”

¹⁵ An open API can provide both easy access to openly available data (such as a bank's product offerings) and secure shared access to private data (such as a third party's access to a user's transaction history). These APIs would be established by banks and could be integrated with third-party technologies to carry out specific functions related to the banking data. For instance, apps could allow customers to compare banking services to choose what products best suits their needs (Staff, 2017^[18]).

¹⁶ The GPHIN, developed by Health Canada in collaboration with the World Health Organization, is a secure Internet-based multilingual early-warning tool that continuously searches global media sources such as news wires and websites to identify information about disease outbreaks and other events of potential international public health concern. See www.who.int/csr/alertresponse/epidemicintelligence/en/ (accessed 7 May 2015).

¹⁷ HealthMap, developed at Boston Children's Hospital in 2006, uses online informal sources for disease outbreak monitoring and real-time surveillance of emerging public health threats (Boston Children's Hospital, n.d.^[41]).

¹⁸ The extent to which data control can raise competition issues has been discussed in OECD (2015^[1]).

¹⁹ Markets featuring a series of disruptive innovations can lead to patterns in which firms rise to positions of temporary monopoly power but are then displaced by a competitor with superior innovation.

²⁰ For example, as the public sector makes its data available for science and research, new scientific insights can be used as evidence for informing policy makers and regulators. Australia's Department of Industry, Innovation, Science, Research and Tertiary Education, through its APS200 Project on "the place of science in policy development in the public service" has identified a number of practical and useful strategies to maximise the use of science in policy development based in particular on public-sector data (Department of Industry, Innovation, Science, Research and Tertiary Education (Australia), 2012^[42]).

²¹ The "super-additive" nature of linked data is of course not without its challenges as well. In particular, linked datasets can undermine confidentiality and privacy protection measures such as anonymisation and pseudonymisation.

4. Risks and challenges of data access and sharing

This chapter describes the major challenges facing policy makers when enhancing access to and sharing of data. These include balancing its benefits and risks, strengthening users' trust and making it easier for them to share and re-use data, and creating data market incentives and sustainable business models.

Individuals, businesses, and governments face common challenges when data are accessed and shared. Many of these challenges were identified based on discussions at the Copenhagen Expert Workshop, the Joint CSTP-GSF and the Stockholm Open Government workshops.

This chapter provides an overview of the major challenges to be addressed by policy makers to facilitate and encourage enhanced access and sharing. They have been grouped around the following three major issues discussed in the following sections respectively:

1. Balancing the benefits of enhanced data access and sharing with the risks, while considering legitimate private, national, and public interests. This may require reducing unjustified barriers to cross-border data flows.
2. Reinforcing trust and empowering users through pro-active stakeholder engagements and community building to facilitate data sharing and help maximise the value of data re-use. This may involve significant costs including for the development of data-related skills, infrastructures and standards as well as for maintaining community engagement.
3. Encouraging the provision of data through coherent incentive mechanisms and sustainable business models while acknowledging the limitations of (data) markets. This may require addressing uncertainties about data ownership and clarification of the role of privacy, intellectual property rights (IPRs) and other ownership-like rights, which ideally should be undertaken by appropriate expert agency and organisations.

These issues are interrelated. For instance, trust can be reinforced by empowering users so that they can address some of the risks of enhanced access and sharing. And private and public interests need to be reflected in incentive mechanisms to assure the coherence of these mechanisms.

When addressing these policy issues, policy makers need to avoid the “data policy pitfall”, which, according to discussions at the Copenhagen Expert Workshop, is the tendency to look for one silver-bullet solution to a multidimensional problem. Flexible data-governance frameworks that take due account of the different types of data and the different context of their re-use, while doing justice to domain and cultural specificities, are crucial.

Need for balancing the benefits of data “openness” with other legitimate interests, policy objectives and risks

As described in Chapter 3, enhancing access to and sharing of data (EASD) can provide social and economic benefits and support good public governance. However, data access and sharing also comes with several risks to individuals and organisations. These include the risks of confidentiality and privacy breaches and the violation of other legitimate private interests, such as commercial interests.

The pursuit of the benefits of EASD therefore needs to be balanced against the costs and the legitimate national, public and private interests, in particular the rights and interests of the stakeholders involved (the protection of their privacy, IPRs, and national security). This is especially the case where sensitive data are involved.

Privacy and IPRs and other legitimate commercial and non-commercial interests need to be protected, otherwise incentives to contribute data and to invest in data-driven innovation may be undermined, in addition to the risks of direct and indirect harm to right holders, including data subjects. Evidence confirms that risks of confidentiality breach, for instance, have led users to be more reluctant to share their data, including providing personal data, and in some cases to use digital services at all.¹ Where multiple right holders may be affected simultaneously, as in the case of large-scale personal data breaches, the scale and scope of the potential impact can become a systemic risk with detrimental effects for society.

**Box 4.1. Balancing the benefits with the risks:
Australia's data-sharing and release legislation**

On 1 May 2018, in response to recommendations in the Productivity Commission's "Data Availability and Use Inquiry" report, the Australian government committed to reforming its national data-governance framework with the development of new DS&R legislation. The DS&R legislation aims to: i) promote better sharing of public-sector data; ii) build trust in the use of this data; iii) establish consistent and appropriate data safeguards that dial up or down depending on the sensitivity of the data; iv) enhance the integrity of the data system; and v) establish institutional arrangements.

The DS&R legislation is based on the recognition that greater sharing of data can lead to more efficient and effective government services for citizens, better informed government programmes and policies, greater transparency around government activities and spending, economic growth from innovative data use, and research solutions to current and emerging social, environmental, and economic issues. In order to balance these benefits with the risks and enhance trust in data sharing and re-use, the issues paper on the DS&R legislation (Department of the Prime Minister and Cabinet [Australia], 2018^[1]) proposes a number of institutional arrangements, including:

The **Office of the National Data Commissioner (NDC)** will provide oversight and regulation of the new data-sharing and release framework, including monitoring and reporting on the operation of the framework and enforcing accompanying legislation. The NDC will also be responsible for the criteria and process for accreditation. This includes the accreditation of "trusted users" and "Accredited Data Authorities (ADAs)":

- Trusted users are the end users of data shared or released by data custodians. To streamline data-sharing arrangements, trusted users would be accredited by demonstrating they can safely use and handle data under the requirements of the DS&R Bill. Trusted users who are employees or contractors of government entities and companies, as well as external users, would be able to seek accreditation in line with the scope of the DS&R Bill.
- ADAs are entities that have strong experience in data curation, collation, linkage, de-identification, sharing and release (for example Australian Bureau of Statistics [ABS], Australian Institute of Health and Welfare) (see the paragraph on the Data Integration Partnership for Australia [DIPA] below). The National Data Advisory Council (NDAC) will advise the NDC on ethical data use, community expectations, technical best practice, and industry and international developments. NDAC will help find the right balance between streamlining the sharing and release of data and ensuring the protection of privacy and confidentiality.

The **DIPA** is a co-ordinated, Australian Public Service-wide investment to maximise the use and value of the government's vast data, allowing cost-effective and timely insights into data that is already available, while ensuring the safe use of data in secure and controlled environments. AUD 130 million over three years will be invested to build a robust, secure, and scalable whole-of-government data integration, policy analysis and evaluation capability. A core component of the DIPA is to establish a central analytics "hub" and issue-specific data analytics units that can integrate and link data assets to solve complex policy issues that cross over multiple portfolios.

Some OECD countries have put in place institutional arrangements to balance the risks and benefits of enhanced access and sharing with other legitimate interests and policy objectives. Australia's data-sharing and release legislation (DS&R legislation) is one example (Box 4.1). The OECD (2016^[2]) *Recommendation of the Council on Health Data Governance* (Recommendation on Health Data Governance) also provides an example of how to address the risks of data sharing and re-use in health care. It calls upon countries to develop and implement health data-governance frameworks that secure privacy while enabling health data uses that are in the public interest.²

Digital security risks and confidentiality breaches in particular

The following subsections describe the increased digital security risks faced by individuals and organisations, before focussing on personal data breaches – more precisely, the breach of the confidentiality of personal data as a result of malicious activities or accidents.

Digital security risks of more data openness

Enhanced access and sharing typically requires opening information systems so that data can be accessed and shared. This may further expose parts of an organisation to digital security threats that can lead to incidents that disrupt the availability, integrity or confidentiality of data and information systems on which economic and social activities rely. Consequently, organisations' assets, reputation and even physical activities can be affected to a point where their competitiveness and ability to innovate are undermined. More importantly, where data is shared among suppliers and customers, these incidents may have a negative impact along an entire supply chain. If critical information systems are concerned, they could undermine the functioning of essential services.

The risk of digital security incidents is growing with the intensity of data use (OECD, 2017^[3]). The actual proportion of the impact varies significantly, depending on the motivation and form of the incidents. Organised crime groups may target valuable assets that they can sell on illegal markets. And as innovation becomes more digital, industrial digital espionage is also likely to further rise. In some cases, the motive may be political, or the attacks may be designed to damage an organisation or an economy (OECD, 2017^[3]).³

Increasing impact of (personal) data breaches

Where data can be accessed and is shared, personal data breaches⁴ are more likely to occur. They will not only cause harm because of the privacy violation of the individuals whose personal data have been breached. They can also cause significant economic losses to the business affected, including loss of competitiveness and reputation. In addition, further consumer detriment may result from a data breach, such as harm caused by identity theft.⁵

Personal data breaches are less frequently experienced compared to other types of digital security incidents, such as malware, phishing and social engineering, or denial of service (DoS)⁶ attacks. However, evidence from Privacy Rights Clearinghouse suggests that although the total number of identified incidents may be relatively small compared to other security incident types, their impact is increasing drastically as large-scale data breaches, i.e. data breaches involving more than 10 million records, become more frequent. This is confirmed by available evidence suggesting that data breaches have increased with the collection, processing and sharing of large volumes of personal data (OECD, 2017^[3]).

In 2005, for example, ChoicePoint, a consumer data aggregation company, was the target of one of the first high-profile data breaches involving over 150 000 personal records.⁷ The

company paid more than USD 26 million in fees and fines. Data breaches have since become almost commonplace. In October 2018, Facebook was fined GBP 500 000, the maximum fine possible by the Information Commissioner's Office (ICO) of the United Kingdom, for "unfairly process[ing] personal data" and "fail[ing] to take appropriate technical and organisational measures against unauthorised or unlawful processing of personal data" (Information Commissioner's Office, 2018^[4]).⁸ This incident involved more than 87 million personal records that were used by Cambridge Analytica (Granville, 2018^[5]; Cadwalladr and Graham-Harrison, 2018^[6]; Hern and Pegg, 2018^[7]). Data breaches are not limited to the private sector, as evidenced by the theft in 2015 of over 21 million records stored by the US Office of Personnel Management, including 5.6 million fingerprints, and by the Japanese Pension Service breach that affected 1.25 million people (Otaka, 2015^[8]).

The violation of privacy, intellectual property rights and other interests

The risks of enhanced access and sharing go beyond digital security and personal data breaches. They include most notably risks of violating contractual and socially agreed terms of data re-use, and thus risks of acting against the reasonable expectations of users. This is true in respect to individuals (data subjects), their consent and their privacy expectations, but also in respect to organisations and their contractual agreements with third parties and the protection of their commercial interests.

In the case of organisations, these risks can negatively affect incentives to invest and innovate. This is true even in cases where these risks may be the unintended consequences of business decisions. For small and medium-sized enterprises (SMEs), identifying which data to share and defining the scope and conditions for access and re-use is perceived as a major challenge. Inappropriate sharing of data can lead to significant costs to the organisation, including fines due to privacy violations and opportunity costs due to a lower ability to innovate. For example, it has been noted that sharing data too prematurely can undermine the ability to obtain IPR (e.g. patent and trade secret) protection.

The following subsections focus on i) the risks of violation of agreed terms of data re-use, which goes hand in hand with ii) the increasing loss of control of individuals and organisations over their data, and iii) the increasing limitations and costs of anonymisation through stronger capabilities to infer information not intended to be shared.

Violations of agreed terms and of expectations in data re-use

Even where individuals and organisations agree on and consent to specific terms for data sharing and data re-use, including the purposes for which the data should be re-used, there remains a significant level of risk that a third party may intentionally or unintentionally use the data differently. The case of Cambridge Analytica discussed above illustrates this risk: personal data of Facebook users was used, not for academic purposes as some users had consented to, but for a commercially motivated political campaign, and this although Facebook explicitly prohibits data to be sold or transferred "to any ad network, data broker or other advertising or monetisation-related service" (Granville, 2018^[5]).⁹

The case of Cambridge Analytica is just one of many occurrences where data are re-used in a different context in violation with agreed terms and conditions. The violation of these terms may not always be the result of malicious intentions. As shown at the Copenhagen Expert Workshop, data access and sharing are about taking data from one context and transferring it to another context. Referring to Nissenbaum (2004^[9]) on privacy as contextual integrity, experts have argued that the change of context made it challenging to ensure that existing rights and obligations were not undermined, for instance, when privacy

assumptions and expectations that were implicit in the initial usage no longer applied in subsequent uses. This is coherent with the observation above that information derived from data is context dependent and so are thus the risks associated to data re-use.

Some of these concerns have been framed as ethical, to underscore the need to recognise the importance of issues such as fairness, respect for human dignity, autonomy, self-determination, the risk of bias and discrimination, in guiding policy on enhanced access and sharing and as complementary to regulatory action. Data ethics is highlighted in particular in cases where the collection and processing of personal data will be legal under privacy law, but may generate moral, cultural and social concerns with potential direct or indirect adverse impacts on individuals or social groups. This could be for example the case when risk profiles are created on the basis of biased and opaque data collections and used for decision-making (see Box 4.2 and Box 4.3, which focus on data ethics in Denmark and the United Kingdom respectively).¹⁰

Recognising and responding to the ethical dimension of research is a fundamental part of the research governance process (OECD, 2016^[10]). In the particular context of research in health care (OECD, 2015^[11]), data ethics has been highlighted as a complementary means to serve and balance the interests of both individuals and societies. Central to an ethical approach to research is understanding how different groups define and value the public benefits of better health data use. So, too, is making sense of where an acceptable balance between risks and benefits may lie. Ethics may provide an additional promising venue in light of the major challenges affecting data access, sharing and re-use discussed further below: i) the loss of control over data and the role of consent; and ii) the increasing limitations and costs of anonymisation.

Box 4.2. Data ethics: Government initiatives in Denmark

In April 2018, the Danish Government established an Expert Group on Data Ethics with the objective of developing recommendations for ensuring citizens trust in the digital economy. The expert group was comprised of high-level representatives from large, medium and small companies. Its objective was to create recommendations that would safeguard consumer trust without creating unnecessary burdens for companies or stifle innovation. The overall *raison d'être* for the group was that data ethics should become a competitive advantage rather than a barrier for Danish and European companies in the global marketplace.

The Expert Group on Data Ethics handed over nine recommendations to the Danish government in November 2018 (The Expert Group on Data Ethics, 2018^[12]). The Danish government evaluated the group's findings and translated them into concrete policy proposals and priorities (Ministry of Industry, Business and Financial Affairs [Denmark], 2019^[13]), such as potential legal provisions that would require statements regarding data ethics policies to be included in annual reports. The Danish government is also co-operating with industry bodies to explore the possibility of creating a “national seal” for digital security and responsible data use that will increase transparency and make it easier for consumers to choose companies, products and solutions that live up to certain data security and ethics standards. In the longer term, these initiatives could be lifted to the international level, since most challenges relating to data is by its nature international.

Box 4.3. Data ethics: Government initiatives in the United Kingdom

To assure that public servants from across disciplines understand insights from data and emerging technologies and use data-informed insight responsibly, the UK government has developed a Data Ethics Framework (Department for Digital, Culture, Media and Sport (UK), 2018^[14]). This Framework sets out principles and practical advice for using data, including building and procuring advanced analytics software for designing and implementing policies and services. The framework is aimed broadly at anyone working directly or indirectly with data in the public sector, including data practitioners (statisticians, analysts and data scientists), policymakers, operational staff and those helping produce data-informed insights. The framework includes a Data Ethics Workbook with questions to probe ethical, information assurance and methodological considerations when building or buying new technology.

As part of its Industrial Strategy and the AI Sector Deal, which commits almost GBP 1 billion to support the artificial intelligence (AI) sector, the UK government has also established its new Centre for Data Ethics and Innovation (Department for Digital, Culture, Media and Sport (UK), 2018^[15]). The Centre will identify the measures needed to strengthen and improve the way data and AI are used and regulated. This will include articulating best practice and advising on how we address potential gaps in regulation. The Centre's role will be to help ensure that those who govern and regulate the use of data across sectors do so effectively. By ensuring data and AI are used ethically, the Centre will promote trust in these technologies, which will in turn help to drive the growth of responsible innovation and strengthen the United Kingdom's position as one of the most trusted places in the world for data-driven businesses to invest in.

Loss of control over data and the role of consent

Once data are accessed or shared, unless specific data stewardship and processing provisions are in place, that data will move outside the information system of the original data holder (data controller) and thus out of his/her control. The same is true for individuals who provide their data and give their consent for their re-use and sharing. Data holders and individuals then lose their capabilities to control how their data are re-used. To object to or oppose such uses, they must rely solely on law enforcement and redress. The risks of loss of control are multiplied where the data are further shared downstream across multiple tiers, in particular when these tiers are located across multiple jurisdictions.¹¹

Lack of control over data is perceived as a major issue for both organisations and individuals. Some SMEs, for instance, have not only refrained from engaging in data sharing, but have even avoided using certain digital technologies such as cloud computing out of concerns of losing control over their data (OECD, 2017^[3]).¹² Similar concerns have been expressed by individuals. According to a 2014 Pew Research Centre poll, 91% of Americans surveyed agreed that consumers had lost control of their personal information and data (Madden, 2014^[16]).¹³ Similarly, in the European Union, “two-thirds of respondents (67%) are concerned about not having complete control over the information they provide online” (European Commission, 2015^[17]). Meanwhile, “roughly seven out of ten people are concerned about their data being used for a different purpose from the one it was collected for”.

Consent has been highlighted as a major mechanism to allow individuals to control the collection and (re-)use of their personal data. It requires clear provision of information to individuals about what personal data are being collected and used, and for what purpose –

as specified in the data protection and privacy laws of most countries and in the OECD *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (hereafter the “OECD Privacy Guidelines”) (OECD, 2013^[18]). Within the medical/scientific field, informed consent generally presumes the ability to indicate clearly to the participant the use and purpose of the particular research activity. However, within the digital environment, data are resources that can be used and re-used, often in ways that were inconceivable at the time the data was collected.

To assure the maximum level of flexibility in compliance with privacy legislations, some organisations have thus come to rely on one-time general or broad consent as the basis for data collection, use and sharing. One-time general consent respects the wishes of people to control the use of their data without mandating that they decide the specific projects for which the data are used. It can be used to achieve an appropriate balance between participant rights to determine the future use of their personal data and the social benefits that may accrue when such use involves unspecified investigators and research aims. Data subjects must be given “reasonable means to extend or withdraw their consent over time” (OECD, 2015^[11]). Broad consent is still subject to the provision of details of the nature, storage, maintenance, and future uses of an individual’s identifiable data. However, these practices have been criticised for posing ethical challenges as data subjects may not realise the full implications of giving a broad consent, particularly in the context of AI and big data. As highlighted in Chapter 2 (subsection on “The manner data originates: Reflecting the contribution to data creation”), it is more and more the case that individuals cannot be fully aware of how the observed, derived, inferred personal data about them can be used and shared between data controllers and third parties. They also cannot predict if their data will be used for purposes that may transgress their moral values.

