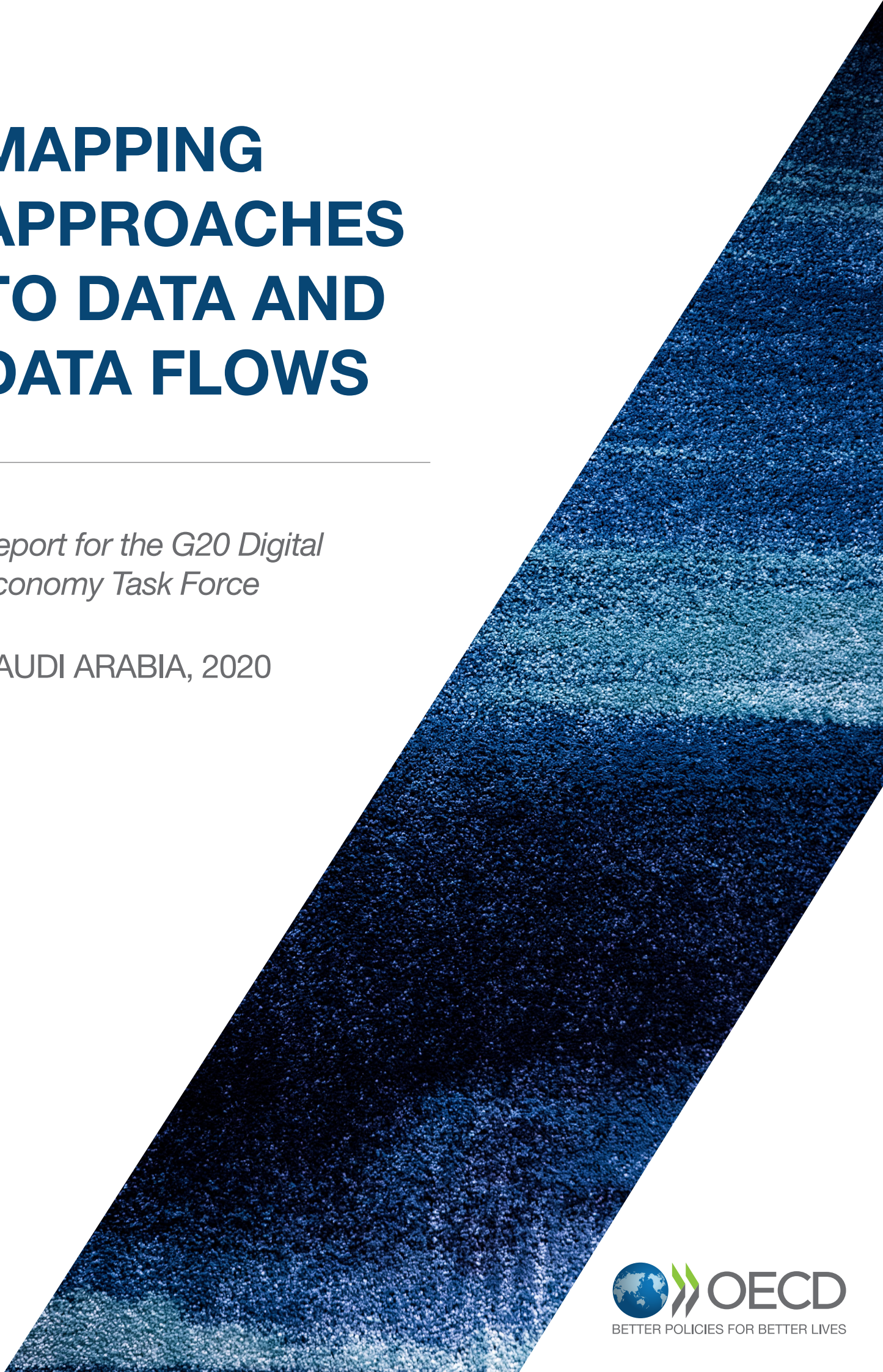
MAPPING APPROACHES TO DATA AND DATA FLOWS

Report for the G20 Digital Economy Task Force

SAUDI ARABIA, 2020



OECD

BETTER POLICIES FOR BETTER LIVES

This document was prepared by the Organisation for Economic Co-operation and Development (OECD) Directorate for Science, Technology and Innovation (STI) and Trade and Agriculture Directorate (TAD), as an input for the discussions in the G20 Digital Economy Task Force in 2020, under the auspices of the G20 Saudi Arabia Presidency in 2020. The opinions expressed and arguments employed herein do not necessarily represent the official views of the member countries of the OECD or the G20.

Contents

Executive Summary

In today's digitalised and globally interconnected world, data has become the lifeblood of economic and social interactions. Indeed, the proliferation of devices and sensors, the exponential growth in computing power, the plummeting costs of data storage and the growing ability to deliver more data at greater speed, have altered how we conduct our lives and how businesses operate. Data has transformed how people interact; altered the configuration of global value chains (GVCs); given rise to new information industries (e.g. cloud computing, 3D printing or the Internet of Things); changed how services are produced and delivered; and, even affected how we grow and trade food.

However, as we become increasingly reliant on data for daily activities, new concerns arise. These relate not only to issues of privacy and security but also to economic development or the need to protect intellectual property and maintain the reach of regulatory and audit bodies, especially in the context of data crossing different jurisdictions.

As a result, governments have started updating and adapting data-related policies to the digital age. For example, some measures restrict the movement of data across borders or mandate that data be stored locally. Digital infrastructures such as the Internet operate globally and although they offer new opportunities for people and countries around the world, they also raise considerable challenges for domestic and international policy in a world where borders and regulatory differences between countries remain.

In this evolving environment, it is increasingly clear that the benefits of the digital transformation for our economies and societies require trust in the activities of different players operating in the digital space, and the ability to scale economic and social activities globally. Individuals will not engage with businesses they do not trust and businesses will struggle to reap the benefits of scale unless they can operate globally. The challenge for governments is to create this environment of trust where regulatory objectives can be met, even when data is moved across jurisdictions – an environment of "data free flow with trust". These issues were discussed by the G20 during the 2019 Japanese presidency.

Against this backdrop, and in an effort to assist G20 members in their ongoing dialogue, this paper aims to strengthen the foundation for further G20 discussions by filling some existing information gaps and mapping the different issues at stake in this debate . Indeed, understanding this environment is a first step in promoting interoperable approaches that meet the dual goal of ensuring that a) data can flow across borders; and b) that it can do so with trust. This involves looking at new issues raised by data and data-sharing, identifying the evolving policies that affect the movement of data, and discussing approaches that have emerged to enable data transfers across jurisdictions.

The key takeaway lessons from this report include:

- Data is a critical resource, creating many benefits, sometimes in ways that cannot be foreseen when first generated. Its value is not necessarily intrinsic, it is contextual, and emerges only after processing and analysis. Data can also be easily copied, used and re-used by an unlimited number of users and for an unlimited number of purposes. This means that there are considerable increasing returns to scale and scope to be had from sharing data, including across borders.
- But data is not monolithic. Different data domains are subject to different data governance frameworks and overlapping domains can raise new challenges.
- Moreover, where data cross international borders issues related to privacy protection, regulatory control or audit, national security, data security, data integrity, economic development and digital industrial policies are heightened.
- These concerns have led countries to update their data-related policies, often placing conditions on the transfer of data across borders or requiring that data be stored locally.

- The multiplicity of applicable data regimes creates uncertainty for governments, businesses and individuals, including with regard to the applicable rules in a given situation. While there are legitimate reasons for diversity, it is important to alleviate possible tension and conflict between them.
- To this end, governments and other stakeholders are increasingly using a range of approaches to ensure that data can flow across borders with trust. These can help businesses find grounds for data transfers while ensuring that, upon crossing a border, data is granted the desired degree of protection.
- These approaches can be grouped into 4 broad categories. Traditionally, data flow policies were focused on the transfer of personal data and were discussed in *plurilateral arrangements* as might be the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. However, increasingly, cross-border data flows are also being discussed and addressed in *trade agreements*, via *unilateral instruments* such as adequacy decisions or contractual frameworks or in the context of *private-led or technology driven initiatives* as might be ISO standards or sandboxes.
- Challenges related to regulatory differences across countries are not new. Overcoming these differences requires engaging in a process of international dialogue to identify policy responses that address the varying concerns that regulatory differences raise and help find pathways for convergence and/or interoperability.

To date, it has proven challenging to promote international discussions that bridge the very different positions of countries on these sensitive issues. However, building up from areas where there is comfort, the G20 Digital Economy Taskforce (DETF) under the Saudi Presidency can provide leadership by supporting further dialogue on data and cross-border data flows in the following areas.