New consent models have thus been proposed in the scientific literature, including “adaptive” or “dynamic” forms of consent (OECD, 2015^[11]). These also include time-restricted consent models, where individuals consent to the use of their personal data only for a limited period. These models typically enable participants to consent to new projects or to alter their consent choices in real time as their circumstances change and to have confidence that these changed choices will take effect (Kaye et al., 2015^[19]). Where obtaining informed consent is impossible or impracticable, however, lawful alternatives consistent with privacy legislation may apply. These in turn typically require the involvement of “designated authorities, such as independent research ethics committees, advisory or institutional research boards and the outcomes of their decisions often require sponsor or public access” (OECD, 2015^[11]).

Limitations of anonymisation and the increasing power of data analytics

The use of anonymisation and similar techniques such as aggregation is often proposed as means of protection in some cases. As described in Chapter 2 (subsection on “Personal data and the degrees of identifiability: Reflecting the risk of harm”), the degree of anonymisation can determine the extent to which legal and technical protection may be necessary and the level of access control required. The less data can be linked to an identity (of an individual or an organisation), because it is effectively anonymised and sufficiently aggregated for example, the more it is expected that the data can be freely shared and re-used.

However, developments in data analytics (and AI) combined with the increasing volume and variety of available data sets, and the capacity to link these different data sets, have made it easier to infer and relate seemingly non-personal or anonymised data to an identified or identifiable entity, even if the entity never directly shared this information with anyone (OECD, 2015^[20]). Once linked with sufficient other information, the likelihood that an individual

will possess certain characteristics can be predicted to build a profile. The inferences may not be accurate, but even where correct, there remains a risk that they could be used against an individual's best interests, wishes or expectations, as highlighted in Chapter 2 (subsection "The manner data originates: Reflecting the contribution to data creation").

All this seriously limits the use of anonymisation as means of protection, in particular if the population size is sufficiently small. When combined with enforceable commitments to not re-identify, anonymisation may still have considerable value, even if there are no fail-proof (technical) guarantees of privacy protection. The situation is exacerbated in cases where anonymised data are considered out of the scope of privacy protection legislation.

The use of anonymisation thus needs to be complemented by other data-governance mechanisms that provide added security to protect de-identified information. These may include independent review bodies that evaluate data use proposals for public benefits and adequacy of data security; contractual agreements that bind data receivers to required data security and disclosure practices; and security audits and follow-up mechanisms to ensure compliance with these contractual obligations; as well as the use of data sandboxes discussed in Chapter 2 (subsection "Data-access control mechanisms: Protecting the interests of data holders").¹⁴ Recent developments in privacy preserving technologies such as distributed machine learning and homomorphic encryption (i.e. encryption that allows processing of encrypted data without revealing its embedded information) could also help protect identifiable information.

The difficulty of applying a risk management approach

Risk management has become a widely accepted practice that is conducted by many types of organisations. It has become, for instance, the recommended paradigm for addressing challenges related to digital security risks as reflected in the OECD (2015^[21]) *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity* (hereafter the "OECD Recommendation on Digital Security Risk Management").

The overall objective of risk management is to facilitate decision-making by taking into account the effect of uncertainty on the organisation's objectives and thereby increase the likelihood of success. Depending on the context and the nature of the organisation, these objectives may be expressed in legal, financial, social or other terms.

Risk management assumes that there is always some level of risk associated with carrying out an activity and that risk is not a binary concept. Activities cannot be simply characterised as "risky" or "risk-free". Rather, the goal of risk management is to reduce the risk to a level that is acceptable in light of the potential benefits and taking context (e.g. values, mission, etc.) into account.

The following two subsections discuss the extent to which risk management approaches can help address the issues highlighted above, in particular digital security risks and the violations of private interests including in particular privacy.

Low adoption of digital risk management practices in organisations

Despite the recognition that digital security issues should be addressed through a risk-based approach, many stakeholders continue to adopt an approach that leverages nearly exclusively technological solutions to create a secure digital environment or perimeter to protect data. However, this approach would likely close the digital environment and stifle the innovation enabled by enhanced access and sharing, which relies on a high degree of data openness, including with a potentially unlimited number of partners outside the perimeter.

A more effective approach would consider digital security risk management and privacy protection as an integral part of the decision-making process rather than separate technical or legal constraints. As called for in the OECD Recommendation on Digital Security Risk Management, decision makers would need to work in co-operation with security and privacy experts to assess the digital security and privacy risk related to opening their data. This would enable them to assess which types of data should be opened and to what degree, in which context and how, considering the potential economic and social benefits and risks for all stakeholders.

However, applying risk management to digital security and other digital risks is still challenging for most organisations, in particular where the rights of third parties are involved (e.g. the privacy rights of individuals and the IPRs of organisation and individuals). The share of organisations with effective risk management approaches to security still remains much too low, although there are significant variations across countries and by firm size.¹⁵ A number of obstacles preventing the effective use of risk management for addressing trust issues have been identified, the biggest one being insufficient budget and a lack of qualified personnel (OECD, 2017^[3]) as further discussed in the subsection “Capacity building: Fostering data-related infrastructures and skills” below.

Challenges of managing the risks to third parties

Applying a risk-based approach for the protection of the rights and interests of third parties, in particular with respect to the privacy rights of individuals and the IPRs of organisations, is more complex. In the context of privacy protection, the need for a risk-based approach is increasingly being recognised. The OECD Privacy Guidelines, for instance, recommend taking a risk-based approach to implementing privacy principles and enhancing privacy protection. Risk management frameworks such as the Privacy Risk Management Framework proposed by the US National Institute of Standards and Technology (2017^[22]) are being developed to help organisations apply a risk management approach to privacy protection. In the specific context of national statistics, frameworks such as the Five Safes Framework have been used for balancing the risks and the benefits of data access and sharing (Box 4.4).

Most initiatives to date tend to see privacy risk management as a means of avoiding or minimising the impact of privacy harms, rather than as a means of managing uncertainty to help achieve specific objectives. Focussing on harm is challenging because, unlike in other areas where risk management is widely used, such as health and safety regulation, there is no general agreement on how to categorise or rate privacy harms, i.e., on the outcomes one is trying to avoid. Also, many organisations still tend to approach privacy solely as a legal compliance issue. Organisations often tend to not recognise the distinction between privacy and security risk, even when privacy risk may be unrelated to security, for example when personal data is processed by the organisation in a manner that infringes on individuals’ rights. This is consistent with findings by a study of business practice in Canada funded by Canada’s Office of the Privacy Commissioner, which notes that privacy risk management is much talked about but poorly developed in practice (Greenaway, Zabolotniuk and Levin, 2012^[23]).¹⁶

Where the rights and interests of third parties are involved (e.g. the privacy rights of individuals and the IPRs of organisation and individuals), applying risk management typically requires defining the proportionate level of risk acceptable to all relevant stakeholders and treating the risk accordingly based on a full risk assessment. However, how to allocate responsibility and how to define the acceptable level of risk, when the rights of third parties may be affected, can be challenging to implement (see section below “Trust and empowerment for the effective re-use of data across society” below).

Box 4.4. Managing the risk of disclosure: The Five Safes Framework

Managing disclosure risk involves assessing not only the data itself, but also the context in which the data are released. Once the context is clearly understood, it is much easier to determine how to protect against the threat of disclosure.

The Five Safes Framework provides a structure for assessing and managing disclosure risk that is appropriate to the intended data use. This framework has been adopted by the ABS, several other Australian government agencies and statistical organisations such as the Office of National Statistics (United Kingdom) and Statistics New Zealand.

The Five Safes Framework takes a multidimensional approach to managing disclosure risk. Each “safe” refers to an independent but related aspect of disclosure risk. The framework poses specific questions to help assess and describe each risk aspect (or safe) in a qualitative way. This allows data custodians to place appropriate controls, not just on the data itself, but on the manner in which data are accessed. The framework is designed to facilitate safe data release and prevent over-regulation. The five elements of the framework are:

- Safe People: Is the researcher appropriately authorised to access and use the data?
- Safe Projects: Is the data to be used for an appropriate purpose?
- Safe Settings: Does the access environment prevent unauthorised use?
- Safe Data: Has appropriate and sufficient protection been applied to the data?
- Safe Outputs: Are the statistical results non-disclosive?

Source: Australian Bureau of Statistics (2017^[24]), *Managing the Risk of Disclosure: The Five Safes Framework*, www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017.

Barriers to cross-border data access and sharing

Individuals and organisations rely more than ever on data collected, stored, processed and transferred from other entities, often located abroad. A significant share of the global volume of data and its processing will rarely be located within just one national border. They will instead be distributed around the globe, reflecting the global distribution of economic and social online activities. Transborder data flows are not only a condition for information and knowledge exchange, but also a vital condition for the functioning of globally distributed data markets and societies. In addition, transborder data flows can also facilitate collaboration between governments to improve their policy making at international levels and to address global challenges as defined by the United Nations’ Sustainable Development Goals.

As stated already in the OECD (1985^[25]) *Declaration on Transborder Data Flows*, “these flows acquire an international dimension, known as Transborder Data Flows”, and enable trade between countries and global competition among actors in data markets, and they can help strengthen collective commitment and efforts across borders to support greater public-sector transparency, reduce corruption and contribute to economic growth as highlighted in the 2015 *G20 Open Data Principles for Anti-Corruption* (G20, 2015^[26]). Restrictions to cross-border data flows could therefore restrict the functioning of markets and the prosperity of societies by restricting the benefits of sharing and re-use of data, information and knowledge across countries (OECD, 2015^[20]).

Transparency in regulations affecting cross-border data access and sharing

Concerns have been expressed about restrictions of cross-border data flows such as data localisation requirements (including for data other than personal data), which force organisations to restrict data access, sharing and re-use within national borders. During the Group of Seven (G7) ICT Ministerial Meeting in Takamatsu, Japan, on 29-30 April 2016, ministers agreed “except for cases with legitimate public policy objectives, [...] to oppose data localisation requirements that are likely to hinder the free flow of information” (G7, 2016^[27]). This is in line with the OECD Privacy Guidelines (OECD, 2013^[18]), which recommend that “any restrictions to transborder data flows of personal data should be proportionate to the risks presented, taking into account the sensitivity of the data, and the purpose and context of the processing”.

The legal grounds of these restrictions can vary across countries and are not limited to privacy protection. Digital security and the protection of IPRs, as well as national security and law enforcement, are grounds for restricting cross-border data flows. The extent to which these restrictions are proportionate to the risks requires a case-by-case assessment that includes human rights and the rule of law, transparency, fair process, and accountability as articulated in the OECD (2011^[28]) *Recommendation of the Council on Principles for Internet Policy Making*.

Discrimination between foreign and domestic entities

Data localisation requirements have raised concerns where “they create situations, not only de jure but also de facto, where foreign services and services suppliers are treated less favourably than domestic firms”. The extent to which countries make distinctions between foreign and domestic entities can vary significantly between sectors. In the context of health data, for instance, OECD (2015, p. 87^[29]) notes that:

Some countries make no distinction between foreign and domestic applicants for secondary data use, subjecting both to the same set of rules. Nonetheless, many countries are reticent to approve foreign applications for access to data, due to the inability to impose sanctions on a foreign entity for non-compliance with legal requirements or with the requirements within their data sharing agreement. Some countries will not consider any foreign applications; some will consider only applications for access to de-identified personal health data; while others will consider the approval of the sharing of identifiable personal health data if there is a strong justification for the project.

Towards the interoperability of existing legal and regulatory frameworks

Lack of common approaches and rules for sharing data across countries, in particular personal and other confidential data, has limited the ability of cross-border data access and sharing. This remains an issue despite the wide recognition for the need for international arrangements and legal interoperability as articulated in the Principle on “International access and use” of the OECD *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information* (hereafter the “OECD PSI Recommendation”) (OECD, 2008^[30]). The Principle calls for “seeking greater consistency in access regimes and administration to facilitate cross-border use and implementing other measures to improve cross-border interoperability, including in situations where there have been restrictions on non-public users. [...]”.

Progress is being made in the field of privacy protection. In June 2016, Ministers and Representatives of 42 countries plus the European Union¹⁷ agreed in the OECD (2016^[31]) *Ministerial Declaration on the Digital Economy* (Cancún Declaration) to share experiences

and work collaboratively to “support the development of international arrangements that promote effective privacy and data protection across jurisdictions, including through interoperability among frameworks”. In April 2017, the Group of Twenty (G20) Ministers¹⁸ responsible for the digital economy also recognised in the *G20 Digital Economy Ministerial Declaration* “the importance of promoting interoperability between privacy frameworks of different countries” (G20, 2017^[32]).

Countries have engaged in several initiatives to enhance global interoperability through co-ordination, harmonisation, and interoperability of privacy frameworks. For example, many countries actively participate in international fora (such as the Global Privacy Enforcement Network [GPEN] and Asia-Pacific Economic Cooperation [APEC]) and increasingly engage in bilateral agreements (such as the EU-Japan Economic Partnership Agreement on the mutual recognition of an equivalent level of privacy protection by the European Union and Japan).

Trust and empowerment for the effective re-use of data across society

Trust plays an essential role in data access, sharing and re-use across organisations, sectors, and countries. It can be abused or erode over time, and restoring it can be challenging.

This section discusses means through which trust can be enhanced, which goes hand in hand with the empowerment of users. After i) discussing the role of communities, including the communities of data users and of holders, this section highlights the enabling factors needed for the effective re-use of data across society. These include ii) data-related skills and infrastructures; and iii) data-related standards.

Supporting and engaging communities of stakeholders

Supporting the creation of communities of stakeholders (data users, data holders and third parties) around data sharing and re-use is considered a major success factor for building trust. Active community engagement can help allocate responsibilities and define the acceptable risk levels.

The way these communities are structured and governed varies according to the degree of data openness as well as the expected potential value to be derived from data re-use. The composition and heterogeneity can be leveraged for more differentiated approaches to data access and sharing and a more effective management of the associated risks and incentives mechanisms (e.g. accreditation/certification of data users, providers, and intermediaries). However, they also reflect the complexity and the opposing interests that policy makers need to manage when developing data-governance frameworks.

Partnerships can encourage and help sustain data sharing between data holders, including between the public and private sectors, and therefore deserve special attention by policy makers. However, fostering and maintaining communities’ engagement can involve significant costs, in particular where there are opposing interests and expectations to be considered and reconciled.

The costs of facilitating and engaging communities of stakeholders

Case studies discussed at the Copenhagen Expert Workshop and the Stockholm Open Government Workshops confirmed that even when made available through open access, there was no guarantee that data would be re-used effectively. Further measures (in addition to technical measures such as the development and maintenance of application programming interfaces [APIs]) are often needed for effectively engaging users. According to Deloitte (2017^[33]), for instance, the costs for Transport for London of publishing open data was estimated to be around GBP 1 million per year. A significant part of the costs was related to maintenance and facilitation of communities of data users.¹⁹

Hackathons are also often used to engage communities of data users. These are special events in which software developers and data scientists are involved in the development of applications in a competitive and collaborative manner. Examples include Challenge.gov in the United States, “a listing of challenge and prize competitions, all of which are run by more than 100 agencies across federal government” (US General Services Administration, 2018^[34]), and the European Big Data Hackathon, an event organised by the European Commission and Eurostat gathering teams from all over Europe to compete for the best data product combining official statistics and big data to support policy makers in pressing policy questions facing Europe (European Commission, 2018^[35]).

Anti-competitive data-sharing agreements (collusion)

Where competitors are involved there is a risk that data partnerships and the use of trusted third parties could lead to implicit collusion between businesses, i.e. agreements that would limit open competition by e.g. fixing prices. This would be in particular the case when data on competition-relevant information such as on production capacity would be shared in a rather closed environment (see subsection “Other restricted data-sharing arrangements” in Chapter 2) (OECD, 2010^[36]).²⁰ But even if this is not the case, data sharing among competitors may still enable “more insidious ways of collusive co-operation” that could result in anti-competitive effects. For example, data sharing could facilitate collusion among competitors by “allowing them to establish coordination, monitor adherence to coordinated behaviour and effectively punish any deviations” (OECD, 2010^[36]).

To deal with these risks, the use of “safe harbours” have been proposed as means for increasing legal certainty and for providing predictability for businesses (OECD, 2010^[36]). These arrangements are based on several possible factors that must be considered to mitigate the risk of anti-competitive data sharing. These include, for example, the market shares of the parties, the characteristics of the information exchanged, and the nature of the sector affected. OECD (2010^[36]) also calls for a careful assessment of the legality of data sharing within the context of competition law prohibitions against cartels.

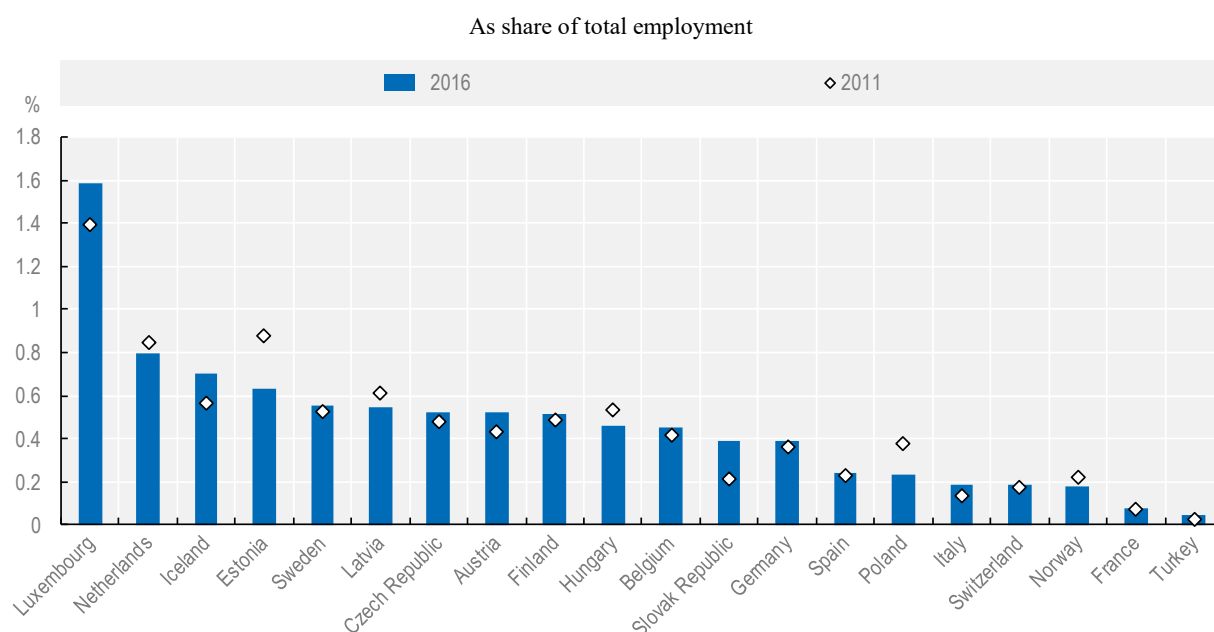
Capacity building: Fostering data-related infrastructures and skills

At the Copenhagen Expert Workshop, concerns have been expressed that enhanced access and sharing (including data portability and open data) may benefit primarily data-savvy firms and individuals (Department for Business Innovation and Skills (UK), 2012^[37]), exacerbating existing inequalities.

Lack of data-related skills and competences and poor access to computation and storage capacities can become bottlenecks preventing the effective re-use of data, even where data are made available through enhanced access and sharing. Fostering data-related infrastructures and skills would not only help assure that all can benefit from enhanced access and sharing. It would also help maximise the re-use and thus the value of data.

Data-related skills and competences

There is growing evidence that the demand for data specialist skills exceeds the supply on the labour market (OECD, 2015^[20]). According to the Harvey Nash/KPMG CIO Survey, “big data and analytics” are top of the list of critical skills shortages (Rae, 2018^[38]). Estimates suggest that data specialists in 2013 accounted for less than 1% of total employment in OECD countries, with Luxembourg, the Netherlands, the United States, Australia, and Estonia reporting the highest share (Figure 4.1).