First, by promoting international cooperation through a greater sharing of experiences and good practices for data policy, in particular in the area of interoperability and transfer mechanisms and identifying commonalities between existing approaches and instruments used to enable data to flow across borders with trust. Second, by reaffirming the importance of the interface between trade and digital economy, including by noting the ongoing negotiations under the Joint Statement Initiative on electronic commerce and restating the importance of the Work Programme on electronic commerce at the WTO. Third, by exploring the use of technological solutions such as privacy enhancing technologies (PETs).

While the concerns raised by the digital transformation of an increasingly global economy are multiple, challenges related to regulatory differences across countries are not new. Overcoming differences will require engaging in a process of international dialogue to identify appropriate and proportionate policy responses that address the different concerns that such regulatory differences raise, allowing to deliver an environment of data flows with trust.

1. Motivation and aim of the work

In today's digitised and globally interconnected world, data has become the lifeblood of economic and social interactions. It has changed how business operates; altered the configuration of global value chains (GVCs), given rise to new information industries (e.g. cloud computing, 3D printing or the Internet of Things), changed how services are produced and delivered and even affected how we trade and grow food.

Digitalisation has also led to the emergence of new challenges, especially around access and sharing of data, including in the context of international data flows. For instance, while businesses are increasingly reliant on data for their day-to-day operations, supply-chain management or for tailoring products to consumers, sharp increases in the amount of personal information gathered by businesses have fuelled concerns about privacy protection and about how personal data is being used and monetised.

It is increasingly clear that the benefits of the digital transformation for economies and societies are contingent on the degree of trust placed on the activities of different players operating in the digital space and on the ability to scale economic activity and operate globally. In recognition of the importance of data and cross-border data flows, the G20 Trade Ministers and Digital Economy Ministers concluded, in June 2019, the following on their discussion on "data free flow with trust":

> *"Cross-border flow of data, information, ideas and knowledge generates higher productivity, greater innovation, and improved sustainable development. At the same time, we recognize that the free flow of data raises certain challenges. By continuing to address challenges related to privacy, data protection, intellectual property rights, and security, we can further facilitate data free flow and strengthen consumer and business trust. In order to build trust and facilitate the free flow of data, it is necessary that legal frameworks both domestic and international should be respected. Such data free flow with trust will harness the opportunities of the digital economy. We will cooperate to encourage the interoperability of different frameworks, and we affirm the role of data for development."*

This statement underscores the emerging challenge for governments which consists of ensuring that businesses and individuals can operate in an environment of trust. Where regulatory objectives such as privacy protection can be met, even when data is moved across jurisdictions, and where businesses are able to operate in a way that enables the benefits of digitalisation.

Against this backdrop, the aim of this paper is to facilitate G20 members in their ongoing dialogue on data and cross-border-data flows by mapping the different issues at stake in this debate, identifying the new issues that data and data-sharing raises, looking at the evolving regulation that affects the movement of data across borders, and discussing the approaches that are being used to enable data transfers across jurisdictions. To this end, the next section provides some context, setting the scene and highlighting the importance of data for modern economic and social activity. Section 3 then maps different characteristics of data, including their nature. Section 4 discusses the growing regulation affecting the movement of data looking at broad approaches. Section 5 identifies international approaches to help data flow with trust across borders. Section 6 concludes.

2. Context and scene setting

With the proliferation of devices, services, and sensors throughout the economy and society, the increasing adoption of artificial intelligence and the Internet of Things, the volume of (digital) data is growing at an unprecedented pace.[1] The rapid growth in the power of computing, the large decrease in storage costs over the last 20 years and cloud computing have also played a significant role in the increase in data storage and processing capacity.

With data increasingly supporting economic and social interactions, understanding what it is, how it is different and how it supports economic activity is critical for harnessing the new opportunities it enables and addressing the challenges it raises.

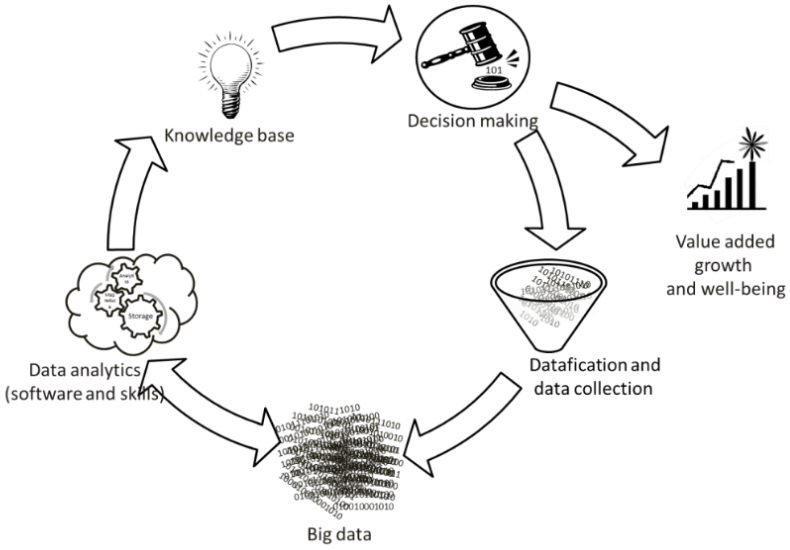
2.1. Data is an increasingly critical resource, generating benefits across society

The use, and in particular the re-use, of data across the economy underlines the importance of data as a new form of capital for 21st century knowledge economies. Data is an increasingly critical resource, generating benefits across society in ways that cannot be foreseen when it is first generated. Its value is not necessarily intrinsic, it is contextual and emerges through processing and analysis. Techniques including software, artificial intelligence (AI) and visualisation tools are needed to extract insights, convert data into information and ultimately generate knowledge to support decision-making. Analysis of data can reveal information that firms can use to modify and improve their business practices and produce new goods and services. The value of data also depends on the timeliness of its use. This implies that while data cannot be depleted, it can depreciate, in particular when it becomes less relevant for the particular purpose of its intended use.

Data value creation can best be described as a process through which data is transformed into knowledge and innovation (OECD, 2015[1]). This data-driven innovation process is not linear, it involves feed-back loops at several phases of the value creation process (Figure 1). The social and economic value of data is mainly reaped across two stages: first when data is transformed into knowledge (gaining insights) through analysis and then when it is used for decision-making (taking action), a phase which is most important for businesses.

[1] The term “data” can have multiple meanings. For the purposes of this report, the term “data” covers only electronic versions of data (digital data) and distinguishes between (raw) data and information, where the latter is understood as “the meaning resulting from the interpretation of data” (OECD, 2015).

Figure 1. The data value cycle



Source: (OECD, 2015[1])

For many of the steps in the value creation process along the data value cycle presented above, organisations will have to involve third parties around the world, because they lack the experience, technological resources and/or talent to deal with the multidisciplinary aspects of data and analytics on their own. This means that the process often involves data moving across countries.

2.2. The effective use of data boosts productivity and enables economic activity across all sectors

The effective use of data can help boost productivity and improve or foster new products, processes, organisational methods and markets. Although there is still little reliable quantification of the economic effects of data use, firm level studies suggest that firms that use data exhibit faster labour productivity growth than those that do not by approximately 5% to 10% (see (OECD, 2015[1]).