Figure 4.1. Data specialists in selected OECD countries

Notes: Data specialists are defined by ISCO-08 codes 212 “Mathematicians, actuaries and statisticians” and 252 “Database and network professionals”. Countries for which detailed data were unreliable or not available were not included. For Luxembourg data are 2015 instead of 2011. For France and Turkey data are 2014 instead of 2011. For the Netherlands data are 2013 instead of 2011. For Germany data are 2012 instead of 2011.

Source: OECD based on EU Labour Force Survey, November 2017.

Lack of data-related skills is an issue across all sectors and may prevent the effective re-use of data, even if made available via open access. Available evidence from open government data initiatives, for instance, shows that “open data literacy programmes” are essential for the pro-active engagement of all stakeholders. As Johnson et al. (2017^[39]) note:

Through its focus on data, the open data literature reinforces a perspective that, for example, mapping tools are ubiquitous and easy-to-use. In reality, even the simplest visualisation and analysis tools like Google Earth may not be simple enough for a broad section of society to use [...] These data literacy issues pose a challenge in using open data as a vector for civic participation. [...] Scholars have noted that placing data in an open data catalogue is no guarantee of its use [...].

There is also evidence of poor levels of skills and competences to manage, create, curate and re-use data in the scientific community. Data stewards sometimes even lack skills to apply the relevant standards for data curation (OECD, 2018^[40]).

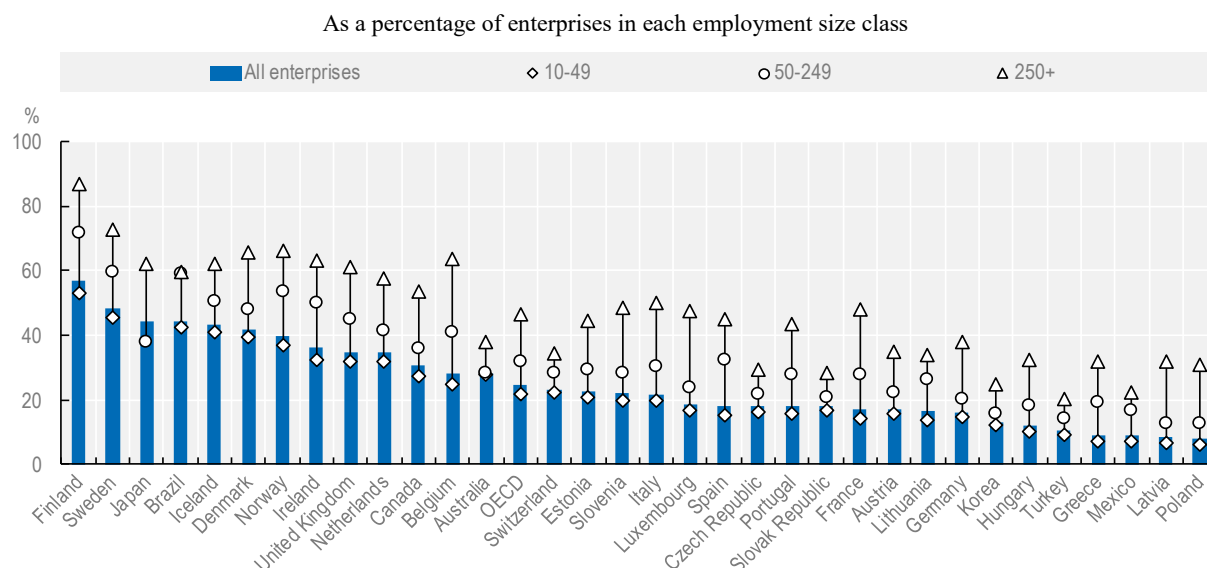
Data-related skills are not only a condition for effective re-use. They also affect trust and thus the willingness to provide data. This underlines the importance of awareness-raising and skills development for a better management of risks and the reinforcement of trust.

Data processing and analytic infrastructures

An important condition for the effective re-use of data is access to data processing and analytic infrastructures. This is particularly critical for SMEs, but also for individuals, including scientists. The diffusion of cloud computing has been a major catalyst for the re-use of data and big data in particular (OECD, 2015^[20]). But the adoption of cloud computing by firms remains much below expectations. In 2016, over 24% of businesses used cloud computing

services. This share ranges from over 57% in Finland down to 8% in Poland. In most countries, uptake is higher among large businesses (close to 50%) compared to small or medium-sized enterprises, which record around 22% and 32%, respectively (Figure 4.2).

Figure 4.2. Enterprises using cloud computing services, by firm size, 2016



Notes: Data refer to manufacturing and non-financial market services enterprises with ten or more persons employed, unless otherwise stated. Size classes are defined as small (10-49 persons employed), medium (50-249) and large (250 and more).

Source: OECD (2017_[41]), *ICT Access and Usage by Businesses* (database), <http://oe.cd/bus> (accessed in June 2017).

The lack of sustainable funding for open data infrastructures, such as research data repositories, remains a source of concern, in particular for public and scientific data. At the Copenhagen Expert Workshop, experts indicated that poor availability of open data infrastructures may have contributed to the erosion of scientific data. This situation is exacerbated by the lack of sustainable business models for data repositories, especially since there is often the assumption that access to open data is to be granted free of costs (OECD, 2017_[42]). But research institutions, including funding agencies, are struggling to keep up with demand for help providing or funding the infrastructure needed for data stewardship and data sharing, as well as the necessary training to support these activities (see subsection “Limitations of current business models and data markets”).

Lack of common standards for data sharing and re-use

One of the most frequently cited barriers to data sharing and re-use is the lack of common standards, or the proliferation of incompatible standards. For example, inconsistent data formats are impediments to the creation of longitudinal data sets, as changes in measurement and collection practices make it hard to compare and aggregate data.

This subsection highlights the importance of standards for data re-use across systems (interoperability). It discusses in particular whether enhanced access and sharing can facilitate the interconnection and interaction of distinct social and information systems through interoperability. It then focusses on the role of data intermediaries in defining standards, and on data quality standards, which are recognised as a condition for the efficient re-use of data.

Interoperability: Facilitating the interconnection and interaction of social systems

Some multinational science and research consortia have defined standards to enhance the level of interoperability. The Critical Path Institute and the Clinical Data Interchange Standards Consortium, for instance, released their Alzheimer's disease Therapeutic Area Standard (SDTM AD/Mild Cognitive Impairment User Guide) to facilitate analysis and learning from clinical studies for treatment or risk reduction (OECD, 2014^[43]; OECD, 2015^[44]). The User Guide outlines a standardised set of data elements so that pharmaceutical companies and other medical researchers can more easily, and consistently, collect data that can be reliably pooled and compared.

Standards are a condition for interoperability. Even when commonly used machine-readable formats are used for accessibility, interoperability is sometimes not guaranteed. These common formats may enable “syntactic” interoperability, i.e. the transfer of “data from a source system to a target system using data formats that can be decoded on the target system”. But they do not guarantee “semantic” interoperability, “defined as transferring data to a target such that the meaning of the data model is understood”.²¹ Both, syntactic and semantic interoperability are needed. Besides being accessible and interoperable, data need to be findable. This may require that data be catalogued and/or searchable.

Enhanced access and sharing are in some cases motivated by interoperability considerations. This is the case with data portability, which, in the case of the General Data Protection Regulation (GDPR), gives data subjects the right to receive the data provided in a structured, commonly used and machine-readable format, and to transmit those data to another controller (see subsection “Data portability” in Chapter 2). Although not a legal obligation under the GDPR, data portability may foster interoperability of data-intensive products, and as a result reduce switching costs to such an extent that businesses can no longer fully exploit the “stickiness” of their products to reinforce their market positions (lock-in effects).²²

The standard-setting role of data intermediaries

Data intermediaries have played, and continue to play, an important role for the development and promotion of data-related standards. In 2017, Google, Facebook, Microsoft, and Twitter joined forces in a new standard-setting initiative for data portability called the Data Transfer Project (DTP), most likely in anticipation of the GDPR right to data portability (Box 4.5).

Box 4.5. The Data Transfer Project: A private-sector initiative for data portability

The DTP was formed in 2017 as an open-source, service-to-service data portability platform allowing “all individuals across the web [to] easily move their data between online service providers whenever they want”. It was motivated by the recognition that “portability and interoperability are central to innovation”. Current contributors are Facebook, Google, Microsoft and Twitter.

The DTP uses services’ existing APIs and authorisation mechanisms to access data. It then uses service specific adapters to transfer that data into a common format, and then back into the new service’s API.

Use cases for DTP include porting data directly between services for “(i) trying out a new service; (ii) leaving a service; and (iii) backing up your data”.

Sources: DTP (n.d.^[45]), *Data Transfer Project*, <https://datatransferproject.dev/>; (DTP, n.d.^[46]), DTP Github website, <https://github.com/google/data-transfer-project> (accessed 5 February 2019).

The lack of a common data format across municipalities is a reason why end users (including businesses) may rely on data brokers, instead of using open government data directly. These data intermediaries often provide their data in common data formats that assure syntactic and semantic interoperability – driving their adoption across industries. One example is Google’s General Transit Feed Specification (GTFS), a common format for public transportation schedules and associated geographic information.

Data quality

The information that can be extracted from data depends on the quality of the data. Poor-quality data will almost always lead to poor data analysis and results. Therefore, data cleaning is often emphasised as an important step before the data can be analysed. This in turn involves significant costs, as it can account for 50% to 80% of a data analyst’s time together with the actual data collection (Lohr, 2014^[47]).

But data quality may not only affect the ability and the cost to re-use data. It can also prevent stakeholders from participating in data-sharing arrangements. According to some studies, uncertainties about data quality may explain, for instance, why open data repositories are used at far lower rates than most scholars and practicing data curators would expect.²³ As noted in OECD (2017^[42]), “many data sets are not of requisite quality, are not adequately documented or organised, or are of insufficient (or no) interest for use by others”. The lack of a common understanding what quality means in the context of data has also been a major source of uncertainty among organisations.²⁴ Some authors have therefore argued that data quality should be considered a key determinant of trust for data sharing (Federer et al., 2015^[48]; Sposito, 2017^[49]; Wallis et al., 2007^[50]).

Data quality is a challenging concept as it typically depends on the intended use of the data: data that are of good quality for certain applications can be of poor quality for other applications. The OECD (2012^[51]) *Quality Framework and Guidelines for OECD Statistical Activities* defines data quality as “fitness for use” in terms of user needs: “If data is accurate, they cannot be said to be of good quality if they are produced too late to be useful, or cannot be easily accessed, or appear to conflict with other data.” In other words, even if data are of good general quality, their use can lead to wrong results if the data are irrelevant and do not fit the business or scientific questions they are supposed to answer.²⁵

Data quality needs to be viewed as a multi-faceted concept, which is why data quality standards need to take into account the specific context of data use. The OECD (2012^[51]) defines seven data quality dimensions, the first two²⁶ – listed below – reflect the context of use:

1. **Relevance** “is characterised by the degree to which the data [serve] to address the purposes for which they are sought by users. It depends upon both the coverage of the required topics and the use of appropriate concepts”.
2. **Accuracy** is “the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure”.

The OECD Privacy Guidelines provide similar criteria for data quality in the context of privacy protection. The Guidelines state that “personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date”. This suggests that “completeness” should be considered another important dimension of data quality.²⁷

Misaligned incentives, and limitations of current business models and markets

The marginal costs of transmitting, copying and processing data can be close to zero. However, substantial investments are often required to collect data and enable data sharing and re-use (Johnson et al., 2017^[39]; Robinson and Johnson, 2016^[52]). Firms are investing a significant share of their capital in the acquisition of start-ups to secure access to data potentially critical for their business. Additional investments may be required to integrate and re-use all data sets. They may also be needed for data cleaning and data curation, which is often beyond the scope and time frame of the activities for which the data were initially collected and used (OECD, 2016^[10]).²⁸

The investments required for effective access to and sharing of data are not limited to data itself or to securing the engagement of all relevant stakeholders. In many cases complementary investments are needed in metadata, data models and algorithms for data storage and processing, to secure information technology infrastructures for shared data storage, processing, and access.²⁹ The overall total up-front costs and spending can be very high.

The following subsections discuss the root cause of the incentives problems faced by stakeholders, namely i) the externalities of data sharing and re-use and the “free-rider” dilemma; ii) the limitations of existing business models and data markets to meet the full range of demand for data; iii) the misaligned incentive structures, which exist in particular in science and research, and the risk of mandatory access to data; and iv) uncertainties about “data ownership”, an often misunderstood concept.

Externalities of data sharing and re-use and the misaligned incentives

The root cause of the incentive problems of data access and sharing can be attributed to a positive externality issue: data access and sharing may benefit others more than it may benefit the data holder and controller, who may not be able to privatise all the benefits of data re-use. Since data are in principle non-exclusive goods for which the costs of exclusion can be high, there is the possibility that some may “free ride” on others’ investments. The argument that follows is that if data are shared, free-riding users can “consume the resources without paying an adequate contribution to investors, who in turn are unable to recoup their investments” (Frischmann, 2012^[53]).

Data holders and controllers may not have the incentives to share their data, especially if the costs are perceived to be higher than the expected private benefits. In other words, where organisations and individuals cannot recuperate a sufficient level of the return on their data-related investments, for instance through revenues arising from granting and facilitating data access and sharing against fees, there is a high risk that data access and sharing will not occur at a sufficient level.

The situation poses even more incentive problems for scientists and researchers, who traditionally compete to be first to publish. They might not enjoy or even perceive the benefits of disclosing the data they could further use for as yet uncompleted research projects (OECD, 2016^[10]). Some have noted that the incentives in current reward and evaluation systems could be the main reason why researchers are reluctant to share scientific data: researchers are primarily rewarded for their scientific papers and not for the data they share with the scientific community.

However, the assumption that positive externalities and free riding always diminish incentives to invest cannot be generalised. This needs careful case-by-case scrutiny. This view has been supported by Frischmann (2012, p. 161^[53]), who notes:

There is a mistaken tendency to believe that any gain or loss in profits corresponds to an equal or proportional gain or loss in investment incentives, but this belief greatly oversimplifies the decision-making process and underlying economics and ignores the relevance of alternative opportunities for investment.

Free riding is sometimes the economic and social rationale for providing enhanced access to data. Open data initiatives, for example, are motivated by the recognition that users will free ride on the data provided, and in so doing will be able to create a wide range of new goods and services that were not anticipated and otherwise would not be produced.³⁰

Limitations of current business models and data markets

Market-based approaches are essential for encouraging data access and sharing. Data markets and platforms that provide added-value services such as a payment-and-data exchange infrastructure can facilitate data sharing, including the commercialisation of data. This was recognised by the G7 ICT and Industry Ministers in Turin in September 2017, where ministers stated that “open public-sector data, as well as market-based approaches to access and sharing of data are important to foster innovation in production and services, entrepreneurship and development of SMEs” (G7 Information Centre, 2017^[54]).

To enhance the functioning of existing markets, several challenges need to be acknowledged and, where possible, addressed. These challenges go beyond the issues of trust, data ownership, and standards discussed above. The following subsections focus on the extent to which: i) the pricing schemes of many data markets and platforms can appear opaque; and ii) data markets may not be able to fully serve social demand for data, i.e. where in particular data is used to produce public or social goods (e.g. scientific knowledge and democratic participation).

Lack of transparency and the limitations of market-based pricing

The value of data depends on the context of their use and the information and knowledge that can be drawn (OECD, 2015^[20]).³¹ This challenges the applicability of market-based pricing, in particular where the monetary valuation of the same data set can diverge significantly among market participants. For example, while economic experiments and surveys in the United States indicate that individuals are willing to reveal their social security numbers for USD 240 on average, the same data sets can be obtained for less than USD 10 from US data brokers such as Pallorium and LexisNexis (OECD, 2013^[55]).

The pricing schemes in many data market platforms may thus appear opaque as prices may vary depending on the type of client (e.g. researcher, firm or government), the size of the client, the markets in which the client is active, and the purpose for which the data are expected to be used. Furthermore, while businesses and consumers can benefit from the services of data markets and platforms, they are at the same time exposed to many risk factors not only due to lack of transparency, but also due to the often sensitive nature of the data.

There have been calls in the past for enhancing transparency in data markets, partly motivated by empirical work showing that high levels of transparency in markets are found to be associated with lower risk and costs of capital and higher trading volumes or liquidity (PriceWaterhouseCoopers, 2001^[56]; Bhattacharya, Daouk and Welker, 2003^[57]). Limited transparency also increases the risk of information asymmetry and thus the risk of consumer detriments. This has led to efforts such as in the United States where there has been a focus on promoting greater transparency of data brokers’ practices (Federal Trade Commission (US), 2014^[58]). These concerns have also encouraged a number of initiative that provide

greater consumers' access to data a company holds on them (Brill, 2013^[59]; Acxiom, 2014^[60]) (see also subsection "Data portability" in Chapter 5).

Unmet demand for data and challenges in sustaining the provision of open data

The functioning of data markets and platforms is also challenged by the limitation of data markets to fully serve social demand for data. This is sometimes characterised as a demand manifestation problem, where a resource (such as data) is used to produce a public or a social good (e.g. scientific knowledge and democratic participation) (Frischmann, 2012^[53]). In these cases, market prices are not able to fully reflect the value of data. In addition, the context dependency of data and the dynamic environment in which some data are used (e.g. research) make it almost impossible to fully evaluate ex ante the economic potential of data and would exacerbate a "demand manifestation" problem. All this may lead to a high risk of market failure and thus an under-provision of data or the prioritisation of access and use for a narrower range of uses than would be socially optimal (OECD, 2015^[20]).

In these cases, open data have traditionally been recognised as more appropriate. The provision of open data through data markets and platforms raises however other challenges, given that open data are expected to be provided for free as is often the case for public-sector data and research data generated from public funding. This raises the question of how the provision of data can be funded sustainably, given that costs are most often borne by data providers, while benefits accrue to data users (OECD, 2018^[40]). Sustainable business models and funding for open data provision (e.g. open data portals and research data infrastructure) as noted previously are therefore critical (OECD, 2017^[42]).

Given their public good characteristics, open data are often significantly, if not fully, publicly funded – in particular in science and in the public sector. However, increasing costs due to the growing volume and variety of data, combined with budgetary constraints in governments and resistances within the research community to redirect existing funds from research to infrastructure services, have challenged the sustainability of data repositories (OECD, 2017^[42]). Alternative revenue sources have thus become a necessity. The analysis of the business model of 47 research data repositories (OECD, 2017^[42]), shows that research data repositories, which are primarily serving public good objectives, have on average more than two revenue sources. "Typically, repository business models combine structural or host funding with various forms of research and other contract-for-services funding, or funding from charges for access to related value-added services or facilities" (OECD, 2017^[42]).

The risks of mandatory access to data

Mandatory access to data has been proposed (Mayer-Schönberger and Ramge, 2018^[61]; Villani, 2018^[62]), and in some cases legislated (see e.g. the right to data portability in Australia and the European Union), in particular where refusal to provide access (through licensing) constitutes an abuse of a dominant market position. This was the case in e.g. IMS Heath vs. NDC Health, two health care information companies, where access to a copyright protected data scheme was mandated.³² Some have therefore suggested that under "exceptional circumstances" data may be considered an "essential facility" in some jurisdictions (Box 4.6). In addition, some countries have started to define and regulate access to data of public interest (see subsection "Data of public interest" in Chapter 5).

However, while regulation may impose data access, it may also undermine incentives to invest in data in the first place, in particular when data commercialisation and licensing are not viable options. For instance, for organisations and individuals, including researchers, which build their competitive advantage based on data lock-in, mandatory data access and sharing

could undermine their ability to compete, to a point where their incentives to invest in data may be too low to enter a particular market. For some start-ups this could mean that they lose their attractiveness as acquisition targets of larger firms, and thus their economic value.

Box 4.6. The “essential facility” concept

An “essential facilities doctrine” (EFD) specifies when the owner(s) of an “essential” or “bottleneck” facility is mandated to provide access to that facility at a “reasonable” price. For example, such a doctrine may specify when a railroad must be made available on “reasonable” terms to a rival rail company or an electricity transmission grid to a rival electricity generator. The concept of “essential facilities” requires there to be two markets, often expressed as an upstream market and a downstream market. [...] Typically, one firm is active in both markets and other firms are active or wish to become active in the downstream market. [...] A downstream competitor wishes to buy an input from the integrated firm but is refused. An EFD defines those conditions under which the integrated firm will be mandated to supply. While essential facilities issues do arise in purely private, unregulated contexts, there is a tendency for them to arise more commonly in contexts where the owner/controller of the essential facility is subject to economic regulation or is state-owned or otherwise state-related. Hence, there is often a public policy choice to be made between the extension of economic regulation and an EFD under competition laws.

Source: OECD (1996_[63]), *The Essential Facilities Concept*, www.oecd.org/competition/abuse/1920021.pdf.

Some authors have therefore argued that mandatory access and sharing could end up having anti-competitive effects. Swire and Lagos (2013_[64]), for instance, argue that the specific provisions of Article 20 of the GDPR may actually have perverse anti-competitive effects as it may put start-ups and SMEs under the heavy obligation of investing in data portability.³³ At the Copenhagen Expert Workshop, some experts also warned that the portability of (non-personal) proprietary data, in particular if mandatory, could be more beneficial to large data-intensive businesses in the long run, because it would facilitate their access to data in niche markets, the markets typically served by start-ups.

Uncertainties about “data ownership”

The increasing social and economic importance of data often leads to the question about who owns data. This question is motivated by the recognition that ownership rights provide a “powerful basis for control” (Scassa, 2018_[65]) as “to have legal title and full property rights to something” (Chisholm, 2011_[66]) implies “the right to exclusive use of an asset” and “the full right to dispose of a thing at will” (Determann, 2018_[67]). It thus comes as no surprise that this question has led to controversial discussions across all sectors. The tensions between farmers and agriculture technology providers (ATPs) in the United States illustrates current issues related to “data ownership” (Box 4.7).

However, the concept of “data ownership” is used in different contexts with a different meaning.³⁴ The rights to control access, copy, use and delete data – what can be seen as the main rights associated with “data ownership” – are affected by different legal frameworks differently. IPRs, in particular copyright and trade secrets, can be applicable under certain conditions. In the case of personal data, privacy protection frameworks will be relevant for the question of data ownership as well. And there is even an opinion among some scholars that, in certain specific instances, data cannot be owned (Determann, 2018_[67]). In certain jurisdictions, cyber-criminal law may have the effects of conferring ownership-like rights

to data holders, while data ownership related questions emerging between firms can also be regulated by competition law as discussed above (Osborne Clarke, 2016^[68]).

Box 4.7. Data ownership controversies in data-driven agriculture

Farmers' ability to access and use agricultural data has become a key determinant for innovation and success. Major ATPs such as John Deere, DuPont Pioneer and Monsanto have recognised this trend when they started taking advantage of the Internet of Things (IoT) by integrating sensors with their latest equipment. In doing so they have been able to generate large volumes of data, which are being considered as an important data source for biotech companies (for example, to optimise genetically modified crops), crop insurance companies and traders on commodity markets.

The control of agricultural data by the major ATPs has led to controversial discussions on the potential harm to farmers from discrimination and financial exploitation and the question of who owns agricultural data (Bunge, 2014^[69]; The Economist, 2014^[70]; Poppe, Wolfert and Verdouw, 2015^[71]; Wolfert, 2017^[72]; Sykuta, 2016^[73]). For farmers, the benefits of data-intensive equipment became also less clear, because there was a sense that farmers would “degrade” to become local caretakers of land, animals and equipment, and only act like a contractor making sure that the interactions between the supply and demand sides of the agricultural system work together properly (OECD, 2015^[20]). This situation has been exacerbated by uncertainties about the question of data ownership (Banham, 2014^[74]; Igor, 2015^[75]).

In April 2014, major providers of precision farming technologies met with the American Farm Bureau Federation to discuss the future of the governance of agricultural data. The question of data ownership was central to this discussion. The result was the “Privacy and Security Principles for Farm Data” (Ag Data Transparent, 2016^[76]), signed by 39 organisations as of 1 April 2016 (see Section 5.2 for the principles). In July 2017, the US House of Representatives Agriculture Committee held a hearing on “The Future of Farming: Technological Innovations” in preparation for the 2018 Farm Bill or the Agriculture Improvement Act of 2018 (United States, 2018^[77]), which addresses some of the questions related to data governance after replacing the Agriculture Act of 2014 (P.L. 113-79), which expired at the end of fiscal year 2018.

The following subsections discuss the extent to which existing legal and regulatory frameworks affect the rights to control access, copy, use and delete data. These include: i) IPRs, in particular copyright and trade secrets; ii) privacy rights, and iii) contractual law. The last subsection on iv) “data commons” discusses to what extent some data can be governed as commons (including “data trusts”).

The role of intellectual property rights

Intellectual property regimes such as copyright and trade secrets are applicable under certain conditions. The extent to which the data itself, and separately, its arrangement or compilation, or its expression as information, are protected depends on the definition of data and in some cases remains controversial (Determann, 2018^[67]; Scassa, 2018^[65]). To the extent that data includes protectable works (e.g. electronic maps, photographs, and text), that data will be protected (by copyright in the case of the provided examples).

- **Copyright** typically “protects and rewards literary, artistic and scientific works, whatever may be the mode or form of their expression, including those in the form of computer programmes” (OECD, 2015^[78]).³⁵ The protection afforded to databases (as collections of data or other elements) is established – or confirmed – by both

Art. 10(2) of the WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) (WTO, 1994^[79]) and the almost identical Art. 5 of the World Intellectual Property Organization (WIPO) Copyright Treaty: “Compilations of data or other material, whether in machine-readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creation shall be protected as such [...]”³⁶ The arrangement or selection thus provides a separate layer of protection without prejudice to any rights to the content of the database itself. With the increasing use of APIs, which are implemented via software code, copyrights have gained further in importance as legal means for controlling data access and re-use (see subsection “Data-access control mechanisms: Protecting the interests of data holders” in Chapter 2).

- **Trade secrets** encompass “confidential business and technical information and know-how that a firm makes reasonable efforts to keep secret and that has economic value as a result” (OECD, 2015^[78]).³⁷ Trade secrets may protect the information conveyed by data, but only under some conditions, the most important one being that the information must be kept secret.³⁸ Not all data can thus be protected as trade secret. But even where data can be protected, the dissemination of the data will only be possible to authorised persons (subject to confidentiality agreements) to a very limited extent. That said, “by offering a measure of protection for valuable information and relieving businesses of the need to invest in more costly security measures, some trade secret laws may encourage businesses to invest in the development of such information” (OECD, 2015^[78]).
- **Sui generis database right:** In some jurisdictions, such as the European Union, Japan and Korea, databases are also protected by a so-called sui generis database right (SGDR), which provides an additional layer of protection for databases regardless of the intellectual creation (i.e. “selection or arrangement”) that may or may not be present. In other words, protection under the SGDR is granted without the requirement for human creativity or originality – contrary to IPRs such as copyright. What is protected more specifically is the investment in generating the database, i.e. in the obtaining, verification or presentation of the data.³⁹ This type of right, which is found for instance in the European Union (1996^[80]) Directive on the legal protection of databases (EU Database Directive), offers protection beyond the protection of arrangement or selection as it protects against the extraction and/or re-use of substantial parts of the database, and thus extends, at least to some extent, to the data themselves (OECD, 2015^[78]).

The co-existence of privacy protection frameworks

While there may be expectations among individuals that they own their personal data, the reality, in many, if not most, jurisdictions, is that they do not legally own their personal data. Data collected by an organisation (including personal data) will typically be considered the intellectual property of that organisation (i.e. proprietary personal data, see Figure 2.2 in Chapter 2). Scassa (2018^[65]), for example, discusses the court decision in Canada (in *McInerney v. MacDonald*),⁴⁰ where “one of the theories considered, and ultimately rejected, by the court was that a patient owned their personal medical information”. Instead, the court found that the physician, institution or clinic compiling the medical records owns the physical records.

However, in the case of personal data, the ownership rights of the organisation will hardly be comparable to other (intellectual) property rights. As Scassa (2018^[65]) states about the same court case, “the court also recognised an ‘interest’ on the part of the patient amounting

to a degree of control over the information”. In fact, most privacy regulatory frameworks give data subjects particular control rights over their personal data, which may interfere with “the right to exclusive use of an asset” and “the full right to dispose of a thing at will” (Determann, 2018^[67]), typically associated with ownership. Some authors have therefore stressed that privacy protection frameworks may have some characteristics similar to the ones of a property right (e.g. right to restriction of processing, right to data portability) (Banterle, 2018^[81]; Drexler, 2018^[82]; Purtova, 2017^[83]).

The Individual Participation Principle of the OECD Privacy Guidelines, for example, recommends that individuals have “the right: a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; b) to have communicated to him, data relating to him within a reasonable time; [...] and d) to challenge data relating to him [...]”. The GDPR in addition has extended this right to the right of data portability (Art. 20).⁴¹

The co-existence of privacy protection frameworks has led some experts to suggest that multiple owners had to be assumed (co-ownership), including in particular the data subject, “as neither the ‘data producer’ nor the ‘data gatherer’ could claim an exclusive right over the data”.⁴² These experts argue that multiple stakeholders were often involved in the contribution, collection and control of personal data, including the data subject him/herself. Assuming some kind of legal ownership rights of data subjects over their personal data may be over-interpreting the nature of privacy protection frameworks. It is sufficient from a data-governance perspective to acknowledge that privacy and other legal frameworks do coexist, and as a result, data and personal data in particular can be subject to multiple overlapping rights and obligations.

The intricate net of existing legal frameworks and the freedom to contract

The different legal frameworks do not preclude each other; in fact, they overlap. However, Determann (2018^[67]) notes that the “intricate net of existing legal frameworks” combined with the involvement of multiple parties in the creation of data (and its value) may explain current uncertainties related to data ownership. The challenge is exacerbated where data are created, and expected to be accessed and shared, across national borders.

As a result, stakeholders have come to rely on contract law as the primary legal vehicle for determining rights related to data control, access and use. These contractual arrangements often can better suit the individual context of data access, sharing and use (freedom of contract).

While freedom of contract may give stakeholders the ability to construct well-suited contractual arrangements, existing uncertainties may also increase transaction costs, and expose particularly those that are in a weaker position to negotiate fair terms and conditions for data access, sharing and re-use. These are typically individuals (consumers) and SMEs (see Box 4.7 on the conflict between farmers and ATPs). As a result, there are little incentives to share data, including with third parties, and where data-sharing arrangements exist, they may be perceived as unfair. This situation has motivated a number of government initiatives aiming at providing guidance for business-to-business data-sharing agreements (see subsection “Voluntary and collaborative approaches” in Chapter 5).

Public-private partnership and the implications on data ownership

Additional issues may also arise in the case of data partnerships, in particular where partners have different market power. Data ownership configurations may be changed as a result of these partnerships. Public-sector data may end up being “privatised” as result, of for example,

confidentiality agreements and private-sector data may end up in the public domain as it becomes subject to public sector information (PSI) frameworks. In science and research, concerns have been raised about the risk of privatisation of data from publicly funded research, in particular when copyright-based licensing schemes, such as those developed by Creative Commons, are not used in public-private partnerships (PPPs).

The risk of privatisation of publicly funded data is not limited to PPPs. The privatisation of organisations in data-intensive sectors including telecommunication services, financial services, transportation, and utilities has implications on the availability of data from these sectors. As public-sector organisations are privatised, they move out of the scope of PSI frameworks, making it challenging for third parties to access the data. The social and economic implications remain underexplored, despite the fact that the private sector is increasingly performing public services traditionally performed by the government. The long-term effects are a further narrowing of the relative scope of PSI frameworks. This trend is expected to accelerate with the deployment of applications such as “smart” cities and “smart” transportation, just to name a few. To address these risks, some countries have started to define and regulate access to data of public interest (see subsection “Data of public interest” in Chapter 5).

Data commons: The governance of shared resources of common interests

The concept of “commons” has been used to describe natural resources that are managed and used for collective benefit, as well as the governance mechanisms (including informal norms and values) affecting their consumption (Hess and Ostrom, 2007^[84]). Commons are collective goods, in which stakeholders have common interests, and which are characterised by the governance mechanisms surrounding their production and consumption. Applied to data, commons imply formal or informal governance institutions to enable the sustainable shared production and/or use of data, be it public data as in the case of open data, or shared data as in the case of more restricted data-sharing arrangements (Madison, 2014^[85]).⁴³

The concept of data commons has been particularly relevant for public data where there is a need for governance institutions that enable their sustainable sharing and use. This is typically the case with open data. It is however also used for where data is no longer privately owned by a single entity, and thus becomes a collective resource that requires management and governance institutions. Elinor Ostrom demonstrated this in her analysis of economic governance of natural resources, for which she was awarded the 2009 Nobel Prize in Economic Sciences (Ostrom et al., 2012^[86]).

References

- Axiom (2014), *Make Data Work for You - Know what data says about you and how it is used*, [60]
<http://www.aboutthedata.com> (accessed on 10 September 2017).
- Ag Data Transparent (2016), *Ag Data's Core Principles: The Privacy and Security Principles for Farm Data*, [76]
<http://www.agdatatransparent.com/principles/>.
- Australian Bureau of Statistics (2017), *Managing the Risk of Disclosure: The Five Safes Framework*, [24]
<http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017>.
- Bakhoun, M. et al. (eds.) (2018), *The Interface Between Data Protection and IP Law: The Case of Trade Secrets and the Database sui generis Right in Marketing Operations, and the Ownership of Raw Data in Big Data Analysis*, Springer, Berlin, Heidelberg, [81]
http://dx.doi.org/10.1007/978-3-662-57646-5_16.
- Banham, R. (2014), *Who Owns Farmers' Big Data?*, [74]
<http://www.forbes.com/sites/emc/2014/07/08/who-owns-farmers-big-data/>.
- BBC (2014), "Sony Pictures computer system hacked in online attack", *BBC News*, [88]
<http://www.bbc.com/news/technology-30189029>.
- Bhattacharya, U., H. Daouk and M. Welker (2003), "The World Price of Earnings Opacity", *The Accounting Review*, Vol. 78/No. 3, pp. 641-678. [57]
- Brill, J. (2013), *Reclaim Your Name*, [59]
<http://www.ftc.gov/speeches/brill/130626computersfreedom.pdf>.
- Bunge, J. (2014), *Agricultural Firms, Farm Groups Strike Deal on Crop Data*, [69]
<http://www.wsj.com/articles/agricultural-firms-farm-groups-strike-deal-on-crop-data-1415854870>.
- Cadwalladr, C. and E. Graham-Harrison (2018), "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach", *The Guardian*, [6]
<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
- Chisholm, M. (2011), *What is data ownership?*, <http://www.b-eye-network.com/view/15697> [66]
 (accessed on 11 February 2019).
- CIGI and Ipsos (2014), *2014 CIGI-Ipsos Global Survey on Internet Security and Trust*, [93]
<http://www.cigionline.org/internet-survey-2014> (accessed on 5 February 2019).
- Deloitte (2017), *Assessing the value of TfL's open data and digital partnerships*, [33]
<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> (accessed on 2 March 2018).

- Department for Business Innovation and Skills (UK) (2012), *Midata: Impact assessment for midata*,
http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32689/12-944-midata-impact-assessment.pdf. [37]
- Department for Digital, Culture, Media and Sport (UK) (2018), *Centre for Data Ethics and Innovation Consultation*, <http://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation> (accessed on 1 October 2018). [15]
- Department for Digital, Culture, Media and Sport (UK) (2018), *Guidance Data Ethics Framework*, <http://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework> (accessed on 1 October 2018). [14]
- Department of the Prime Minister and Cabinet [Australia] (2018), *New Australian Government Data Sharing and Release Legislation: Issues Paper for Consultation*,
<http://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation>. [1]
- Determann, L. (2018), “No One Owns Data”, *UC Hastings Research Paper* No. 265,
<http://dx.doi.org/10.2139/ssrn.3123957>. [67]
- DTP (n.d.), *Data Transfer Project*, <https://datatransferproject.dev/> (accessed on 1 October 2018). [45]
- DTP (n.d.), *google/data-transfer-project*, <https://github.com/google/data-transfer-project> (accessed on 5 February 2019). [46]
- ECJ (2004), *IMS Heath GmbH & Co. OHG vs. NDC Health GmbH & Co. KG*,
<http://curia.europa.eu/juris/showPdf.jsf?docid=55399&doclang=EN>. [97]
- ENISA (2009), “Cloud Computing - SME Survey”,
http://www.enisa.europa.eu/publications/cloud-computing-sme-survey/at_download/fullReport. [92]
- European Commission (2018), *European Big Data Hackathon 2017*,
https://ec.europa.eu/eurostat/cros/EU-BD-Hackathon_en (accessed on 24 October 2018). [35]
- European Commission (2018), “Evaluation of Directive 96/9/EC on the legal protection of databases”, *Commission Staff Working Document*,
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51764. [94]
- European Commission (2015), “Data protection”, *Special Eurobarometer 431*,
http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf. [17]
- European Union (1996), *Directive 96/9/EC of The European Parliament and of the Council of 11 March 1996 on the legal protection of databases*, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009>. [80]

- Federal Trade Commission (US) (2014), *Data brokers: A call for transparency and accountability*, <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. [58]
- Federer, L. et al. (2015), “Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff”, <http://dx.doi.org/10.1371/journal.pone.0129506>. [48]
- Franceschi, A. and R. Schulze (eds.) (2018), *Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy*, Max Planck Institute for Innovation & Competition Research Paper, <https://ssrn.com/abstract=3274519>. [82]
- Frischmann, B. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press. [53]
- Frischmann, B., M. Madison and K. Strandburg (eds.) (2014), *Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo*, Oxford University Press. [85]
- G20 (2017), *G20 Digital Economy Ministerial Conference*, <http://www.bmwi.de/Redaktion/DE/Downloads/G/g20-digital-economy-ministerial-declaration-english-version.pdf> (accessed on 11 February 2019). [32]
- G20 (2015), *Introductory Note to the G20 Anti-Corruption Open Data Principles*, <http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>. [26]
- G7 (2016), *Outcomes of the G7 ICT Ministers’ Meeting in Takamatsu, Kagawa*, http://www.soumu.go.jp/joho_kokusai/g7ict/english/about.html (accessed on 1 October 2018). [27]
- G7 Information Centre (2017), *G7 ICT and Industry Ministers’ Declaration: Making the Next Production Revolution Inclusive, Open and Secure*, <http://www.g8.utoronto.ca/ict/2017-ict-declaration.html>. [54]
- Goodin, D. (2015), “Pay or we’ll knock your site offline – DDoS-for-ransom attacks surge”, *ars Technica*, <https://arstechnica.com/information-technology/2015/11/pay-or-well-knock-your-site-offline-ddos-for-ransom-attacks-surge/>. [99]
- Granville, K. (2018), “Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens”, *The New York Times*, <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>. [5]
- Greenaway, K., S. Zabolotniuk and A. Levin (2012), “Privacy as a risk management challenge for corporate practice”, *Ted Rogers School of Management, Ryerson University, Privacy and Cyber Crime Institute*, http://www.ryerson.ca/content/dam/tedrogersschool/privacy/privacy_as_a_risk_management_challenge.pdf. [23]
- Hern, A. and D. Pegg (2018), “Facebook fined for data breaches in Cambridge Analytica scandal”, *The Guardian*, <https://www.theguardian.com/technology/2018/jul/11/facebook-fined-for-data-breaches-in-cambridge-analytica-scandal>. [7]