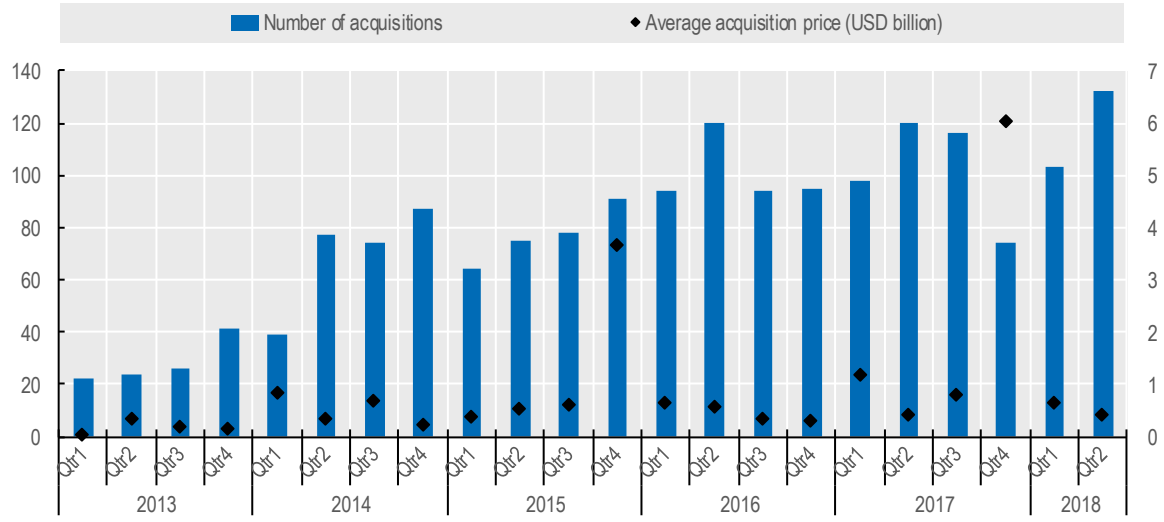
In manufacturing, the data used is typically obtained through sensors that are increasingly used to monitor and analyse the efficiency of machines, to optimise their operations and to provide after sale services, including preventative maintenance. The data are sometimes also used to work with suppliers, and are, in some cases, even commercialised in the form of new services (for example, to optimise production control) (OECD, 2017[2]). Increasingly, manufacturing activities rely on data flows that connect geographically dispersed stages of production across global value chains (OECD, 2019[5]). In the United States, for instance, (Brynjolfsson and McElheran, 2019[6]) estimate that being at the frontier of data-driven decision in manufacturing is linked with improvements in revenue-based productivity of 4-8%.

In agriculture, data captured by sensors on farm equipment are combined with weather, climate, and soil data, to provide information about production processes. This often involves transfers of different types of data, including personal or commercially sensitive information, from and to other countries. The use of all this data together with data analytics (i.e. precision agriculture) provides productivity gains by optimising the use of agriculture-related resources. These include, but are not limited to, savings on seed, fertiliser and irrigation as well as farmers' savings in time (OECD, 2017[2]). By some estimates the economic benefits from precision agriculture can be around USD 12 billion annually for the US. This represents about 7% of the total value-added of USD 177 billion contributed by farms to the GDP of the United States in 2014 (Schimmelpfennig and Ebel, 2016[7]).

The economic importance of data access and analysis is reflected also in the growing number of mergers and acquisitions (M&A) of data intensive firms often meant to assure access to business-critical data and the ability to analyse such data. Examples of such M&As 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); and Alibaba's total investment of USD 4 billion between 2016 and 2018 to acquire Lazada, a leading e-commerce platform. The annual number of acquisitions increased from more than 100 acquisitions in 2013 to more than 400 in 2017, with the average price paid exceeding USD 1 billion in some quarters (Figure 2).

Figure 2. Trends in the acquisition of big data and analytics firms, Q1 2013 - Q2 2018

Number of acquisitions (left scale), and average acquisition price in USD billion (right scale)



Source: OECD based on Crunchbase data.

2.3. The data ecosystem value chain is global

Individuals and organisations rely more than ever on data collected, stored, processed, and transferred from other entities, often located abroad. Data may be collected from individuals or devices located in one country through devices and apps developed in another country. They may then be processed in a third country and used to improve marketing to the individual in the first country and/or to other individuals around the globe. As a result, 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.[2]

Whether large or small firms, goods or services traders, businesses across all industries are therefore increasingly reliant on data transfers in support of their activities, including in global value chains ((OECD, 2017[8]) (OECD, 2018[3]) (OECD, 2019[5])). This means that data increasingly underpins international trade transactions, reducing trade costs which support growing trade in goods and enabling trade in services

[2] As stated in the OECD (1985) 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. Transborder data flows are thus not only a condition for information and knowledge exchange, but also a vital condition for the functioning of globally distributed data markets and societies.

that were previously considered non-tradable. Data has also allowed new bundles that have blurred distinctions between goods and services.

Cross-border data flows are especially important for micro, small and medium-sized enterprises (MSMEs), enabling a new breed of 'micro multinationals' which is 'born global' (MGI, 2016[11])and is constantly connected. Data flows allow MSMEs to access IT services, such as cloud computing, reducing the need for costly upfront investment in digital infrastructure. This allows them to be more nimble, quickly scaling-up IT functions in response to changes in demand. Better and faster access to critical knowledge and information also helps MSMEs overcome informational disadvantages, notably with respect to larger firms, reducing barriers to engaging in international trade and allowing them more readily to compete with larger firms.

In addition, transborder data flows can also facilitate collaboration between governments to improve their policy-making at international level. They can help strengthen collective commitment and efforts across borders to support greater public sector transparency, contribute to addressing global challenges as defined for instance by Sustainability Development Goals (SDGs) as well as reduced corruption as highlighted in the 2015 G20 Open Data Principles for Anti-corruption.

2.4. Data sharing enables increasing returns to scale and scope

Data, in theory, is an infinite resource and can be re-used by an unlimited number of users and for an unlimited number of purposes as an input to produce other goods and services (OECD, 2015[1]). This means that there are considerable increasing returns to scale and scope through data access and sharing, including across borders. 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]).

This is particularly evident in data-rich sectors such as the financial sector, health care, transportation and public administration, as well as in new production platforms in manufacturing and services. But even in traditionally less data-intensive fields, organisations are starting to leverage the large volumes of data generated from today's myriad transactions and production and communication processes. With the increasing use of AI and the IoT, the supply of, and demand for, data will thus increase. A single self-driving car, for example, can generate between one and five terabyte (TB) of data per hour according to some estimates (Grzywaczewski, 2017[12]; Nelson, 2016[13]).[3]

Overall, the evidence shows that enhancing data access and sharing can generate positive social and economic benefits for data holders, their suppliers and data users, and for the wider economy thanks to: (i) greater transparency, accountability and empowerment of users, for instance, when open data is used for (cross-subsidising) the production of public and social goods; (ii) new business opportunities, including for the creation of start-ups and in particular for data intermediaries and mobile application (app) developers; (iii) competition and co-operation within and across sectors and nations, and including the integration of value chains, (iv) crowdsourcing and user-driven innovation and (v) the increasing efficiency thanks to linkage and integration of data across multiple sources.

Recent available studies provide a rough estimate of the magnitude of the relative effects of enhanced access and sharing. They suggest that enhancing access to and sharing of data can create value for data holders (direct impact), but it can help create 10 to 20 times more value to data users (indirect impact), and 20 to 50 times more value for the wider economy (induced impact). In some cases, however, enhanced access and sharing may also reduce the producer surplus of data holders. Overall, these studies suggests

[3] As a comparison, an average person is estimated to generate up to 1.5 GB per day by 2020. (Nelson, 2016[13])

that enhanced access and sharing can help generate social and economic benefits worth between 0.1% and 1.5% of 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. Where sharing data across borders is concerned, estimates suggest that cross-border data flows contribute USD 2.8 trillion to global economic activity (MGI, 2016[12]).

2.5. Measures that affect data sharing and cross-border data flows can restrict the functioning of markets

Despite the growing evidence of the economic and social benefits from data re-use and linkage, data access and sharing remains below its potential (OECD, 2019). Risks associated with the possible revelation of confidential information (e.g. personal data and trade secrets) are often indicated as the main rationale for individuals and organisations not to share their data. Measures affecting access and sharing also remain in cases where commercial and other private interests would not oppose data sharing and re-use.

At the same time, the growing exchange of data has raised new challenges. These relate not only to ensuring the privacy and security of individuals but also to national security considerations, the need to protect intellectual property, to maintain the reach of regulatory and audit bodies, and concerns related to economic development, especially in the context of data crossing different jurisdictions. This has led countries to adapt data policies to the digital age and, in many instances, introduce measures that restrict the movement of data across-borders or which mandate that data be stored locally.

Digital infrastructures such as the Internet operate globally and although they offer new opportunities for people and countries around the world, they also raise considerable challenges for domestic and international policy in a world where borders and regulatory differences between countries remain. Given the huge potential of data in generating spill-over effects, measures affecting data sharing, including across borders, can provide significant opportunity costs not only for those controlling the data, but also for society at large. Restrictions to cross-border data flows, in particular, could severely affect the functioning of global value chains and international trade more broadly (OECD, 2015, (OECD, 2019[5])) and the globally distributed data markets.

In this evolving environment, it is increasingly clear that the benefits of the digital transformation for our economies and societies are contingent on the degree of trust placed on the activities of different players operating in the digital space and on the ability to scale economic and social activities globally. The challenge for governments is to create this environment of trust where regulatory objectives can be met, even when data is moved across jurisdictions – an environment of “data free flow with trust”.