- Hess, C. and E. Ostrom (eds.) (2007), *Understanding Knowledge as a Commons: From Theory to Practice*, MIT Press, Cambridge, Mass. [84]
- Hofheinz, P. and D. Osimo (2017), *Making Europe a Data Economy: A New Framework for Free Movement of Data in the Digital Age*, http://www.lisboncouncil.net//index.php?option=com_downloads&id=1338. [87]
- IBM (2018), *IBM Study: Hidden Costs of Data Breaches Increase Expenses for Businesses*, <https://newsroom.ibm.com/2018-07-11-IBM-Study-Hidden-Costs-of-Data-Breaches-Increase-Expenses-for-Businesses>. [95]
- Igor, I. (2015), *How To Approach Data Ownership In AgTech?*, <https://medium.com/remote-sensing-in-agriculture/how-to-approach-data-ownership-in-agtech-486179dc9377>. [75]
- Information Commissioner's Office (2018), *ICO issues maximum £500,000 fine to Facebook for failing to protect users' personal information*, <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/10/facebook-issued-with-maximum-500-000-fine/> (accessed on 25 October 2018). [96]
- Information Commissioner's Office (2018), *Monetary Penalty Notice to Facebook Ireland Ltd*, <https://ico.org.uk/media/action-weve-taken/mpns/2260051/r-facebook-mpn-20181024.pdf>. [4]
- Johnson, P. et al. (2017), "The Cost(s) of Geospatial Open Data", *Transactions in GIS*, Vol. 21/3, <http://dx.doi.org/10.1111/tgis.12283>. [39]
- Kaye, J. et al. (2015), "Dynamic consent: a patient interface for twenty-first century research networks", *European Journal of Human Genetics*, Vol. 23/2, pp. 141-146, <http://dx.doi.org/10.1038/ejhg.2014.71>. [19]
- Keogh, M. (ed.) (2015), "A European Perspective on the Economics of Big Data", *Farm Policy Journal*, Vol. 12/No. 1, Chapter 2, http://www.researchgate.net/profile/Sjaak_Wolfert/publication/278300518_A_European_Perspective_on_the_Economics_of_Big_Data/links/557e8e3c08aeea18b777cb11/A-European-Perspective-on-the-Economics-of-Big-Data.pdf. [71]
- Lohr, S. (2014), "For big-data scientists, 'janitor work' is key hurdle to insights", *New York Times*, <http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>. [47]
- Madden, M. (2014), "Public perceptions of privacy and security in the post-Snowden era", *Pew Research Center*, <https://www.pewinternet.org/2014/11/12/public-privacy-perceptions/>. [16]
- Marcus, G. and E. Davis (2014), "Eight (no, nine!) problems with big data", *New York Times*, <http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html>. [89]
- Mayer-Schönberger, V. and T. Ramge (2018), "A Big Choice for Big Tech: Share Data or Suffer the Consequences", *Foreign Affairs* September/October 2018 Issue, <http://www.foreignaffairs.com/articles/world/2018-08-13/big-choice-big-tech>. [61]

- Ministry of Industry, Business and Financial Affairs [Denmark] (2019), *Dataetiske tiltag for erhvervslivet*, https://em.dk/media/12932/faktaark_dataetiske-initiativer.pdf. [13]
- National Institute of Standards and Technology (US) (2017), *An Introduction to Privacy Engineering and Risk Management in Federal Systems*, <http://dx.doi.org/10.6028/NIST.IR.8062>. [22]
- Nissenbaum, H. (2004), “Privacy as Contextual Integrity”, *Washington Law Review*, Vol. June, <https://www.nyu.edu/projects/nissenbaum/papers/washingtonlawreview.pdf> (accessed on 24 March 2018). [9]
- OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-en. [40]
- OECD (2017), “Business models for sustainable research data repositories”, *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, <http://dx.doi.org/10.1787/302b12bb-en>. [42]
- OECD (2017), *ICT Access and Usage by Businesses*, (database), OECD, Paris, <http://oe.cd/bus> (accessed on June 2017). [41]
- OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264276284-en>. [3]
- OECD (2016), *Declaration on the Digital Economy: Innovation, Growth and Social Prosperity (Cancún Declaration)*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0426>. [31]
- OECD (2016), “Health Data Governance Recommendation”, in *Recommendation of the Council on Health Data Governance*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433>. [2]
- OECD (2016), “Research ethics and new forms of data for social and economic research”, *OECD Science, Technology and Industry Policy Papers*, No. 34, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jln7vnp32-en>. [10]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [20]
- OECD (2015), *Enquiries into Intellectual Property’s Economic Impact*, OECD, Paris, <https://www.oecd.org/sti/ieconomy/intellectual-property-economic-impact.htm>. [78]
- OECD (2015), *Health Data Governance: Privacy, Monitoring and Research*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264244566-en>. [11]
- OECD (2015), *Health Data Governance: Privacy, Monitoring and Research*, (policy brief), OECD, Paris, <https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdf>. [29]

- OECD (2015), *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0415>. [21]
- OECD (2015), “The evolution of health care in a data-rich environment”, in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-12-en>. [44]
- OECD (2014), “Unleashing the power of big data for Alzheimer’s disease and dementia research: Main points of the OECD Expert Consultation on Unlocking Global Collaboration to Accelerate Innovation for Alzheimer’s Disease and Dementia”, *OECD Digital Economy Papers*, No. 233, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jz73kvmvbw-b-en>. [43]
- OECD (2013), “Exploring the economics of personal data: A survey of methodologies for measuring monetary value”, *OECD Digital Economy Papers*, No. 220, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k486qtxldmq-en>. [55]
- OECD (2013), *New Data for Understanding the Human Condition: International Perspectives*, OECD, Paris, <http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf>. [100]
- OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>. [18]
- OECD (2012), *Quality Framework and Guidelines for OECD Statistical Activities*, OECD, Paris, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs\(2011\)1&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs(2011)1&doclanguage=en). [51]
- OECD (2011), *Recommendation of the Council on Principles for Internet Policy Making*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0387>. [28]
- OECD (2011), “The evolving privacy landscape: 30 years after the OECD Privacy Guidelines”, *OECD Digital Economy Papers*, No. 176, OECD Publishing, Paris, <https://dx.doi.org/10.1787/5kgf09z90c31-en>. [101]
- OECD (2010), *Information Exchanges Between Competitors under Competition Law*, OECD, Paris, <http://www.oecd.org/competition/cartels/48379006.pdf>. [36]
- OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/122/122.en.pdf>. [30]
- OECD (1996), *The Essential Facilities Concept*, OECD, Paris, <http://www.oecd.org/competition/abuse/1920021.pdf>. [63]
- OECD (1985), *Declaration on Transborder Data Flows*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0216>. [25]

- Olenski, S. (2018), *3 Barriers To Data Quality And How To Solve For Them*, [98]
<http://www.forbes.com/sites/steveolenski/2018/04/23/3-barriers-to-data-quality-and-how-to-solve-for-them/#7399561429e7>.
- Osborne Clarke (2016), *Legal Study on Ownership and Access to Data*, A Study prepared for the [68]
 European Commission DG Communications Networks, Content & Technology,
<https://publications.europa.eu/en/publication-detail/-/publication/d0bec895-b603-11e6-9e3c-01aa75ed71a1>.
- Ostrom, E. et al. (2012), *The Future of the Commons: Beyond Market Failure and*, Indiana [86]
 University, Bloomington School of Public & Environmental Affairs Research Paper No.
 2012-12-02, <http://ssrn.com/abstract=2267381>.
- Otaka, T. (2015), “Japan Pension Service hack used classic attack method”, *The Japan Times*, [8]
<http://www.japantimes.co.jp/news/2015/06/02/national/social-issues/japan-pension-service-hack-used-classic-attack-method>.
- PriceWaterhouseCoopers (2001), *Investigating the costs of opacity: Deterred foreign direct [56]
 investment*.
- Purtova, N. (2017), “Do Property Rights in Personal Data Make Sense after the Big Data Turn?: [83]
 Individual Control and Transparency”, *Journal of Law and Economic Regulation* 10(2),
<https://ssrn.com/abstract=3070228>.
- Rae, S. (2018), “Big data skills shortages – and how to work around them”, *ComputerWeekly*, [38]
<https://www.computerweekly.com/opinion/Big-data-skills-shortages-and-how-to-work-around-them>.
- Robinson, P. and P. & Johnson (2016), “Civic hackathons: New terrain for local government- [52]
 citizen interaction?”, *Urban Planning*, Vol. 1/2, pp. 65-74,
<http://dx.doi.org/10.17645/up.v1i2.627>.
- Scassa, T. (2018), “Data Ownership”, *CIGI Papers* No. 187, [65]
https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf.
- Shapiro, C. and H. Varian (1999), *Information Rules: A Strategic Guide to the Network [90]
 Economy*, Harvard Business Press.
- Sposito, F. (2017), *What do data curators care about? Data quality, user trust, and the data [49]
 reuse plan*, <http://library.ifla.org/1797/1/S06-2017-sposito-en.pdf>.
- Swire, P. and Y. Lagos (2013), “Why the Right to Data Portability Likely Reduces Consumer [64]
 Welfare: Antitrust and Privacy Critique”, *Maryland Law Review*, Vol. 72:335,
<http://dx.doi.org/10.2139/ssrn.2159157>.
- Sykuta, M. (2016), “Big Data in Agriculture: Property Rights, Privacy and Competition in Ag [73]
 Data Services”, *International Food and Agribusiness Management Review*, Vol. 19/Special
 Issue - Issue A, <http://www.ifama.org/resources/Documents/v19ia/320150137.pdf>.

- The Economist (2014), “Digital disruption on the farm”, *The Economist*, [70]
<http://www.economist.com/business/2014/05/24/digital-disruption-on-the-farm>.
- The Expert Group on Data Ethics (2018), *Data for the Benefit of the People: Recommendations from the Danish Expert Group on Data Ethics*, <https://eng.em.dk/media/12209/dataethics-v2.pdf>. [12]
- United States (2018), *Agriculture Improvement Act of 2018*, [77]
<http://www.govtrack.us/congress/bills/115/hr2/text>.
- US General Services Administration (2018), *About Challenge.gov*, [34]
<https://www.challenge.gov/about/> (accessed on 1 October 2018).
- Villani, C. (2018), “For a Meaningful Artificial Intelligence: Towards a French and European Strategy”, *AI For Humanity*, http://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf. [62]
- Wallis, J. et al. (2007), “Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries”, *Center for Embedded Network Sensing* 4675, [50]
http://dx.doi.org/10.1007/978-3-540-74851-9_32.
- Wallis, J., E. Rolando and C. Borgman (2013), “If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology”, *PLoS One* 8(7), [91]
<http://dx.doi.org/10.1371/journal.pone.0067332>.
- Wolfert, S. (2017), “Big Data in Smart Farming – A review”, *Agricultural Systems*, Vol. 153, [72]
<http://dx.doi.org/10.1016/j.agsy.2017.01.023>.
- WTO (1994), *Agreement on trade-related aspects of intellectual property rights*, WTO, Geneva. [79]

Notes

¹ Today, for example, 34% Internet users in the European Union say that they are less likely to give personal information on websites (OECD, 2017^[3]). Over one-third (37%) use software that protects them from seeing online adverts and more than a quarter (27%) use software that prevents their online activities from being monitored. Overall, 65% of respondents have taken at least one of these actions. The changing behaviour is confirmed by a recent survey of 24 000 users in 24 countries in 2014 commissioned by the Centre for International Governance Innovation (CIGI), which reveals that only 17% of users said they had not changed their online behaviour in recent years (CIGI and Ipsos, 2014^[93]).

² The OECD (2016^[2]) *Health Data Governance Recommendation* is structured according to 12 high-level principles, ranging from engagement of a wide range of stakeholders, to effective consent and choice mechanisms to the collection and use of personal health data, to monitoring and evaluation mechanisms. These principles set the conditions to encourage greater cross-country harmonisation

of data-governance frameworks so that more countries are able to use health data for research, statistics and health care quality improvement, as well as for international comparisons.

³ This was, for example, the case with the attack that targeted Sony Pictures Entertainment at the end of 2014, exposing unreleased movies, employee data, emails between employees, and sensitive business information like sales and marketing plans (BBC, 2014^[88]). This incident resulted in the exposure of 103 million records and as a consequence to a 23-day closure of the PlayStation Network. According to Sony's executives, this data breach cost the company at least USD 171 million.

⁴ A data breach is "a loss, unauthorised access to or disclosure of personal data as a result of a failure of the organisation to effectively safeguard the data" (OECD, 2011^[101]).

⁵ The severity and impact of data breaches have also increased (OECD, 2017^[3]). According to a study released in 2018 by the data security research organisation the Ponemon Institute, the total average cost of a data breach is now USD 3.9 million, compared to USD 3.5 million in 2014 (IBM, 2018^[95]).

⁶ DoS incidents affect an organisation by flooding its online service or bandwidth with spam requests, knocking it offline for hours or days (Goodin, 2015^[99]).

⁷ The Choicepoint breach became public because of a 2003 California law requiring notification to an individual when their personal information was wrongfully disclosed. This contributed to the adoption of similar laws in many other jurisdictions. The OECD (2013^[18]) *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (paragraph 15(c)) call for controllers to provide notifications in cases where there has been a significant security breach affecting personal data.

⁸ As the Information Commissioner's Office (2018^[96]) noted: "The ICO's investigation found that between 2007 and 2014, Facebook processed the personal information of users unfairly by allowing application developers access to their information without sufficiently clear and informed consent, and allowing access even if users had not downloaded the app, but were simply "friends" with people who had. Facebook also failed to keep the personal information secure because it failed to make suitable checks on apps and developers using its platform. These failings meant one developer, Dr Aleksandr Kogan and his company GSR, harvested the Facebook data of up to 87 million people worldwide, without their knowledge. A subset of this data was later shared with other organisations, including SCL Group, the parent company of Cambridge Analytica who were involved in political campaigning in the United States. Even after the misuse of the data was discovered in December 2015, Facebook did not do enough to ensure those who continued to hold it had taken adequate and timely remedial action, including deletion. In the case of SCL Group, Facebook did not suspend the company from its platform until 2018."

⁹ Researchers in collaboration with Cambridge Analytica had asked users to take a personality survey via an app, which collected personal data not only of the 270 000 Facebook users, who had given their consent to participate in what was believed to be a research related activity, but also personal data of their "friends", leading a total of 50 million profiles being collected and used by Cambridge Analytica (Granville, 2018^[5]).

¹⁰ This could for instance be the case, where more general consent is used as the basis for data collection, use and sharing, without the *explicit* knowledge of data subjects, or where there are risks that the inferences could be used in ways that individuals do not desire or expect or which adversely affect them – for example, when it results in unfair discrimination.

¹¹ At the Copenhagen Expert Workshop, for example, experts highlighted that data portability could increase the level of security and privacy risks because data would be more likely to be accessed and shared inappropriately by a third party. Experts argued that a lot depended on the identification, authentication and other security measures companies would put into place to respond to data portability requests.

¹² In a survey of European SME perspectives on cloud computing, the security of corporate data and potential loss of control featured highly among the concerns for SME owners (ENISA, 2009^[92]). Loss of control in the case of cloud computing is partly related to uncertainties about the location of

the data, which is perceived across countries as significant a barrier to cloud computing adoption as the risk of security incidents.

¹³ 80% of social networking site users in the United States are concerned with third party access by businesses, and 70% with third party access by governments.

¹⁴ Examples include supervised research data centres, where authorised researchers analyse data within a physically secure location; and secure remote data access services, where authorised researchers enter a secure portal (OECD, 2013^[100]; OECD, 2016^[10]).

¹⁵ Results from the Eurostat Community Survey on ICT usage and e-commerce in enterprises consistently indicate that SMEs were less likely to have a formally defined ICT security policy across all reporting EU countries in 2015. In almost all countries, the differential between SMEs and large enterprises was approximately 30 percentage points.

¹⁶ While the study by Greenaway, Zabolotniuk and Levin (2012) may indicate a lack of understanding of how to implement privacy regulatory requirements, it may also reflect a lack of organisational strategies on how to deal with privacy risk and a gap in the assignment of responsibilities.

¹⁷ These include Argentina, Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, the Czech Republic, Denmark, Ecuador, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, the United States and the European Union.

¹⁸ The G20 comprises 19 countries plus the European Union. These countries are Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, Korea, Turkey, the United Kingdom and the United States.

¹⁹ At the Copenhagen Expert Workshop, some experts also stressed the significance of co-ordination costs due to frictions and poor interoperability as well as possible opportunity costs due to unrealised innovation.

²⁰ The OECD (2010^[36]) notes that “[g]enerally, information exchanges among competitors may fall into three different scenarios under competition rules: i) as a part of a wider price fixing or market sharing agreement whereby the exchange of information functions as a facilitating factor; ii) in the context of broader efficiency enhancing co-operation agreements such as joint venture, standardisation or R&D agreements; or iii) as a stand-alone practice, whereby the exchange of information is the only co-operation among competitors”.

²¹ This would typically require the use of mutually understood ontologies and metadata such as Web Ontology Language (OWL) and the Dublin Core Schema. The Dublin Core Metadata Terms were endorsed in the Internet Engineering Task Force (IETF) RFC 5013 and the International Organization for Standardization (ISO) Standard 15836-2009 (see Endnotes 18 and 19 in Chapter 2 as well as Box 2.6 in OECD (2015^[20])).

²² Data portability in theory would enable data users (incl. consumers and businesses) to easily change to new and potentially better data-intensive goods and services and possibly foster competition and innovation. Data portability is therefore often compared to “number portability” – a concept that is now an established part of OECD country’s telecommunications policy – although data portability is much more complex so that comparisons remain mainly theoretical.

²³ Through a survey of 190 clinical and basic science researchers in the Intramural Research Program at the National Institutes of Health in the United States, Federer et al. (2015^[48]) show that while 71% of respondents reported sharing data directly with colleagues, only 39% reported ever having used a data repository service for data sharing. See also Sposito (2017^[49]) and Wallis, Rolando and Borgman (2013^[91]).

²⁴ As Olenski (2018^[98]) notes: “Measuring and benchmarking data quality and accuracy in digital continues to be a hurdle. This challenge is driven by fragmentation across vendors and a lack of standards by key third-party auditors like Nielsen and comScore.”

²⁵ Experts recognise that it is often too tempting to think that with big data one has sufficient information to answer almost every question and to neglect data biases that could lead to false conclusions, because correlations can often appear statistically significant even if there is no causal relationship. Marcus and Davis (2014^[89]) give the illustration of big data analysis revealing a strong correlation of the United States murder rate with the market share of Internet Explorer from 2006 to 2011. Obviously, any causal relationship between the two variables is spurious.

²⁶ The remaining five include: 1. Credibility – “the credibility of data products refers to the confidence that users place in those products based simply on their image of the data producer, i.e. the brand image. Confidence by users is built over time. One important aspect is trust in the objectivity of the data.” 2. Timeliness – “reflects the length of time between their availability and the event or phenomenon they describe, but considered in the context of the time period that permits the information to be of value and still acted upon. [...] Real-time data [are] data with a minimal timeliness.” 3. Accessibility – “reflects how readily the data can be located and accessed”. 4. Interpretability – “reflects the ease with which the user may understand and properly use and analyse the data.” The availability of metadata plays an important role here, as they provide for example “the definitions of concepts, target populations, variables and terminology, underlying the data, and information describing the limitations of the data, if any.” And 5. Coherence – “reflects the degree to which they are logically connected and mutually consistent. Coherence implies that the same term should not be used without explanation for different concepts or data items; that different terms should not be used without explanation for the same concept or data item; and that variations in methodology that might affect data values should not be made without explanation. Coherence in its loosest sense implies the data are ‘at least reconcilable’” (OECD, 2012^[51]).

²⁷ One could also argue for an addition ninth dimension “cost efficiency” with which data are collected could also be considered as a measure for data quality. “Whilst the OECD does not regard cost-efficiency as a dimension of quality, it is a factor that must be taken into account in any analysis of quality as it can affect quality in all dimensions” (OECD, 2012^[51]).

²⁸ Data curation embodies data-management activities necessary to assure long-term data quality across the data life cycle, is needed to assure to sustainability of data-related investments.

²⁹ See Endnote 7 in Chapter 2.

³⁰ According to Frischmann (2012^[53]), “free riding is pervasive in society and a feature, rather than a bug”.

³¹ In fact, information – more than any other good – is an experience good, i.e. a good that consumers must experience in order to value. “Virtually any new product is an experience good”; however, “information is an experience good every time it’s consumed” (Shapiro and Varian, 1999^[90]).

³² IMS Health had developed a copyright protected data scheme for compiling information on sales of prescription pharmaceutical products. On the basis of its copyright the company had refused to licence the data scheme to its competitors. In its ruling, the ECJ considered that the “refusal by an undertaking which holds a dominant position and owns an intellectual property right in a brick structure indispensable to the presentation of regional sales data on pharmaceutical products in a Member State to grant a license to use that structure to another undertaking which also wishes to provide such data in the same Member State, can constitute an abuse of a dominant position within the meaning of Article 82 EC” (see ECJ (2004^[97])).

³³ Swire and Lagos’s critique applies to the draft regulation proposed by the European Commission in 2012. In the meanwhile, the regulation was approved in 2016 with modifications, which may partially address some of the authors’ concerns.

³⁴ Within businesses, for example, data ownership is often used to assign responsibility and accountability for specific databases (the “data owners”). In this context, ownership is perceived as a means of assuring data quality and curation, as well as data protection and security. Some authors have therefore suggested replacing the term “ownership” with “stewardship” (Scofield, 1998; Chisholm, 2011). However “stewardship” refers to a management function and not to the underlying ownership rights (or lack thereof). It is therefore not an appropriate substitute for *ownership*, but it may be used as a separate category to acknowledge that some entities (licensees, users) may have access to or use data without having ownership rights.

³⁵ As further explained in OECD (2015^[78]), “copyright laws provide for certain exceptions and limitations. Their protections typically last 50 to 70 years after the death of the creator (and for shorter periods for those works whose term of protection is based on the date of fixation or communication to the public). Copyrights stimulate creativity by assuring individuals and businesses, large and small, that the original, expressive material they create will not be reproduced, adapted, communicated to the public, displayed, distributed or performed without their permission or otherwise used in a manner that violates the exclusive rights of the copyright owners.”

³⁶ See Art. 10(2) TRIPS (www.wipo.int/wipolex/en/other_treaties/text.jsp?file_id=305907, accessed 5 February 2019).

³⁷ The term “trade secret” may vary across jurisdictions and may include various types of information, irrespective of the means of storage and the forms of expression of this information. In the United States, for instance, US Code 18 USC 1839(3) defines the term “trade secret” as “all forms and types of financial, business, scientific, technical, economic, or engineering information, including patterns, plans, compilations, programme devices, formulas, designs, prototypes, methods, techniques, processes, procedures, programmes, or codes, whether tangible or intangible, and whether or how stored, compiled, or memorialised physically, electronically, graphically, photographically, or in writing if: (A) the owner thereof has taken reasonable measures to keep such information secret; and (B) the information derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable through proper means by, the public” (see <https://www.govinfo.gov/app/details/USCODE-2011-title18/USCODE-2011-title18-partI-chap90-sec1839>, accessed 11 February 2019).

³⁸ See endnote 5 in Chapter 2.

³⁹ The European Union (1996^[80]) Directive on the legal protection of databases (EU Database Directive), for instance, requires that “there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents [...]” [Art. 7 (1)]. That said, the European Commission (2018^[94]) assumes that “the sui generis right does not apply to databases that are the by-products of the main activity of an organisation. This means that the sui generis right does not apply broadly to the data economy (machine-generated data, IoT devices, big data, AI, etc.); it only covers databases that contain data obtained from external sources (for example industries like publishers, who seek out data in order to commercialise databases).”

⁴⁰ See *McInerney v MacDonald*, [1992] 2 SCR 138, 1992 CanLII 57 (SCC), available at <http://canlii.ca/t/1fsbl> (accessed 5 February 2019).

⁴¹ The right to data portability in the GDPR is composed of three different rights: i) the right to receive (without hindrance from the data controller) data concerning data subject which he/she has provided; ii) the right to transmit (without hindrance from the data controller) those data to another controller; and iii) the right to have the personal data transmitted directly from one controller to another.

⁴² According to Hofheinz and Osimo (2017^[87]), co-ownership was essentially about the rights and responsibilities that a limited number of stakeholders have vis-à-vis each other. To help understand the implications for data ownership, the authors advise to compare data “ownership” to the rights parents have over their children.

⁴³ Therefore, data commons should not be misunderstood, and used as a synonym for “open data”, which are a particular type of data commons.

5. Policy initiatives enhancing data access and sharing

This chapter assesses current trends in policies aimed at enhancing data access and sharing. It is based on two country surveys including a total of 205 policy initiatives across 37 countries. Four priority areas are revealed through the analysis of these policy initiatives and are presented in this chapter: i) enhancing access to and sharing of public-sector information/data; ii) facilitating data sharing within the private sector; iii) increasing data analytic capacities across society; and iv) developing national data strategies.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

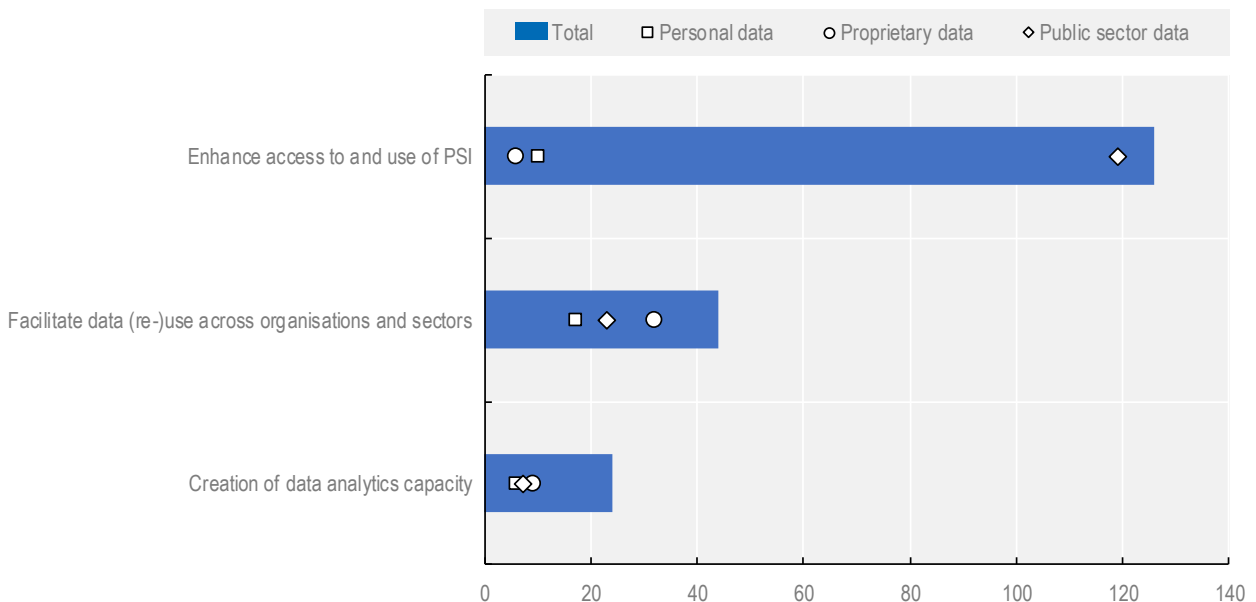
Governments have a major role to play in encouraging, facilitating and enhancing data access and sharing through policy action and governance frameworks. “Effectively safeguarding the public interest is another important function of governments”, as articulated in the OECD (2016^[1]) *Recommendation of the Council on Health Data Governance* (hereafter “Recommendation on Health Data Governance”). The leadership role of governments is also reflected in their ability to foster and enhance access to and sharing of public-sector data.

All OECD countries and most partner economies had in 2018 one or more initiatives to enhance access to and sharing of data in their economies. The scope of these initiatives may vary significantly, however. While all these countries have initiatives that foster and enhance access to and sharing of public-sector data, significantly fewer countries target private-sector data, for instance.

The analysis of current policy trends related to enhanced access and sharing in this chapter is based on a policy questionnaire, which was conducted between June and September 2018 and covered 20 countries¹ plus the European Union. This survey was complemented by the responses to the OECD (2017^[2]) *Digital Economy Outlook (DEO) Policy Questionnaire*, which included an additional 16 countries,² many of which are partner economies. As a result, a total of 205 policy initiatives across 37 countries were analysed (Figure 5.1).

These policy initiatives revealed the following four priorities: i) enhancing access to and sharing of public-sector information/data; ii) facilitating data sharing within the private sector; iii) increasing data analytic capacities across society; and iv) developing national data strategies. These four policy action areas are discussed in the following sections.

Figure 5.1. Number of government policy initiatives enhancing data access and sharing



Notes: PSI = public sector information.

Source: This figure is based on two country surveys, the most recent of which was conducted between June and September 2018 and covered 20 countries plus the European Union. This survey was complemented by the responses to the OECD (2017^[2]) *Digital Economy Outlook (DEO) Policy Questionnaire*, which included additional 16 countries. As a result, a total of 205 policy initiatives across 37 countries were analysed.

Governments leading by example in enhancing access to and sharing of public-sector data

The large majority of government initiatives focus on access to and sharing of public-sector data (almost 65% of all initiatives), with the majority of these initiatives aiming at enabling open access to government data. There was also a noticeable trend towards facilitating data sharing within the public sector (almost 15% of all initiatives on public-sector data). Opening geospatial data (e.g. maps) and transportation data ranked high on the agenda of public-sector data initiatives (representing almost 8% of the initiatives). The following subsections discuss each of these initiatives in more detail, by presenting a selection of related government initiatives.

Access to open government data and public-sector information

As highlighted in Chapter 2, open access to public-sector data (open government data) is the most prominent approach used to enhance access data, and to public-sector data in particular (OECD, 2015^[3]; Ubaldi, 2013^[4]; Vickery, 2012^[5]). Even before the emergence of open data initiatives such as data.gov (United States), data.gov.uk (United Kingdom), data.gouv.fr (France), or data.go.jp (Japan), it was recognised that public-sector data should be provided “at the lowest possible cost, preferably at no more than the marginal cost” as stated in the OECD PSI Recommendation (OECD, 2013^[6]). This motivated the establishment of public sector information (PSI) initiatives, which in many countries were legally backed by freedom of information legislations, and therefore were broader in scope compared to open data initiatives.³

As a result, many countries have PSI initiatives, while others have open data initiatives or both. This is the case, in particular, for EU member states which are subject to the Public Sector Information Directive (Directive 2003/98/EC), currently under revision (Box 5.1).

That said, a general trend towards the establishment of open data portals can be observed across all OECD countries.

In Canada, the government launched its Open Government Portal in 2014. The objective is to offer a one-stop access to information provided by departments. The Open Government initiative provides greater access to government data and information to the Canadian public and the businesses community through an “open by default” policy. The objective is to maximise the release of government data of value to support transparency, accountability, citizen engagement, and socio-economic benefits. This initiative is ongoing and evolving.

Mexico’s Open Data Initiative was established in 2015. The initiative mandates to make all public data available to citizens under open standards via the website datos.gob.mx. To date, the portal has almost 16 000 data sets from 217 public-sector institutions in Mexico. This initiative is backed by a strong regulatory framework that has institutionalised open data at the national, federal and local level. According to the OECD’s *Open Data Review of Mexico* (OECD, 2018^[7]), “the annual net sales [firms using open data is] between USD 28 000 and USD 1.4 million suggesting – and highlighting – the importance of providing support to SMEs for the data-driven economy in Mexico”.

In Denmark, the Agency for Digitisation under the Ministry of Finance have allocated DKK 30 million (USD 4.5 million) between 2016 and 2020 to establish a common architecture for the sharing and re-use of public-sector data, including the commercial use of the data. The objective is to i) establish a better overall picture of existing open data; ii) promote efforts to ensure access to more open data; iii) promote the use of open data, including by ensuring clear framework conditions for use; as well as iv) draw attention to the value of using data, for example through hackathons and the establishment of a public-private data space (Agency for Digitisation [Denmark], 2017^[8]).

Box 5.1. The Revision of the EU Directive Public-sector Information

The Directive on the re-use of public-sector information (PSI), also known as the PSI Directive, focuses on the economic aspects of the re-use of PSI rather than on access to PSI by citizens. It addresses material held by public-sector bodies in the member states, at national, regional and local levels, such as ministries, state agencies and municipalities, as well as organisations funded mostly by or under the control of public authorities (e.g. meteorological institutes).

The Directive covers written texts, databases, audio files and film fragments; it does not apply to the educational, scientific and broadcasting sectors. Since its 2013 revision (European Union, 2013^[9]), content held by museums, libraries and archives also falls within the Directive's scope of application.

On 25 April 2018 the European Commission adopted a proposal for a revision of the Directive, which was presented as part of a package of measures aiming to facilitate the creation of a common data space in the European Union (European Commission, 2018^[10]).

The proposal aims to overcome the barriers that still prevent the full re-use of PSI, which according to the impact assessment include the following:

- Data generated by the utilities and transport sectors has tremendous re-use potential. Yet entities active in these sectors are not covered by the PSI Directive.
- Dynamic data is one of the most commercially valuable types of data. However, the provision of real-time access to this data, for example using application programming interfaces (APIs), is rare.
- Several public-sector bodies continue to charge well above what is needed to cover reproduction and dissemination costs for the re-use of public-sector data.
- Public data holders sometimes enter into arrangements with the private sector to derive value from their data, which creates the risk of lock-in of public-sector data.

In respect of the continued existence of these barriers, the changes proposed are to:

- Reduce market entry barriers, in particular for small and medium-sized enterprises (SMEs), by limiting the exceptions that allow public bodies to charge for the re-use of their data more than the marginal costs of dissemination.
- Increase the availability of data by bringing new types of public and publicly funded data into the scope of the Directive, such as data held by public undertakings in the utilities and transport sectors and research data resulting from public funding.
- Minimise the risk of excessive first-mover advantage, which benefits large companies and thereby limits the number of potential re-users of the data in question, by requiring a more transparent process for the establishment of public-private data arrangements.
- Increase business opportunities by encouraging the dissemination of dynamic data via APIs.

Sources: European Commission (2018^[11]), *Proposal for a revision of the Public Sector Information (PSI) Directive*, <https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive>; European Commission (2018^[12]), *European Legislation on the Re-use of Public Sector Information*, <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>.

Norway has formulated a number of Strategy & Action plans for opening up public-sector data in five prioritised sectors, including culture, finance, research, and geo- and transport data. It is currently in the process of evaluating the need to open up public-sector data in other sectors.

Facilitating data sharing within the public sector

There is a noticeable trend towards facilitating data sharing within the public sector. This trend is motivated by governments' commitment to become more data-driven and to take advantage of technological trends such as big data and artificial intelligence (AI). Australia's data-sharing and release legislation (DS&R legislation) (presented in Box 4.1) and the United Kingdom's Data Ethics Framework and Centre for Data Ethics and Innovation (presented in Box 4.2 in Chapter 4) are examples that facilitate data sharing within the public sector.

Another example is Estonia's Information Sharing Data Sheet (X-Road) initiative. The objective of X-Road is to facilitate data exchange and linkage by inter-connecting the main national databases in Estonia. X-Road enables citizens, government agencies and private-sector organisations to securely use the majority of Europe-wide data that are registered in national registries. It is motivated by the "once-only" principle according to which public agencies should only collect data that is not previously maintained in any other public-sector databases. In other words, if a company or an individual has already submitted data to the public sector, he/she should not be forced to do it twice. At the same time, X-Road allows to verify the quality of the data, which is possible because public and private-sector institutions can connect their information system to X-Road (Information System Authority [Estonia], 2019^[13]).

In Israel, the government has also adopted an "only once policy" with the purpose of improving government public services and reducing bureaucratic load when sharing and re-using data. This initiative is driven by the ICT Authority and the Digital Israel Bureau in the Ministry for Social Equality of Israel. The initiative complements the government's resolution on open data, which allows government databases to be open to the public to encourage technological innovation in the public sector. This initiative was funded via a one-time budget of ILS 8 million (USD 2.2 million) spread over the years 2017 and 2018, for the purpose of creating the secure technological infrastructure and a designated budget of ILS 15 million (USD 4.1 million) as part of the base budget for incentivising government ministries to transfer data.⁴

In 2018, Canada started the development of its national data-governance framework. The objective of the framework is to: i) facilitate the use of data analytics and data for evidence-based solutions and improvements to public services and service delivery; ii) improve access to services through a single identifier; and iii) map new data back to existing data sources. This initiative is to be seen in the context of other initiatives in Canada such as Canada's National Data Strategy (see section "Achieving greater policy coherence through national and sectoral data strategies" below) and the initiative on Service Transformation of Employment and Social Development Canada (ESDC). ESDC identified an inherent need to have a solid foundation to support data access and data sharing in order to achieve desired improvements to service delivery for clients and for more efficient processing and reporting on policy outcomes to Canadians.

Similar initiatives are found in other countries, such as in Italy, where the government has developed a Data & Analytics Framework with the objective of improving and simplifying the interoperability and exchange of data between public administrations, promoting and improving the management and usage of open data, and optimising activities of analysis and knowledge generation. Begun in 2016, this initiative is led by the Digital Transformation Team under the Presidency of the Council of Ministers.

In Colombia, the National Planning Department, in close co-operation with iNNpulsa Colombia, the Management Unit of Business Growth of the Colombian government, is engaging with the Massachusetts Institute of Technology to develop a nation-wide Big Data Strategy for the Colombian government. The objective is to have a general architecture and some pilot projects that showcase the use and benefits of big data analytics for the public sector. USD 1.7 billion are foreseen to be spent over the course of this project.

In 2018, the People's Republic of China (hereafter "China") established its National Data Center for Governance. More than 20 government departments and agencies are participating in this initiative, which started in 2016. The objective of the initiative is to capture long-term activities in all areas that are relevant for assuring people's livelihood security, such as credit, transportation, medical, health, employment, government data, culture and education.

Geospatial and transportation data: A highly valued public-sector data

Geospatial data (geo-data) provide information about specific geographic locations. They are typically used for geographic information systems.⁵ The most prominent example are digital maps, but geo-data may also include data on addresses, cadastral parcels, administrative units, geology, and agriculture and aquaculture facilities, to name just a few. Transportation data can also be considered geo-data, to the extent that geolocation information are covered by the data. This is the case with data on traffic flows and data on public transportation schedules.

The combination of all these data has become the foundation for many location-based services provided by Internet service platforms such as Google, Microsoft, Uber and Waze. The data is also recognised as critical for the functioning of multimodal transport. This may explain why many countries have classified geospatial and/or transportation data among their high value data sets.