2.6. Mapping the evolving environment will help better focus efforts towards more data sharing

Against this backdrop, a better understanding of the evolving landscape would help G20 members in their ongoing dialogue around data and data flows. There remains several misconceptions and lack of clarity on key concepts and terms which clutter the policy debate. A more detailed mapping of the issues at stake, including a discussion of the heterogeneity of data and access and control mechanisms and the emerging measures that restrict or enable data sharing can provide a first step towards identifying interoperable approaches that meet the dual goal of ensuring that data can flow across borders with trust.

3. Understanding data: what data for what purpose?

Data is sometimes treated as monolithic entity, but evidence shows that data are heterogeneous goods whose value depends on the context of their use, with different implications for individuals, businesses, and policy makers (OECD, 2015[1]; OECD, 2013[16]).[4] From a privacy perspective, *personal data*, for instance, typically requires more restrictive access regimes than, for example, (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. This chapter discusses these issues, drawing heavily on OECD (2019).

3.1. Data characteristics and control mechanisms

The following sections present and further discuss four major dimensions that are considered critical for the governance of data and data flows based. These include: (i) *personal and confidential data and the degrees of identification*, given that a higher degree of identification 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 is 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 various data stakeholders including data subjects, data holders and data users can have. Finally, (iv) *the ways in which data is accessed and controlled,* including access control mechanisms, downloads, APIs, and data sandboxes, the last two of which can be considered to enhance access to data while at the same time 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 even across borders, but only within a restricted digital and/or physical environment (data sandboxes) to trusted users in accordance with existing rules on the protection and dissemination of such information (see section below on “Data sandboxes for trusted access and re-use of sensitive and proprietary data”). 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 individuals or business customers may depend on the manner that data originated and whether the data is considered personal and/or proprietary (see section below on “The manner data originates – reflecting the contribution to data creation”). Where data is 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. Expectations 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.

3.1.1. Personal and confidential data, and the degrees of identification – reflecting the risk of harm

Most privacy regulatory frameworks as well as the (OECD[18]) *Council Recommendation concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (OECD Privacy Guidelines) include a definition of personal data. The OECD Privacy Guidelines, for instance, define “personal data” as “any information relating to an identified or identifiable individual (data subject)”. Once data is classified as personal data, access to and sharing of this data become predominantly governed by the applicable privacy regulatory framework. And this remains true irrespective of the sector of data

[4] Referring to most privacy frameworks, WEF (2014[24]), 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.”

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 AI) have made it easier to link and relate seemingly non-personal data to an identified or identifiable individual (Narayanan and Shmatikov, 2006[19]; Ohm, 2009[20]). This is not only blurring the distinction between personal and non-personal data (OECD, 2011[21]), but it is challenging 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 deserve to be distinguished and addressed differently in some cases, given the different context and the different level of risks associated with their collection, processing and use. This is reflected in some privacy regulatory frameworks such as the GDPR (see 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).[5]

To address these shortcomings, OECD (2019[4]) suggests using a more detailed differentiation of personal data based on recognised standards, such as ISO/IEC 19941 (2017[22]).[6] ISO/IEC 19441 distinguishes between five categories or states of data identification, which include (in reverse order to the degree identification)[7]:

- *Identified data*: Data that can unambiguously be associated with a specific person because personal identifiable information (PII) 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.

[5] Art. 9(1) GDPR (European Union, 2016[29]) states that "Processing 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".

[6] ISO/IEC 19941 (2017[22]), 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.

[7] These are aligned with the privacy enhancing de-identification techniques included under the ISO/IEC 20889 (2018[70]) standard.

- *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.

This differentiation is relevant for the governance of data and data flows 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 and confidentiality, which in turn can help determine the degree to which legal and technical protection may be necessary, including the level of access control required. 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 confidentiality, and thus the more openly can the data be made available, including across jurisdictions (see Section 3.3 on privacy-enhancing technologies and methods for further discussion).[8] The preferred level of data openness will therefore depend on the potential impact that data re-use will have on privacy (or more generally on confidentiality).

3.1.2. 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 certain types of public sector data should be made available through open data, free of charge and free of any restrictions from intellectual property right (IPR) – where there are no conflicting interests, such as national security, law enforcement, regulatory personnel, 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), or data.go.jp (Japan) (see section below on “Open data”).

However, the private-public data distinction as often defined in policy circles[9] raises a number of issues, which are rarely fully acknowledged (OECD, 2019[4]). The most important being that 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 the private sector entities (e.g. public-private partnerships). For instance, data generated, created, collected, processed, preserved, maintained, disseminated by the private sector, that are however funded by or for the public sector would classify as public sector *and* private sector data.[10] As a result, some of the existing presumptions about public sector data (e.g. that public sector data generally, when appropriate, should be made available through open data, free of charge and free of any IPR restrictions) have to be questioned.[11]

---

[8] 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.

[9] See for instance the OECD (2008[17]) Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (OECD PSI Recommendation), according to which public sector (government) data is defined as meaning data generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the government or public institutions. In analogy, private sector data can be defined as data that is generated, created, collected, processed, preserved, maintained, disseminated, and 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[18]).

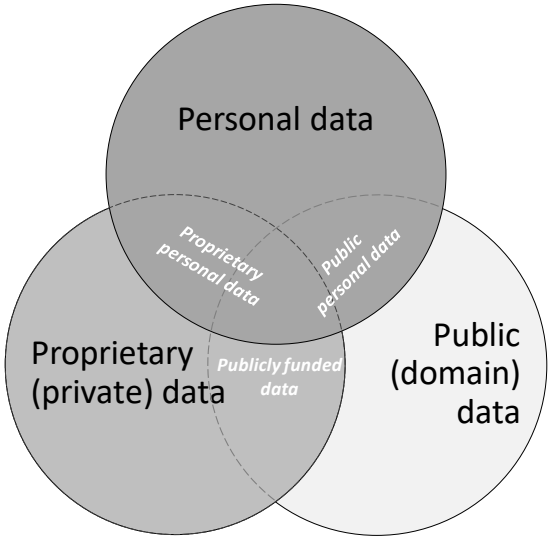
[10] See as another example data funded by the private sector, but collected, processed, preserved, and maintained by the public sector.

[11] Both the OECD (2008[56]) *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* (OECD Recommendation on Research Data), call for

For data governance frameworks to be applicable across society, it seems crucial to rather distinguish between the following three domains of data (Figure 3), as discussed in OECD, 2019:

- the *personal domain*, which covers all data "relating to an identified or identifiable individual" (personal data) for which data subjects have an interest for privacy,
- the *private domain,* which covers all *proprietary data* that are typically protected by IPR (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 certain types of such data are free to access and re-use.

Figure 3. The personal, private and public domain of data



Source: OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies

These three domains not only overlap as illustrated in Figure 3, 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 may not be subject to any specific framework and governed through contractual frameworks, or in some specific instances covered by IPRs. These overlaps may partly explain the potential conflicting views and interests of some stakeholder groups, as reflected for instance in issues related to "data ownership".

Furthermore, these overlaps can explain why data governance is often perceived as complex from a legal and regulatory perspective, in particular when cross-border data flows are concerned. Depending on the jurisdiction, some domains may be prioritised differently over others. This is for example reflected in current privacy and data portability rights, which vary significantly across countries, reflecting various approaches to address the conflicting interests of individuals and organisations over their "proprietary personal data" (Figure 3). Even in the case of data portability which aims to empower individuals and give them more control rights over their personal data, it remains unclear what type of data falls within the scope of data

open access to *publicly funded data*, irrespective of whether the data is controlled by an entity in the public or private sector, acknowledges however the need to protect the commercial interests of the private sector.

portability across the various initiatives. This reflects the (implicit) priorities of personal domain vs. the proprietary domain.

3.1.3. The manner data originates – reflecting the contribution to data creation

Multiple stakeholders are often involved in the contribution, collection and control of data, including the data subject in the case of personal data. The data categories discussed above – in particular the distinction between the personal domain and the proprietary domain – however do not help differentiate how different stakeholders contribute to data co-creation. Data categories that differentiate according to the way data is collected or created (Schneier, 2009[23]; WEF, 2014[24]; Abrams, 2014[25]; OECD, 2014[26]) can provide further clarity in this respect.

Based on Abrams (2014[25]) and the Productivity Commission (2017[27]), OECD (2019[4]) suggests to distinguish data based on how different stakeholders contribute to data co-creation as follows:

- *Volunteered (or surrendered or contributed or provided) data* is 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 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 case of observed data is passive and it is the data controller that 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[26]). 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[19]).
- *Acquired (purchased or licenced) data* are obtained from third parties based on commercial (licencing) contracts (e.g. when data is acquired from data brokers) or other non-commercial means (e.g. when is acquired via open government initiatives). As a result, contractual and other legal obligations may affect the re-use and sharing of the data.

This differentiation is relevant for the governance of data and data flows for several reasons: (i) it helps determine the level of awareness that data subjects can have about the privacy impact of the data collection and process, which is critical when assessing the privacy risks associated to data collection and the level of control data subjects can be expected to have[12] (ii) It also reflects the contribution of various stakeholders to data creation and therefore their rights and interests in accessing and using the data. (iii) Last, but not least, this differentiation can help identify the geographic location and jurisdiction based on data generation and collection, and it can therefore help determine the applicable legal and regulatory frameworks.

[12] Abrams (2014[25]), for instance, notes that "legacy privacy governance regimes are based on a presumption that data is primarily being collected from the individual with some level of their awareness".

3.1.4. Data access control mechanisms – protecting the interests of data stakeholders

There are a wide range of different mechanisms for accessing and sharing data within and across organisational and national borders. OECD (2019[4]) highlights the most commonly used, namely data access via (i) (ad-hoc) downloads, via (ii) application programming interfaces (APIs), and (iii) data sandboxes, which are increasingly recognised as means to access sensitive and proprietary data, while assuring the privacy rights and IPR of right holders.[13]

- (Ad-hoc) downloads: In the case of data access via downloads, the data is stored, ideally in a commonly-used format and made available online (e.g. via a web site). 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. 14 Furthermore, data access via (ad-hoc) downloads can increase digital security and privacy risks since data once downloaded is 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 in respect to the protection of the privacy of data subjects and the IPRs of right holders.
- Application programming interfaces (APIs): 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, their technologies, and services, particularly through the use of cloud computing. 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.
- Data sandboxes for trusted access and re-use of sensitive and proprietary data: The term “data sandbox” is used in OECD (2019[4]) 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 is located). Data sandboxes would typically require that the analytical code is executed at the same physical location as where the data is 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 including across borders. 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 (CDRs) including de-

[13] 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[15]).

[14] 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” OECD (2019[4]). 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.” OECD (2019[4]).

identified low-resolution location data (on nearest tower location). To assure the privacy of their users, mobile operators hosted separate dedicated servers behind their firewalls, on top of which Flowminder.org researchers would conduct their analyses. The data always resided within the operators' servers; only non-sensitive aggregated estimates were exported outside the servers. This configuration allows sensitive data to remain under the operators' control, minimising privacy, security and commercial concerns.

3.2. Approaches to access, sharing and degrees of openness

Different approaches can be employed to enhance data access, sharing and re-use along the data openness continuum, which can range from internal access and re-use (only by the data holder), to restricted (unilateral and multilateral) external access and sharing, and open data as the most extreme form of data openness (Figure 4).

Figure 4. The degrees of data openness and access

Source: (OECD, 2015[1])

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) engagements in *data partnerships*. Many of these approaches are based on voluntary and mutually agreed terms between organisations, while others are mandatory such as the right to data portability under the EU (European Union, 2016[29]) General Data Protection Regulation (GDPR) (Art. 20) or Australia's recently proposed Consumer Data Right (CDR) (see (OECD, 2019[4]) for more examples].

3.2.1. Contractual agreements and data markets

Increasingly, businesses are recognising the opportunities of commercialising their proprietary data (OECD, 2015[1]). While some organisations offer their data for free (via open access) – especially NGOs and governments as highlighted below, many businesses have already engaged bilateral arrangements to sell or licence their data. For example, the French mobile ISP Orange is acting as a data provider by using its *Floating Mobile Data* (FMD) 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.

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 survey by Forester Research of almost 1 300 data and analytics businesses across the globe, only a 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 healthcare are in the bottom of the list. (Belissent, 2017[32])

With 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, the commercialisation of data could become more mainstream, in particular as even less data savvy firms may find it easier to commercialise their data.

3.2.2. Open data

Open data is the most prominent approach used to enhance access to data and the most extreme form of data openness (OECD, 2015). In the public sector, open government data has been promoted for many years by initiatives such as data.gov (United States), data.gov.uk (United Kingdom), data.gov.fr (France), or data.go.jp (Japan) (Ubaldi, 2013[33]).

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 e.g. selling the data or indirectly by e.g. 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 (social) value of data, as it creates new business opportunities and economic and social benefits. Assessing the resulting economic and social benefits of moving towards open data remains however 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".

3.2.3. 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 of businesses (in particular SMEs) over their business data (Productivity Commission, 2017[27]). 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 'My Data' initiatives of the United States initiated in 2010, such as the 'Green Button' (United States Department of Energy, n.d.[37]), to the 'Midata' data portability initiative of the United Kingdom in 2011 (Department for Business Innovation & Skills, 2011[38]), the 'Right to Data Portability' (Art. 20) of the EU (European Union, 2016[29]) GDPR, and most recently Australia's recently proposed Consumer Data Right (CDR).

Data portability initiatives may however 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.

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, and specifically by reducing switching costs. For example, a study on current limitations to move mobile applications (apps) across platforms shows that switching cost can be a barrier for moving from

one platform to another. Enabling *app portability*, for 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[39]; iClarified, 2012[40])

3.2.4. Data partnerships and data for societal objectives

In cases where data is considered too confidential to be shared openly with the public (as open data) or where there are legitimate (commercial and non-commercial) interests opposing such sharing, restricted data sharing arrangements can be more appropriate. This is for instance 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 of data. 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, under voluntary and mutually agreed terms.

It is, for example, common to find restricted data sharing agreements in areas such as digital security, science and research, 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 types of are highlighted in the following sections 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 is shared to support societal objectives.

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[41]). Examples include:

- The pooling of data between Take Nectar, a UK-based program 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[42]).
- The joint venture between DuPont Pioneer and John Deere illustrates. 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[43]).
- And the collaboration of Telefónica with organisations such as Facebook, Microsoft, and 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 some third-party platforms like Facebook Messenger, and in the future through Google Assistant and Microsoft Cortana.

Similar partnerships also exist in the form of public and private partnerships (data PPPs):

- For Transport for London (TfL), a local government body responsible for the transport system in Greater London (United Kingdom), 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 could enhance the efficiency of its planning and operation.

Data partnerships (including data PPPs) however raise several challenges as highlighted in OECD (2019[4]). For instance, ensuring a fair data sharing agreement between the partners can sometimes be challenging, in particular where partners have different 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 sharing arrangements can also be found where private sector data is 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[44]) 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[44]). Examples range from trends in the consumption of goods and services to flu epidemics and employment/unemployment trends, and the monitoring of information systems and networks to identify malware and cyberattack patterns (Choi and Varian, 2009[45]; Harris, 2011[46]; Carrière-Swallow and Labbé, 2013[47]). 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[48]).[15] And some of these arrangements have been instrumental in the fight against pandemics such as COVID-19 (Box 1).

Box 1. Data and the fight against the COVID-19 pandemic

The importance of data sharing linked to social good initiatives is being highlighted in the context of the current COVID-19 pandemic. The spread of the novel coronavirus (COVID-19) is seriously affecting human lives and global economies, and governments are adopting a wide range of measures to track and contain its spread. Timely, secure and reliable data access and sharing – within and outside borders, and between governments, organisations, civil society and individuals – are critical to understanding the epidemic, improving the effectiveness of government policies, and fostering global co-operation in the race to develop and distribute therapies and vaccines.

In particular, lessons from previous outbreaks have underscored the importance of data concerning the spread of virus infections, such as the location and number of new confirmed cases, rates of recoveries and deaths, and the source of new cases (international arrivals or community transmission). The ability to model the pathogenesis, transmission, effective control strategies, and spread of a disease can provide crucial information to those needing to make decisions about the distribution of limited resources.

Knowing how a virus mutates as it moves through a population is also vital to understanding possible changes in disease severity or transmissibility, amenity to diagnosis, and responsiveness to vaccine. This is an issue of global interest and needs to involve scientists from many parts of the world. International data sharing and enlisting technology companies that have the ability to provide data acquisition and processing are important components of a comprehensive response system.

In the current global health emergency scientific discovery has progressed much more rapidly than before. The full genome of COVID-19 was published barely a month after the first patient was admitted into Wuhan

[15] 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.

hospital, as an open-access publication. The release of full viral genome sequences through public access platforms and the polymerase chain reaction assay protocols that were developed as a result made it possible to accurately diagnose infections early in the current emergency.

While global sharing and collaboration of research data has reached unprecedented levels and clinical, and epidemiological and laboratory data about COVID-19 is today widely available, similar efforts may also needed for other types of data.