Australia, for instance, has classified its Geocoded National Address File (G-NAF) as one of its most high-valued data sets ("National Interest Datasets").⁶ G-NAF provides information on addresses street addresses in Australia including the state, suburb, street, number and co-ordinates reference (or "geocode") and contains more than 13 million Australian physical address records. What makes G-NAF so valuable is ability to be used for data linkage to other data sets including Census data from the Australian Bureau of Statistics. This has made G-NAF a kind of standard data set for data integration.

In Switzerland, the Federal Office of Transport is looking into ways to facilitate the exchange of data between the various public and private actors active in the Swiss public transport system. It is therefore focussing on geo-data, price data of transportation services and operational data. Measures considered target the availability of these data and the opening of distribution systems that could be put in place. Action plans will be submitted to the Federal Council by the end of 2018. Regulatory issues are looked into and public consultation on regulatory adaptations is planned for 2019.

In EU member states, access to geospatial data is subject to the Directive 2007/2/EC on establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Based on this Directive, EU member states have established national geoportals. This is, for example, the case with Norway's national geoportal "geonorge.no", the national website for map data and other location information (spatial data) in Norway. In Italy, the Ministry of Transport and Infrastructures has made all data concerning infrastructures and transports available in data catalogues and portals. This also includes data on the number of traffic accidents, public expenditure for transport and strategic infrastructures.

Some countries have gone further by actively supporting the re-use of geospatial and transportation data for “smart” applications. In Germany, the government has established the research initiative mFUND, to support the development of data-based business models for smart mobility (Mobility 4.0). The goal is to provide significant impulses for digital innovations in the transport sector. Innovative applications will be developed on the basis of already available or newly acquired data. A central aspect for the programme is the provision of mobility and geo-data (e.g. transport and traffic data, hydrological data, climate and weather data). For this purpose, data access and sharing are promoted according to open data principles and technically supported by the creation of a central, open data access point for mobility-related data (mCLOUD). This initiative is funded by the German Federal Ministry of Transport and Digital Infrastructure with EUR 150 million to be invested between 2016 and 2020. More than 100 projects were already funded with more than 300 individual project partners, including enterprises, research institutions and universities (Federal Ministry of Transport and Digital Infrastructure [Germany], n.d.^[14]).

Facilitating or regulating data access and sharing within the private sector

Fewer countries had initiatives to facilitate data sharing within the private sector (almost 15% of all initiatives), although sharing and re-use of private-sector data was the most frequently cited emerging challenge (followed by public-private partnerships [PPPs]) among the countries that responded to the EASD Policy Questionnaire. A majority (around 55%) were voluntary initiatives. Among those that were mandatory, most focussed on data-sharing agreements that were restricted to trusted users (restricted data sharing). These included initiatives promoting data sharing i) between the private and public sector with a focus on “data of public interests”; or ii) within network industries such as transportation and energy for ensuring the interoperability of smart services. Data portability with a focus on consumer personal data was another means for promoting access and sharing in the private sector. These different types of initiatives are discussed in the following subsections.

Voluntary and collaborative approaches

Voluntary approaches to data access and sharing tend to be used where the risks of detrimental consequences of mandatory access and sharing outweigh the expected public benefits (see subsection “The risks of mandatory access to data” in Chapter 4). This is the case where data access regulation would undermine incentives to invest in data, or where such regulation would not be granular enough to do sufficiently justice to important specificities, and as a result would reduce innovation and competition.

Against these risk, and to incentivise and co-ordinate actions that facilitate data access and sharing in the private sector, many governments have put in place incentives for voluntary initiatives. Two major types of voluntary government-led initiatives have been identified through the survey: i) contract guidelines; and ii) data partnerships, including PPPs.

Contract guidelines and principles

Contract guidelines define a set of contractual clauses based on defined principles. They constitute the default position that parties can consider when negotiating their data-sharing agreements, with a focus on potentially contentious issues to be addressed. Because they are voluntary, parties are free to deviate from the proposed contractual clauses at their will (freedom of contract). It is typically expected that parties would do so if such deviation would better reflect their common interests and the specific context of their data-sharing agreements. However such deviation would have to be justifiable, which is why contract

guidelines are seen as promising means to assure fair terms and conditions for data access, sharing and re-use, in particular where there is significant power and information asymmetries between parties (including individuals, as well as private and public-sector actors). Because they are based on agreed principles and refer to applicable national and international laws, contract guidelines are expected to also reduce legal transaction costs.

Japan's Ministry of Economy, Trade and Industry has formulated the Contract Guidance on Utilisation of AI and Data, which is a good example and one of the first of its kind. This Guidance summarises the issues and factors to be considered when drafting a contract on the utilisation of AI or data. It is intended to be used as a reference when private businesses conclude contracts related to data re-use or development and use of AI-based software. It differentiates between three different types of data utilisation contracts: i) data provision contracts; ii) data creation contracts; and iii) data-sharing (platform) contracts. The Guidance provides in-depth explanations and illustrates actual examples of data utilisation contracts. It also explains basic concepts concerning rights and responsibilities associated with development and utilisation of AI-based software and the use of data. This includes main legal issues and proper contract preparation processes for each contract type.⁷

Another government initiative is the Netherlands' Dare-2-Share Co-operation Agreement. This initiative aims at helping entrepreneurs "establish agreements in an honest and reliable way in the 'collaboration in innovation' phase – where data are shared between large and small companies". The initiative defines legal standards and references to national and international laws and regulations that parties need to take into account in their agreements. Particular attention is put on the relation between small players and larger entities. "Just like consumers enjoy some level of protection after they have clicked 'I agree' without having read the numerous pages of conditions in English, a Dare-2-Share arrangement must offer small parties some security that they do not forfeit all their rights." Overall, Dare-2-Share is expected to significantly reduce the amount of time spent for setting out the starting points for data-sharing agreements.

In the United States, the American Farm Bureau Federation (AFBF), together with commodity groups, farm organisations, and agriculture technology providers (ATP), helped establish the Privacy and Security Principles for Farm Data. It addresses controversial issues related to the ownership of agricultural data (Box 5.2). As of 1 April 2016, 37 organisations have signed onto the Core Principles, pledging to incorporate them into their contracts with farmers. To verify compliance with the Core Principles, AFBF and the other interested stakeholder groups formed a non-profit organisation, Ag Data Transparency Evaluator, to audit companies' agricultural data contracts and to certify these companies with the Ag Data Transparent seal of approval (Ag Data Transparent, 2016^[15]).

The European Commission in its Communication "Towards a common European data space" (European Commission, 2018^[10]), has proposed to develop Guidance on Private-sector Data Sharing (European Commission, 2018^[16]). The proposal is motivated by the observation that:

Manufacturers of IoT [Internet of Things] objects are usually in a privileged position to determine access to and re-use of the non-personal and automatically generated data from IoT objects. Depending on the market, these manufacturers may or may not grant access and usage rights to the user of the object, who may find him/herself prevented from using data, the generation of which he/she triggered (European Commission, 2019^[17]).

Box 5.2. The Core Principles of Ag Data

Ownership: “We believe farmers own information generated on their farming operations. However, it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders with an economic interest, such as the tenant, landowner, co-operative, owner of the precision agriculture system hardware, and/or ATP. The farmer contracting with the ATP is responsible for ensuring that only the data they own or have permission to use is included in the account with the ATP.”

Collection, Access and Control: “An ATP’s collection, access and use of farm data should be granted only with the affirmative and explicit consent of the farmer. This will be by contract agreements, whether signed or digital.”

Notice: “Farmers must be notified that their data is being collected and about how the farm data will be disclosed and used. This notice must be provided in an easily located and readily accessible format.”

Transparency and Consistency: “ATPs shall notify farmers about the purposes for which they collect and use farm data. They should provide information about how farmers can contact the ATP with any inquiries or complaints, the types of third parties to which they disclose the data and the choices the ATP offers for limiting its use and disclosure.”

Portability: “Within the context of the agreement and retention policy, farmers should be able to retrieve their data for storage or use in other systems, with the exception of the data that has been made anonymous or aggregated and is no longer specifically identifiable. Non-anonymised or non-aggregated data should be easy for farmers to receive back at their discretion.”

Disclosure, Use and Sale Limitation: “An ATP will not sell and/or disclose non-aggregated farm data to a third party without first securing a legally binding commitment to be bound by the same terms and conditions as the ATP has with the farmer. Farmers must be notified if such a sale is going to take place and have the option to opt out or have their data removed prior to that sale. [...] If the agreement with the third party is not the same as the agreement with the ATP, farmers must be presented with the third party’s terms for agreement or rejection.”

Data Retention and Availability: “Each ATP should provide for the removal, secure destruction and return of original farm data from the farmer’s account upon the request of the farmer or after a pre-agreed period of time. The ATP should include a requirement that farmers have access to the data that an ATP holds during that data retention period. ATPs should document personally identifiable data retention, and availability policies and disposal procedures, and specify requirements of data under policies and procedures.”

Unlawful or Anti-Competitive Activities: “ATPs should not use the data for unlawful or anti-competitive activities, such as a prohibition on the use of farm data by the ATP to speculate in commodity markets.”

Liability and Security Safeguards: “The ATP should clearly define terms of liability. Farm data should be protected with reasonable security safeguards against risks such as loss or unauthorised access, destruction, use, modification or disclosure. Policies for notification and response in the event of a breach should be established.”

Source: (Ag Data Transparent, 2016^[15]), *Ag Data’s Core Principles: The Privacy and Security Principles for Farm Data*, www.agdatatransparent.com/principles/.

To address these issues, the European Commission (2018^[10]) articulates a set of principles that are expected to be reflected in contractual agreements to ensure fair and competitive markets related to “non-personal machine-generated data” (Box 5.3). Additional principles have been articulated to support the supply of private-sector data to public bodies under preferential conditions for re-use (see subsection “Data of public interest” below).

Box 5.3. Proposed EC Principles for Contractual Agreements on Non-Personal Data

Transparency: The relevant contractual agreements should identify in a transparent and understandable manner i) the persons or entities that will have access to the data that the product or service generates, the type of such data, and at which level of detail; and ii) the purposes for using such data.

Shared value creation: The relevant contractual agreements should recognise that, where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.

Respect for each other’s commercial interests: The relevant contractual agreements should address the need to protect both the commercial interests and secrets of data holders and data users.

Ensure undistorted competition: The relevant contractual agreements should address the need to ensure undistorted competition when exchanging commercially sensitive data.

Minimised data lock-in: Companies offering a product or service that generates data as a by-product should allow and enable data portability as much as possible. They should also consider, where possible and in line with the characteristics of the market they operate on, offering the same product or service without or with only limited data transfers alongside products or services that include such data transfers.

Source: European Commission (2018^[10]), “Towards a common European Data Space”, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN>.

Data (sharing) partnerships including PPPs

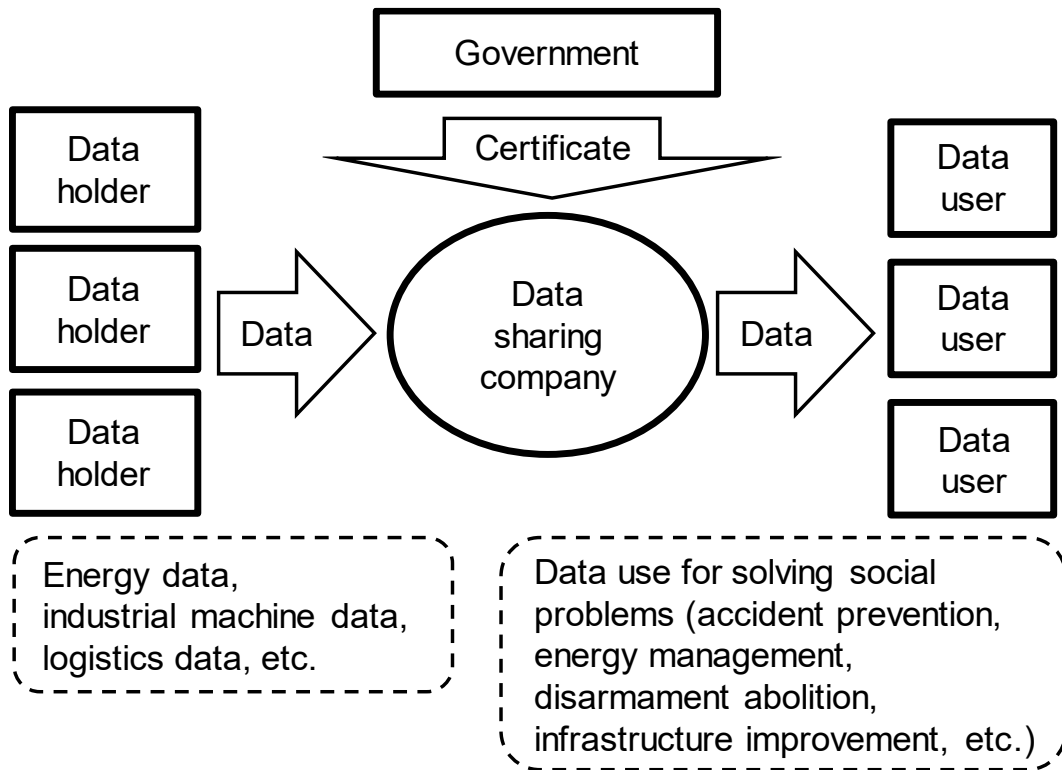
As highlighted in Chapter 2 (see subsection “Other restricted data-sharing arrangements”), data partnerships enable organisations to share and mutually enrich their data sets, including through cross-licensing agreements. A number of governments have focussed on encouraging the establishment of data partnerships within the private sector and/or between the private and public sector. Many of these initiatives are enabled by open access to public-sector data. In Chile, for instance, the government has engaged in academic agreements on open data with academic and research institutions for the re-use of data in open format.

Other data partnerships are incentivised through research-related funding. An example is the Industrial Data Space (IDS) research project, a platform for the commercialisation of data in a business-to-business context. IDS was funded EUR 5 million by the German Ministry of Education and Research between 2015 and 2018, and co-ordinated by the Fraunhofer Institute.

In some other initiatives, the governments’ role has been to incentivise and orchestrate data partnerships either by acting as independent trusted third parties or by engaging with the private sector in PPPs. The Data Integration Partnership for Australia is an example of a government acting as a trusted third party: it is “an investment to maximise the use and value of the Government’s data assets” (Department of the Prime Minister and Cabinet (Australia),

2017^[18]). Another example is Japan's Certification System for data-sharing platforms that support companies that want to share their data (Figure 5.2). This system includes a data request system, i.e. a system that allows data-sharing companies to request data that have been provided to relevant ministries and agencies. The government also provides support through tax incentives and administrative guidance, in particular. It can also revoke accreditation in some cases.

Figure 5.2. Certification system for data-sharing platforms



Source: Slide presented by Naoto Ikegai (Interfaculty Initiative in Information Studies, University of Tokyo).

Digital Hub Denmark is an example of a data PPP, where both public and private-sector actors agree to share their data. It is a PPP between the government, the Confederation of Danish Industry, the Danish Chamber of Commerce and Finance Denmark. The partnership aims to make Denmark one of the main European tech-hubs in AI, Internet of Things and big data. The Digital Hub will improve companies' access to talent and investments, and facilitate the matchmaking between larger companies, start-ups and universities. Access to data thus constitutes just one element of the overall objective of the partnership.

Data portability

The concept of data portability is discussed in Chapter 2 in the context of the General Data Protection Regulation (GDPR). However, there are a number of other data portability initiatives, some of which existed prior to the GDPR. These initiatives sometimes differ in important ways.

The United States, for example, initiated a number of My Data initiatives back in 2010. They aimed at facilitating consumers' access to their own personal data in particular sectors. Some

of these initiatives concerned data held by public bodies: for instance, the Get Transcript initiative streamlines access to data held by the United States Internal Revenue Service (IRS, 2018^[19]) while My Student Data provides access to federal student data.⁸ Other initiatives encourage private enterprises to give consumers access to their data. For example, the Green Button initiative allows consumers to access their electric utility data and this opportunity is now offered to over 60 million homes and businesses in the United States (Honey, Chrousos and Black, 2016^[20]). Blue Button is a hybrid initiative that applies to public and private-sector health care providers. It seeks to expand patients' access to their medical records. As of March 2016, Blue Button has enabled more than 3 million Medicare beneficiaries, service members and veterans to access their government-held medical records. An estimated 150 million Americans will be able to access their health records from private-sector health care stakeholders, including health professionals and retail pharmacy chains, through Blue Button.⁹

In 2011, The United Kingdom introduced its Midata data portability initiative, which was renamed Mydata as part of a broader consumer empowerment strategy (Department for Business Innovation and Skills (UK), 2011^[21]). Midata seeks to give consumers access to the electronic information that companies hold about their transactions in a machine-readable and portable format. This transaction data includes information collected on an individual's browsing history and purchases when logged into a particular website (Department for Business Innovation and Skills (UK), 2012^[22]). However, purchases made on a guest account entailing no user registration, or information about complaints or other such communication with service providers would not constitute individual transaction data. The Midata initiative focuses on three sectors: energy supply; the mobile phone sector; and the financial sector (current accounts and credit cards). Rather than legislating to introduce this data portability obligation, the government preferred to take a power pursuant to the Enterprise and Regulatory Reform Act 2013 (HMSO, 2013^[23]). This allows the Secretary of State to introduce regulations to make Midata compulsory if the government is dissatisfied with the progress made in these sectors on a voluntary basis.¹⁰

In Australia, the government is passing legislation on a data portability right, the Consumer Data Right (CDR), to give consumers the right to safely access certain data about them held by businesses. They will also be able to direct that this information be transferred to accredited, trusted third parties of their choice. The particularity of the CDR, compared to other data portability rights, is that it also grants SMEs the right to data portability. The CDR will be implemented first in the banking, energy, and telecommunications sectors, and then rolled out economy-wide on a sector-by-sector basis. Regulations will provide for the designation of certain data sets. This initiative, which is led by Australia's Treasurer with the Australian Competition and Consumer Commission as lead regulator, the Office of Australian Information Commissioner for privacy issues, and Data61 for issues on technical standards, is funded with AUD 45 million for a period of four years starting in 2018.

Data of public interest

A number of countries have started to specify a new class of data, which is often referred to as **data of public interest**. The scope of this class varies significantly across countries, however. In some countries, data of public interest explicitly refers to private-sector data (of public interest), while in others it refers to public-sector data. Sometimes both private and public-sector data as well as personal and non-personal data are included.

Australia, for instance, is considering the establishment of a framework to identify National Interest Datasets or designated data sets (Department of the Prime Minister and Cabinet

(Australia), n.d.^[24]; Department of the Prime Minister and Cabinet [Australia], 2018^[25]). These data sets would primarily include public-sector data but may also include private-sector data controlled by the public sector under certain conditions. France's (2016^[26]) Law for a Digital Republic (*Loi pour une République numérique*) defines data of general interest (*données d'intérêt général*) as including: i) private-sector data from delegated public services such as utility or transportation services; ii) private-sector data that are essential for granting subsidies; and iii) private-sector data needed for national statistics. And under the concept of "private-sector data for public interest purposes", the European Commission is examining data sharing between the private and public sector in order to guide policy making and improve public services (European Commission, 2018^[10]).

A strong commonality of these initiatives is that the data are needed to fulfil well-defined societal objectives that otherwise would be impossible or too costly to fulfil. These include the development of national statistics, the development and monitoring of public policies, the tackling of health care and scientific challenges of societal importance and in some cases the provision of public services. In this sense, the concept of data of public interest was already articulated in the OECD's Recommendation on Health Data Governance, which recognises that "effectively safeguarding the public interest is an important function of governments".

Data of public interest are typically intended to be used mainly by governments or public-sector institutions. However, in some cases, access to data is regulated based on competition and efficiency considerations. This is particularly true for network industries such as telecommunications, energy and transport (Box 5.4).

Box 5.4. Fostering interoperability through enhanced access and sharing: The case of Finland's (2018) Act on Transport Services

Finland's (2018) Act on Transport Services is a three-stage legislative project that aims to streamline all transport market regulations into one package. The Act introduces significant changes to transport markets that have so far been strictly regulated and steered by public measures. It promotes customer-oriented, market-based transport services on the basis of sound competition. The Act's goals are twofold: firstly, through deregulation it gives more room to develop innovative, digitally enabled services. Secondly, it obliges all service providers to open certain essential data to all as well as ticketing and payment APIs for single trip/ticket to third parties. The Act makes it possible to examine transport as a whole, i.e. as one service.

The Finnish Act on Transport Services is built on the fact that future transport will rely on open access to necessary data, the interoperability of information and information systems through APIs as well as the openness of these interfaces. As of end of 2018, around 5 200 companies in the Finnish transportation sector have made their data available, mostly via APIs, since the adoption of the Act. Current estimates suggest that this covers around 80% of transportation services used in Finland. These include taxi services (with more than 1 400 data sets), on-demand transportation services (around 400 data sets), timetable-bound public transportation services (around 240 data sets), rental services and commercial car-sharing services (around 20 data sets), and commercial parking services. In addition, the most important actors have opened their ticketing and payment system APIs, in particular those operating within the largest cities.¹

1. To support the interoperability of ticketing and payment system APIs, the Lippu Network was established.

Increasing data analytic capacities across society

Increasing data analytic capacities, either in the public or private sector, was addressed by only 12% of all policy initiatives. A quarter of those initiatives focussed on the establishment of data analytic technology centres that provide support and/or guidance in the re-use and analysis of data for public and/or private-sector entities. Some have also supported investments in data-related innovation and research and development (R&D). The following subsections discuss these three types of initiatives respectively, with the caveat that some initiatives address all three aspects such as Colombia's Excellence and Appropriation Centre (CEA) in Big Data Analytics (Box 5.5).

Box 5.5. Colombia's Excellence and Appropriation Centre in Big Data Analytics

The CEA in Big Data Analytics, initiated and supported by the Colombian government since 2015, reunites three key stakeholders: i) high-ranking universities; ii) information and communication technology (ICT) companies which are leaders in big data and IoT technologies; and iii) non-ICT companies which are national leaders in their respective economic sectors. As of today, this CEA now counts 11 organisations.

The CEA drives the development of four aspects which are expected to increase data analytic capacities in Colombia:

1. **Training:** Generation of research capabilities with R&D teams in universities that specialise in big data and IoT respectively. Training decision makers to understand ICT as strategic business assets.
1. **Applied R&D:** Emphasise research capabilities in solving real and proven problems as well as the seizing of business development and creation opportunities. The pairing of non-ICT sector problems and opportunities with ICT-based solutions through analytics and IoT, thus strengthening the position of ICT in strategic business processes.
2. **Entrepreneurship:** Take advantage of knowledge transfer and development to further sectors' development with new ICT-based companies: start-ups and spin-offs.
3. **Innovation and appropriation of ICT:** Ensure use and appropriation of ICT solutions at decision-making business levels: positioning the chief information officer as a relevant actor of business strategy. Strengthen the competitiveness of firms with ICT so they can be leaders in their respective sectors in a digital economy.

USD 1.6 million have already been spent between 2015 and 2016 for the initial development of the CEA in big data and data analytics. Additional USD 1.2 million have been spent each year since 2017.

Supporting the development of data-related skills and infrastructures

Governments have recognised that the availability of data-related skills and competences can constitute a critical bottleneck for the effective re-use as well as provision of data in the private as well as public sector. Some have established dedicated initiatives to support the development of data-related skills and infrastructures.

The United Kingdom, for example, has a number of initiatives aimed at supporting the development of skills in the private and public sector. The Digital Skills Partnership, for instance, brings together public, private and charity sector organisations to boost skills for

a world-leading, inclusive digital economy. Besides initiatives related to data ethics and AI such as the United Kingdom Government Data Ethics Framework and the Centre for Data Ethics and Innovation, the United Kingdom has established a Data Skills Taskforce with the help of the Department of Digital, Culture, Media and Sport, public bodies like The Tech Partnership and private partners such as Accenture to enhance data analytic skills in the workforce.

Estonia's digital solutions seminars, as another example, target industrial companies that are keen to make their production more efficient through the use of data. The aim is to improve knowledge and skills having to do with the collection and use of data and information. The initiative is funded with EUR 200 000 between 2017 and 2020.

China's Ministry of Education has supported the development of data-related skills through data analytics competitions together with Internet firm Alibaba. This data analytic competition has been conducted every year since 2010 and has helped Alibaba and the government identify the most talented data scientists in China.

There is also a significant share of initiatives that address public servants. An example is Slovenia's education and training programmes to increase data-related skills and competencies among public servants, which is funded by Slovenia's Ministry of Public Administration since 2016.

Many of the skills-related initiatives are complemented by, or integrated with, initiatives that establish data infrastructure and/or data analytic support centres as discussed in the next subsection. Austria's initiatives for the provision of a big data infrastructure and technology foundation, for instance, covers both the support of data-related know-how and infrastructural support. This includes support for the development and provision of statistical and analytical methods and tools with the help of data scientists for different target groups (supporting outcome monitoring and measuring cross indicators out of heterogeneous data sources).

In Ireland, the Department of Education and Skills and the Higher Education Authority have offered Springboard ICT Conversion Courses, which increasingly focus on data analytics. This Springboard offers free, part-time and intensive conversion courses in higher education from certificate to degree to postgraduate level. All courses are in enterprise sectors which are growing and need skills personnel. This initiative has been funded since 2013 with EUR 28 million in 2016, the overall budget for Springboard, including courses other than data analytics. This initiative is complemented by the establishment of an Insight Centre for Data Analytics.

Establishing and collaborating with data analytic support centres

Some governments have established data analytic and innovation centres to support government agencies in the sharing and re-use of data. They have also created and strengthened partnerships with such centres. Ireland's Department of Jobs, Enterprise and Innovation opened the Insight Centre for Data Analytics in 2013. One of Europe's largest data analytics research organisations, its objective is to find solutions for issues related to connected health and the discovery economy. The Centre has more than 400 researchers, around 80 industry partners and over EUR 100 million in funding (Insight Centre for Data Analytics, n.d.^[27]).

Australia's data innovation centre, Data61, which is part of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO), has partnered with Australian government agencies to build new technologies that make high-value government data available to more people while preserving privacy. In close collaboration with partner agencies, Data61 has developed a suite of new tools and technologies to enhance open data access, data

sharing between agencies and managing privacy risks with sensitive data. In particular, the Confidential Computing Platform uses distributed machine learning – as well as homomorphic encryption and secure multi-party computing – to enable insights to be provided without organisations disclosing any data. This keeps the source data secure, private and up-to-date (Data61, n.d.^[28]).

The European Commission is working towards the development of a Support Centre for Data Sharing under the Connecting Europe Facility Programme. This Centre is expected to put in place a set of measures to facilitate data sharing in the private and public sectors. “It will offer know-how and assistance on data sharing by providing best practice examples and information on APIs, existing model contracts and other legal and technical aspects” (European Commission, 2017^[29]). This would include further improving the Guidance on Private-sector Data Sharing (European Commission, 2018^[16]) discussed above.

Supporting innovation and R&D in data analytics and related technologies

A number of countries have initiatives to support innovation and R&D in data analytics and related technologies. Many of these policy measures are part of more general initiatives to support the digital economy and innovation. Few dedicated initiatives exist that focus particularly on data analytics and data sharing.

The European Commission, for example, has put in place a number of funding schemes for data-related innovation. These include:

- The funding for data innovation incubators which connect data providers to data users. Three consortia composed of businesses and research organisations have been funded for three years with EUR 15 million.
- The funding of pan-European aggregators of PSI (European Data Portal), with the aim to develop common metadata catalogues of all PSI published in EU member states, searchable in multiple languages. This initiative has been funded with EUR 10 million since 2015 with an ongoing contract until 2020.
- The funding of privacy-enhancing technologies, including five consortia composed of businesses and research organisations, which are funded for three years with EUR 65.5 million.

Achieving greater policy coherence through national and sectoral data strategies

Some countries are in the process of developing national data strategies to assure the coherence and flexibility of their national data-governance frameworks. National data strategies can help address many of the policy issues discussed in this report in a comprehensive manner by incorporating a whole-of-government perspective. Similarly to national privacy strategies, national data strategies can be instrumental in creating the conditions for effective governance frameworks that better protect the private interests of individuals and organisations while providing the flexibility needed for all to benefit from data sharing and re-use.

The most prominent example of a national data strategy is Canada’s current proposal for a Government of Canada Data Strategy to support the management of data as a strategic asset, enabling improved services and informed decision-making. While still under development, key expected benefits of the strategy include:

- Enabling the government to be more purposeful about what data it wants to collect and how they will be used and re-used to support decision-making, service delivery or broader societal outcomes.
- Improving the enterprise-wide governance, oversight and stewardship of data.
- Ensuring that the intended uses of data held by the government are appropriate and ethical, and that individuals' confidentiality and privacy are protected.
- Allowing federal government organisations to better monitor the effectiveness of policies, programmes and/or services in order to adjust and recalibrate them.
- Ensuring that data is collected and deployed in such a way that more relevant, timely and actionable information can be used by ministers, departments, businesses, the not-for-profit sector and Canadians to improve decision-making and outcomes.
- Transforming service delivery to better meet the needs of citizens by providing seamless and easy-to-use online services.

In addition, the Canadian government is working with the private sector to assist businesses to use data in innovative ways to enhance the service experience for Canadians and safeguard their privacy.¹¹

Some national data strategies may have sector-specific elements or they may even focus on selected sectors such as health, transportation, energy, or the public sector. Some of these strategies may not be explicitly referred to as national data strategies, although they could be considered as such. National open data initiatives, for instance, such as Australia's DS&R legislation discussed in Box 4.1, can be considered national data strategies with a focus on public-sector data. As another example, the United States is developing a Federal Data Strategy (<https://strategy.data.gov/>) to provide a co-ordinated and integrated approach to using government data to deliver on mission, serve the public, and steward resources while respecting privacy and confidentiality. The strategy incorporates four areas of exploration:

1. **Enterprise Data Governance:** setting priorities for managing government data as a strategic asset, including establishing data policies, specifying roles and responsibilities for data privacy, security, and confidentiality protection, and monitoring compliance with standards and policies throughout the information lifecycle.
2. **Access, Use and Augmentation:** developing policies and procedures that enable stakeholders to effectively and efficiently access and use data assets by: i) making data available more quickly and in more useful formats; ii) maximising the amount of non-sensitive data shared with the public; and iii) leveraging new technologies and best practices to increase access to sensitive or restricted data while protecting privacy, security, and confidentiality as well as the interests of data providers.
3. **Decision-Making and Accountability:** improving the use of data assets for decision-making and accountability for the federal government, including both internal and external uses. This includes: i) providing high-quality and timely information to inform evidence-based decision-making and learning; ii) facilitating external research on the effectiveness of government programmes and policies which will inform future policy making; and iii) fostering public accountability and transparency by providing accurate and timely spending information, performance metrics, and other administrative data.
4. **Commercialisation, Innovation and Public Use:** facilitating the use of federal government data assets by external stakeholders at the forefront of making government

data accessible and useful through commercial ventures, innovation, or for other public uses. This includes use by the private sector and scientific and research communities, by state and local governments for public policy purposes, for education, and in enabling civic engagement. Supporting the production and dissemination of comprehensive, accurate, and objective statistics on the state of the nation helps businesses and markets operate more efficiently.

Another example includes national health data initiatives, which aim to facilitate sharing and re-use of health-related data such as electronic health record data to inform clinicians, monitor disease outbreaks, conduct research, and monitor and improve the overall quality of health care (OECD, 2015^[30]; OECD, 2015^[31]). Some of these strategies have significantly facilitated and institutionalised the linkage across existing health data sets, as discussed in Chapter 3.

References

- Ag Data Transparent (2016), *Ag Data's Core Principles: The Privacy and Security Principles for Farm Data*, <http://www.agdatatransparent.com/principles/>. [15]
- Agency for Digitisation [Denmark] (2017), *The digitally coherent public sector: White Paper on a common public-sector digital architecture*, https://arkitektur.digst.dk/sites/default/files/white_paper_on_a_common_public-sector_digital_architecture_pdfa.pdf. [8]
- Data61 (n.d.), *Confidential Computing – Insights from data without seeing the data*, <https://data61.csiro.au/en/Our-Work/Safety-and-Security/Privacy-Preservation/Confidential-computing> (accessed on 13 March 2019). [28]
- Department for Business Innovation and Skills (UK) (2012), *Midata: Government response to 2012 consultation*, BIS/12/1283, http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/34700/12-1283-midata-government-response-to-2012-consultation.pdf. [22]
- Department for Business Innovation and Skills (UK) (2011), “Better Choices: Better Deals – Consumers Powering Growth”, <http://www.gov.uk>. [21]
- Department of the Prime Minister and Cabinet (Australia) (2017), *Information about the Data Integration Partnership for Australia*, <http://www.pmc.gov.au/sites/default/files/publications/DIPA-information.pdf>. [18]
- Department of the Prime Minister and Cabinet (Australia) (n.d.), *Designated Datasets – a special class of high-value dataset*, <https://dataavailability.pmc.gov.au/designated-datasets> (accessed on 15 September 2018). [24]
- Department of the Prime Minister and Cabinet [Australia] (2018), *New Australian Government Data Sharing and Release Legislation: Issues Paper for Consultation*, <http://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation>. [25]

- European Commission (2019), *Guidance on private sector data sharing*, [17]
<https://ec.europa.eu/digital-single-market/en/guidance-private-sector-data-sharing> (accessed on 5 February 2019).
- European Commission (2018), *European legislation on the re-use of public sector information*, [12]
<https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information> (accessed on 1 October 2018).
- European Commission (2018), “Guidance on sharing private sector data in the European data economy”, *Commission Staff Working Document*, Accompanying the document, [16]
<https://ec.europa.eu/digital-single-market/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy>.
- European Commission (2018), *Proposal for a revision of the Public Sector Information (PSI) Directive*, <https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive> (accessed on 1 October 2018). [11]
- European Commission (2018), “Towards a common European Data Space”, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* COM(2018)232 final, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN>. [10]
- European Commission (2017), *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Building a European Data Economy”*, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2017%3A9%3AFIN>. [29]
- European Union (2013), *Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information*, [9]
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0037&from=FR>.
- Federal Ministry of Transport and Digital Infrastructure [Germany] (n.d.), *Was ist die mCLOUD?*, <http://www.mcloud.de/web/guest/informationen> (accessed on 5 February 2019). [14]
- Government of France (2016), *Loi pour une République numérique*, [26]
<http://www.senat.fr/leg/pjl15-744.html>.
- HMSO (2013), *Enterprise and Regulatory Reform Act 2013*, [23]
<http://www.legislation.gov.uk/ukpga/2013/24/contents>.
- Honey, K., P. Chrousos and T. Black (2016), *My Data: Empowering All Americans with Personal Data Access*, <http://www.whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access>. [20]
- Information System Authority [Estonia] (2019), *Data Exchange Layer X-tee*, [13]
<http://www.ria.ee/en/state-information-system/x-tee.html>.
- Insight Centre for Data Analytics (n.d.), <http://insight-centre.org> (accessed on 29 September 2019). [27]

- IRS (2018), *Welcome to Get Transcript*, <http://www.irs.gov/individuals/get-transcript>. [19]
- OECD (2018), *Open Government Data in Mexico: The Way Forward*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264297944-en>. [7]
- OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264276284-en>. [2]
- OECD (2016), “Health Data Governance Recommendation”, in *Recommendation of the Council on Health Data Governance*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433>. [1]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [3]
- OECD (2015), *Health Data Governance: Privacy, Monitoring and Research*, (policy brief), OECD, Paris, <https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdf>. [31]
- OECD (2015), “The evolution of health care in a data-rich environment”, in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-12-en>. [30]
- OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>. [6]
- Ubaldi, B. (2013), “Open government data: Towards empirical analysis of open government data initiatives”, *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5k46bj4f03s7-en>. [4]
- US Department of Education (n.d.), *MyStudentData Download*, <https://studentaid.ed.gov/sa/resources/mystudentdata-download> (accessed on 5 February 2019). [32]
- Vickery, G. (2012), *Review of recent studies on PSI reuse and related market developments*, http://www.scb.se/statistik/_publikationer/NR9999_2012A01_BR_X76BR1201.pdf. [5]

Notes

¹ Countries covered by the EASD Policy Questionnaire include: Australia, Austria, Canada, Denmark, Estonia, Finland, Germany, Italy, Japan, Korea, Latvia, the Netherlands, Norway, Portugal, Slovenia, Spain, Switzerland, Turkey, the United Kingdom and the United States.

² Additional countries covered by the OECD (2017^[2]) *DEO* Policy Questionnaire include: Belgium, Brazil, Chile, China, Colombia, Costa Rica, Czech Republic, France, Ireland, Israel, Lithuania, Luxembourg, Mexico, Poland, Singapore and Sweden.

³ PSI typically includes not only data but also digital content, such as (e.g.) text documents and multimedia files.

⁴ This does not include the ILS 1.5 million budget as part of the base budget starting 2019 towards the current costs of the Government's information backbone and a budget of ILS 1 million as part of the base budget starting 2017 for the acquisition of consulting services for the implementation of the "once only" policy, with an emphasis on the expansion of the policy to permits and data concerning corporations.

⁵ These include a database, geodatabase, shape-file, coverage, raster image, or dbf table.

⁶ As of end of 2018, G-NAF was downloaded more than 3 million times.

⁷ The AI Section then proposes "exploratory multi-phased" AI development processes which consists of (1) the assessment phase, (2) the PoC phase, (3) the development phase and (4) the retraining phase. The AI Section further describes types of contracts and factors to be considered in contract preparation, with sample clauses provided.

⁸ According to the Federal Student Aid website, the "MyStudentData Download button allows you to download your federal student grant and/or loan information or your FAFSA [Free Application for Federal Student Aid] information in a plain text file" (US Department of Education, n.d.^[32]).

⁹ These various initiatives allow individuals to access their personal information however, the "my Social Security" programme goes a step further by allowing individuals to create a free online account and to download a copy of their Social Security benefit statement. This benefit statement could then be used for financial planning purposes.

¹⁰ See Sections 89-91 of HMSO (2013^[23]), dealing with "supply of consumer data".

¹¹ The SecureKey Concierge Service, for instance, is a next generation authentication network for conveniently connecting people to government services online using a banking credential they already have and trust. SecureKey Concierge is configured to be "triple-blind", ensuring that no party unnecessarily receives sensitive or personal information from other parties.

Enhancing Access to and Sharing of Data

RECONCILING RISKS AND BENEFITS FOR DATA RE-USE ACROSS SOCIETIES

This report examines the opportunities of enhancing access to and sharing of data (EASD) in the context of the growing importance of artificial intelligence and the Internet of Things. It discusses how EASD can maximise the social and economic value of data re-use and how the related risks and challenges can be addressed. It highlights the trade-offs, complementarities and possible unintended consequences of policy action – and inaction. It also provides examples of EASD approaches and policy initiatives in OECD countries and partner economies.

This publication is a contribution to the OECD Going Digital project which aims to provide policymakers with the tools they need to help their economies and societies prosper in an increasingly digital and data-driven world.

For more information, visit www.oecd.org/going-digital

#GoingDigital



Consult this publication on line at <https://doi.org/10.1787/276aaca8-en>.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases. Visit www.oecd-ilibrary.org for more information.