Data is crucial to monitor the supply and demand of critical services such as food, fuel, healthcare and life-saving medicines. It is needed in front line responses to improve the capacity of health care systems to address the crisis and the effectiveness of containment and mitigation policies that restrict the movement of individuals. Data such as the daily number of hospitalizations, intensive care admissions, ventilator use, and deaths, can be used in forecasting expected epidemic progression and assist with clinical care decisions.

Accurate information on population movements is also valuable for monitoring the progression of an outbreak and predicting its future spread, facilitating the prioritization of interventions and designing effective containment strategies. Vital questions include how the affected regions are connected by population flows, which areas are major mobility hubs, what types of movement typologies exist in the region, and how all of these factors are changing as people react to the outbreak and movement restrictions are implemented. Just a decade ago, obtaining detailed and comprehensive data to answer such questions over large geographical regions would have been impossible. Today, for example, geolocation data produced by telecommunication service providers on mobile telephone calls or other telecommunications transactions, can provide valuable insights into population movements. As network operators serve substantial portions of the population across entire nations, the movements of millions of people at fine spatial and temporal scales can be measured in near real-time. The resulting information and trends can be invaluable for governments seeking to track the COVID-19 outbreak, warn vulnerable communities, and understand the impact of policies such as social distancing and confinement.

However, while geolocation data, and particularly when obtained from mobile call data records, may hold great potential in the early response to crises, this data is generally very difficult to access due to commercial and privacy concerns. Extending geolocation data to the management of natural disasters or outbreaks entails the problem of the nature of the information — often private and/or sensitive — associated with them. The use and the subsequent mass collection and analysis of personal data raise important data governance and privacy concerns. The pandemic has also exposed key weaknesses in data integration across public and private sources; clinical care and public health; local, state, national and international levels.

Recent work by the OECD suggests that few countries have existing policy initiatives to facilitate data sharing within the private sector, and even fewer have data governance frameworks in place to support such extraordinary data collection and sharing measures in ways that are fast, secure, trustworthy, scalable and in compliance with the relevant privacy and data protection regulations (OECD, 2019[16]).

As a result, many countries have recently sought advice from Privacy Enforcement Authorities (PEAs), private sector law firms, civil society, academics and other actors to ensure that their actions are necessary and proportionate, and that they have a full understanding of their potential implications. Many governments have passed or are about to pass laws specifying how data collection will be restricted to a certain population, for what time, and for what purpose. PEAs across many G20 countries have generally endorsed a pragmatic and contextual approach, and exercised enforcement discretion recalling that respect for fundamental data protection and privacy principles do not stand in the way of necessary and proportionate front-line responses to COVID-19.

The European Data Protection Board and the Council of Europe have released similar statements explaining that the General Data Protection Regulation and Convention 108 do not hinder measures taken

in the fight against the pandemic but do require that emergency restrictions on freedoms be proportionate and limited to the emergency period (Council of Europe, 2020[34]; European Data Protection Board, 2020[35]).

The use of privacy enhancing solutions may provide added privacy and data protection (OECD, 2015[1]). This can include homomorphic encryption (i.e. the encryption that allows processing of encrypted data without revealing its embedded information) as well as the use of data sandboxes, through which access to highly sensitive (personal) data is only granted within a restricted digital and/or physical environment to trusted users (see Section 3.3 on Privacy-enhancing technologies and methods).

3.3. Privacy-enhancing technologies and methods

Privacy-enhancing technologies (PETs), such as anonymisation and cryptography technologies and techniques (described further below), are increasingly viewed as promising approaches designed to prevent and mitigate the risk of privacy and confidentiality breaches and enable organisations to better manage data responsibly. PETs may make it possible to balance the respective interests of data users and data subjects by enabling data access and use, while data subjects' privacy remains protected(Acquisti, 2010[1])

PETs typically can help reduce the risk of privacy and confidentiality breaches by *minimising information disclosure* in a number of ways. Some PETs seek to curb "default" data disclosures; that is, they are designed to prevent any disclosure of data, unless strictly necessary to provide the envisaged functionality. The use of PETs in this case is primarily aimed at minimising the risk of inadvertent disclosures (e.g. due to how a specific system operates). A commonly used application of this type of PETs is the Internet security protocol Transport Layer Security (TLS)[16], which is used by e.g. Internet browsers, mobile applications (apps) and web servers to exchange sensitive information such as passwords and credit card numbers. TLS therefore ranks as the most frequently used application of cryptography (OECD, 2017[6]).

Other PETs do not focus on data minimisation, but rather on *obfuscation*. The goal of these PETs is to introduce "noise" into the system, with the aim of disrupting the integrity and reliability of data collection. This type of PET is often highlighted as a tool for individuals to counter or disrupt the continuous surveillance of their online activities, although not without controversy as they can also be used for unlawful and malicious intents. A commonly cited application of this type of PETs is Tor (originally the acronym for The Onion Router), an anonymity network that allows anyone to use the Internet without easily revealing their location or identity.[17]

While a common objective of all PETs is to minimise the risk of privacy and confidentiality breaches, they cannot completely eliminate these risks. Even when using de-identified data, if a sufficient number of an individual's data sources are cross-linked, patterns can emerge that are traceable back to that same individual. De-Montjoye et al. (2018[2]) have shown, for example, that four spatio-temporal points of pseudonymised data may be enough to uniquely identify 95% of people in a mobile phone database of 1.5 million people and to identify 90% of people in a credit card database of one million people (see Section

[16] TLS, as was the case with its predecessor the Secure Sockets Layer (SSL) protocol, relies on a certificate authority, such as those provided by companies like Symantec and GoDaddy, that issues a digital certificate containing a public key and information about its owner, and confirms that a given public key belongs to a specific site. In doing so, certificate authorities act as trusted third parties.

[17] Tor (originally the acronym for The Onion Router) is free software that protects Internet users' privacy, confidentiality of communications and other freedoms (i.e. freedom of expression) by enabling online anonymity. The project was initially sponsored by the US Navy Research Lab, then by the Electronic Frontier Foundation, and now by the Tor Project, which is a US-based research and education not-for profit organisation, with different sources of funds published on the website.

3.3.2). Questions are therefore emerging on the efficacy of PETs to withstand progress in re-identification methods in particular in the absence of complementary technical, organisational and legal measures.

Efficacy also depends on the context in which the data are to be used, as well as on the resources and motivation of those who might try to undermine PETs. The use of PETs therefore requires risk assessment, which should also consider the likely consequences to the data subject in the event that re-identification occurred. Approaches that combine the use of PETs with administrative and legal measures (e.g. enforceable commitments not to re-identify) are to be prioritised to minimise risks of privacy and confidentiality breaches (OECD, 2015[3]; OECD, 2019[4]).

The following sections briefly present key PETs as well organisational methods that can be used to complement their use. More work will be needed to evaluate the reliability of these tools and their applicability to cross-border data flows. A number of possible barriers to the implementation or adoption of PETs include lack of awareness about the existence of these tools, poor usability, and a lack of incentives for organizations to offer or implement these tools. Additional work is also needed to assess the relative strengths and weaknesses of PETs, develop new PETs or improve the effectiveness of existing ones, and better understand the barriers to their deployment and adoption, including their application to cross-border data flows.

3.3.1. Cryptography

Cryptography is a practice that "embodies principles, means, and methods for the transformation of data in order to hide its information content, establish its authenticity, prevent its undetected modification, prevent its repudiation, and/or prevent its unauthorised use" (OECD, 1997[5]). It is the most fundamental technological basis for PETs. Cryptography is increasingly used by businesses and for consumer goods and services to protect the confidentiality of data, such as financial or personal data, whether that data is in storage or in transit (OECD, 2017[6]). It can also be used to verify the integrity of data by revealing whether data have been altered to identify the person or device that sent it. There has been an acceleration of the adoption of cryptography by businesses (including for consumer goods and services) since 2014, with companies such as Apple and Google starting to accelerate their default use of cryptography.[18]

According to a Thales e-Security (2016[7]) 11-year study on encryption application trends,[19] the rate of adoption of cryptography has increased significantly since 2014. In 2015, 41% of surveyed businesses indicated having extensively deployed cryptography, compared to 34% in 2014 and 16% in 2005. A recent and comparable study by the Ponemon Institute (2020[8])[20] shows that the share of companies with an encryption strategy applied consistently across the entire enterprise has increased from 15% in 2006 to 48% in 2020. In both studies, businesses surveyed indicate privacy protection and compliance with regulation,[21] digital security threats targeting in particular intellectual property, as well as employee and customer data as the main reason for this rapid increase in adoption. In particular, firms in heavily regulated

[18] Since then the latest mobile operating systems of Apple and Google provide encryption for nearly all data at rest (in addition to data in transit) by default.

[19] The study covered over 5 000 respondents across 11 countries including (in the decreasing order of the country with the largest number of respondents) the United States, United Kingdom, Germany, France, Australia, Japan, Brazil, the Russian Federation, Mexico, India, Saudi Arabia and United Arab Emirates.

[20] The study which was conducted between December 2019 and January 2020 covered more than 6 000 respondents across 17 countries including Australia, Brazil, France, Germany, Hong Kong, India, Japan, Mexico, the Middle East (which is a combination of respondents located in Saudi Arabia and the United Arab Emirates), Netherlands, the Russian Federation, Southeast Asia, Korea, Sweden, Chinese Taipei, the United Kingdom, and the United States.

[21] As an example, the EU General Data Protection Regulation considers both pseudonymisation and encryption as appropriate measures to be used by data controllers and processors to ensure the security of personal data.

industries dealing extensively with big data rank high among the list of extensive cryptography users. Most notably, these include financial services, healthcare and pharmaceuticals, and digital technology firms. The Ponemon Institute (2020[8]) study also noted a significant increase in manufacturing, hospitality and consumer products in recent years.

Increasingly, the use of homomorphic encryption is being considered as a means to run computations on encrypted data whilst protecting privacy and confidentiality. Thanks to homomorphic encryption, a user can encrypt data, send it to the cloud for processing, and have the results of the computation sent back to him or her or to third parties, all the while maintaining the privacy of the individual and confidentiality of the data (The Royal Society, 2019[9]).

3.3.2. De-identification: from anonymisation to pseudonymisation and aggregation

De-identification covers a range of practices ranging from anonymisation to pseudonymisation and aggregation (see Section 3.1.1). These practices share a common aim of preventing the extraction of identifying attributes (i.e. re-identification), or at least significantly increasing the costs of re-identification. Anonymisation is a process in which an individual's identifying information is excluded or masked so that the individual's identity cannot be, or becomes too costly to be, reconstructed (Pfitzmann and Hansen, 2010[10]; Mivule, 2013[11]). Some research, such as the study by De-Montjoye et al. (2018[2]) highlighted above, suggests that when linked with other data, most anonymised data can be de-anonymised – that is, the identifying information can be reconstructed.[22]

For many applications, however, some kind of identifier is needed because complete anonymity would prevent any useful two-way communication and transaction. Pseudonymisation offers a solution whereby the most identifying attributes (i.e. identifiers) within a data record are replaced by unique artificial identifiers (i.e. pseudonyms). A prominent example is the use of blockchain or distributed ledger technologies (DLTs), including for crypto-currencies such as bitcoin. On public blockchains, user identities are typically anonymous "but their accounts are not, as all of their transactions are visible to all other users" (OECD, 2018[12]). This is in contrast to permissioned blockchains, which typically require a user's identity to be verified before they are able to access or use the blockchain.

3.3.3. Unlinkability and differential privacy

Unlinkability results from processes to ensure that data processors cannot distinguish whether items of interest are related or not. According to (Pfitzmann and Hansen, 2010[10]), unlinkability "ensures that a user may make multiple uses of resources or services without others being able to link these uses together". De-identification is a means to enable unlinkability but cannot guarantee it. Therefore, other means including noise addition, functional separation and distribution (decentralisation) and administrative safeguards are needed to ensure unlinkability.

The addition of "noise" to a data set allows analysis based on the complete data set to remain significant while masking sensitive data attributes. Finding the right balance that protects privacy while minimising the costs to data utility remains a challenge (Mivule, 2013[11]). Noise addition techniques are considered promising means to help protect privacy and confidentiality in databases, while keeping all data sets statistically close to the original data sets. Work on "differential privacy" is one example (Dwork and Roth,

[22] See also (Narayanan and Shmatikov, 2006[17]), who have used the "anonymous" data set released as part of the first Netflix prize to demonstrate how the authors could correlate Netflix's list of movie rentals with reviews posted on the Internet Movie Database (IMDb). This let them identify individual renters, and gave the authors access to their complete rental histories.

2014[13]), which addresses privacy in disclosure rather than in computation, by adding noise to the data so that when a statistic is released, information about an individual is not revealed with it.

Administrative safeguards are particularly important to ensure that data used for a particular purpose (such as scientific research) is not used inappropriately or for a different purpose (such as making business decisions with respect to a particular individual) (EU WP29 [former Article 29 Data Protection Working Party of the European Union], 2014[14]). Functional separation can also support uses related to gaining appropriate insights and knowledge. Policy makers could consider stimulating further research and development in this area, and promote adoption via private-public partnerships, certification schemes and similar initiatives.

3.3.4. Data sandboxes for trusted data access and processing

The term "data sandbox" describes any isolated environment through which data is 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 is located). Data sandboxes would typically require the analytical code to be executed at the same physical location where the data is stored.

Data sandboxes offer a potentially strong level of control and protection of data, as they combine technological means such as the use of cryptography with organisational safeguards such as functional separation. Data sandboxes are, therefore, promising for providing access to very sensitive, personal and proprietary data even across borders (see Box 2 for selected examples).

Box 2. Selected examples of data sandboxes

The Virtual Research Data Center (VRDC) of the Centers for Medicare and Medicaid Services (CMS)

The VRDC is a virtual research environment that provides timely access to Medicare and Medicaid program data (such as beneficiary level protected health information) (ResDAC, n.d.[15]). Researchers working in the CMS VRDC have direct access to approved data files and can conduct their analyses within the CMS secure environment. They can download aggregated reports and results to their own personal workstation. Researchers can also 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

Flowminder.org is an initiative combining new types of data to support, in particular, people in low- and middle-income countries, including in the context of the recent COVID-19 crisis (Flowminder, 2020[16]; OECD, 2019[4]). It relies on secured access to personal data of mobile operators' Call Detail Records (CDRs) including de-identified low-resolution location data (on nearest tower location). 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, thereby minimising privacy, security and commercial concerns.

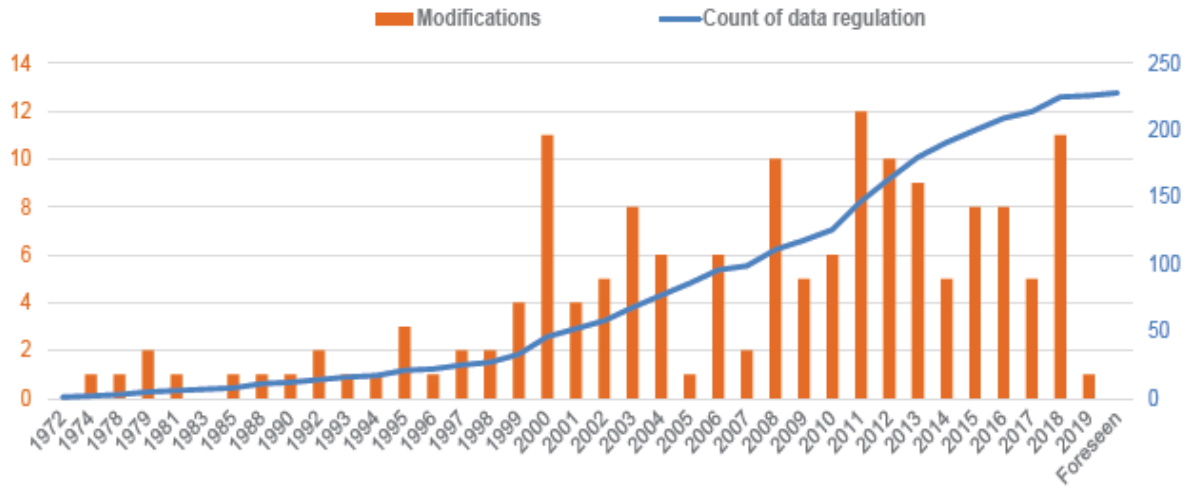
4. Domestic approaches to cross-border data flow policies

Although data-related policies have recently received growing attention from policy makers, academics and the international press (see The Economist, 2017 and The Financial Times, 2018), they have been around for some time. Indeed, the issue of cross border flows of personal data has been acknowledged

since at least the 1970s, when the OECD developed Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (Privacy Guidelines).

In light of emerging regulatory challenges and complexities outlined above, governments have been updating their data-related policies, adapting them to the digital age. This has resulted in a growing number of countries placing conditions on the transfer of data across borders or requiring that data be stored locally (Figure 5).

Figure 5. A growing number of data policies



Note: This figure includes different types of regulation relating to data transfers and local storage requirements.
Source: (OECD, 2019[5])

4.1. Why are cross-border data flow policies emerging?

The previous section reviewed the different types of control mechanisms and data characteristics that can impact the regulatory environment. This environment has additional layers of complexity when moving into cross border data flows. The motivations behind emerging data policy in a globalised economy are manifold but can be grouped into five non-exhaustive categories.

- Much of the debate about data flows revolves around the movement of personally identifiable information, which raises concerns about ***privacy protection***. For some, the challenge is to ensure that when data is transferred outside a specific jurisdiction, this data continues to receive the same protection that it received in the domestic jurisdiction. However, approaches to privacy and personal data protection vary significantly across cultures, which is why regulation also differs.
- Some measures conditioning data flows are aimed at securing access to information for ***regulatory control or audit*** purposes. In this sense, requirements for data to be stored locally can be seen as the online equivalent of a longstanding practice in the offline world of ensuring that information is readily accessible to regulators. Such measures can be sector-specific, reflecting particular regulatory requirements and targeting specific data such as business accounts, telecoms or banking data.
- Measures related to ***national security*** often mandate that data be stored and processed locally for the purpose of protecting information deemed to be sensitive, or securing the ability of national security services to access and review data. The latter in particular can be very broad in nature, providing wide scope of access to any form of data.

- Governments also promote local storage and processing with a view to ensuring ***data security***. The rationale for implementing countries is that data security can best be guaranteed when storage and processing is domestic.
- Finally, other reasons for conditioning the flow of data or mandating that it be stored locally can be motivated by the desire to use a pool of data to encourage or help develop domestic capacity in digitally intensive sectors, a kind of ***digital industrial policy***, including in the context of economic development. This can reflect a view that data is a resource that needs to be made available first and foremost to national producers or suppliers. These approaches can be sector specific or apply to a range of data.

In discussing data policy, it is important to bear in mind the underlying goals of the government as well as data characteristics. Also important are how effective the measures are in achieving their stated aims, the associated costs and trade-offs of such measures, and whether there are alternatives that would enable a better balance among different aims to maximize overall benefits for the population. From a trade policy perspective, it is important that the same policy objective can be fulfilled in a way that has a less restrictive effect on trade.

4.2. How are countries regulating cross border data flows and data storage?

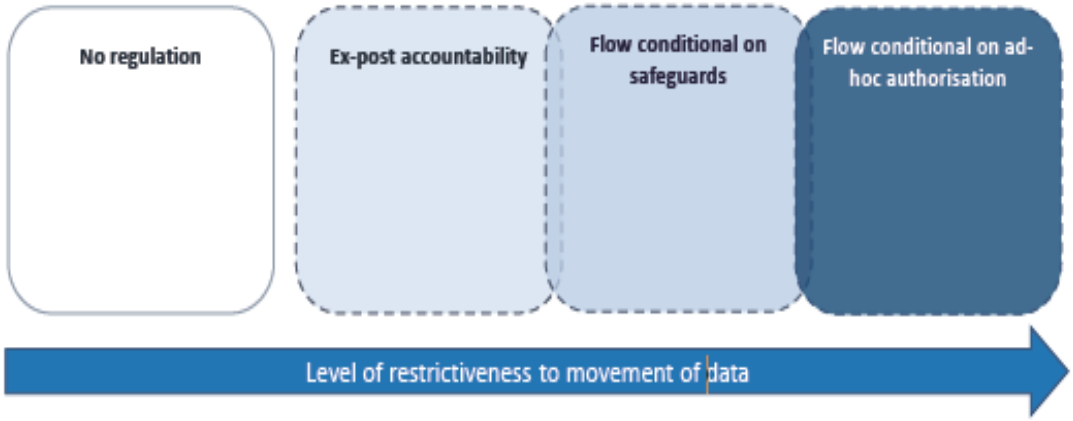
Two broad types of data policies have emerged. Those that condition the movement of data across borders and those that mandate that data is stored locally. Each addresses different and sometimes overlapping policy objectives. Moreover, the manner in which countries approach her data-related policies naturally reflects the underlying preferences, including in relation to trade-offs, of their citizens.

4.2.1. Cross border data flow policies

Although approaches to cross-border data transfers differ across countries, they can be broadly grouped into four categories, albeit with blurred boundaries (Figure 6). These are not mutually exclusive; different approaches can apply to different types of data even within the same jurisdiction (health data, for instance, might be subject to more stringent approaches than data related to product maintenance).

1. The first approach is ***no regulation*** of cross-border data flows, usually because there is no data protection legislation at all.
2. The second type of approach, ***ex-post accountability***, does not prohibit the cross-border transfer of data nor does it require any specific conditions to be fulfilled, but provides for ex-post accountability for the data exporter if data sent abroad is misused (e.g. firms send data but if something goes wrong they are legally accountable).
3. A third approach, ***flows conditional on safeguards***, includes approaches relying on the determination of adequacy or equivalence as ex-ante conditions for data transfer. These rulings can be made by a public authority or by private companies and can include requirements about how data is to be treated. Where an adequacy determination has not yet been made, firms can move data under options such as binding corporate rules, contractual clauses and consent (see Section 4.3.4.).
4. The last broad type of approach, ***flow conditional on ad-hoc authorisation***, relates to systems that only allow data to be transferred on a case-by-case basis subject to review and approval by relevant authorities. This approach relates to personal data for privacy reasons but also to the more sweeping category of “important data”, including in the context of national security.

Figure 6. Approaches to data flow regulation



Source: (OECD, 2019[5])

4.2.2. Local storage requirements

Local storage requirements constitute another type of emerging data-related policy. As their name indicates, measures falling under this category require that certain types of data be stored in local servers, and often also include local processing requirements. Although distinct from cross-border data flow restrictions, a complete prohibition on the transfer of data amounts to a de-facto requirement for local storage and processing. By contrast, a local storage requirement does not always correspond to a complete prohibition of cross-border transfer. Local storage requirements could still affect cross border data flow to the extent that companies switch from a foreign supplier to a domestic supplier to store and process data that is collected in a certain country.

As with regulations on cross-border data transfers, local storage requirements can be grouped into four categories, also with blurred boundaries (Figure 7). Different local storage and processing rules can also apply to different types of data even within a country. They can be aimed at personal data, or can be sectoral, typically targeting regulated sectors such as health, telecoms, banking or payment processing, insurance, or satellite mapping.[23]

1. A default position is where there are ***no requirements*** to store data locally. This is a relatively common category given that the number of local storage requirements remains small and targeted to specific sectors.

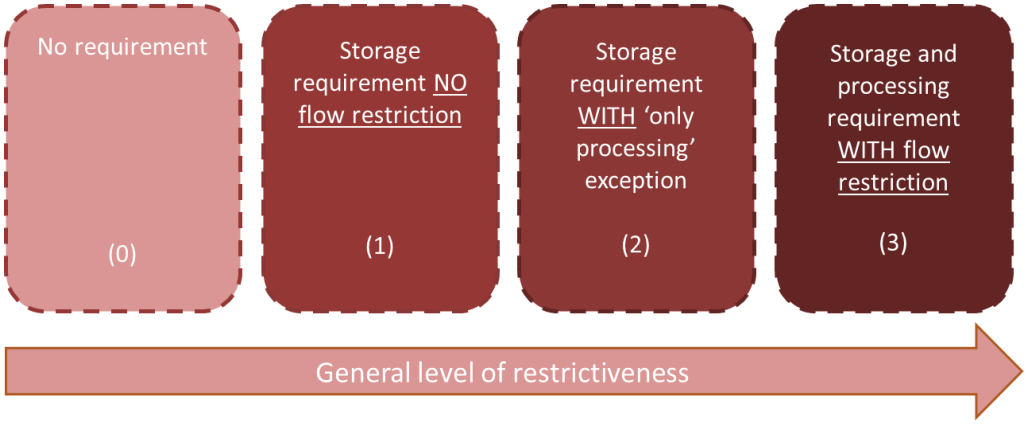
2. Next are approaches that require that ***a copy of the targeted data is stored in domestic computing facilities***. This type of approach has no restrictions in terms of transferring or processing copies of the data abroad and its objective is, more often than not, to ensure that regulators do not encounter issues related to jurisdictional reach. Approaches falling under this category often target telecommunication metadata and business fiscal data, as a continuation of traditional data retention policies. Newer approaches to data retention now establish that data be retained and made accessible to local authorities without prescribing the country where the data has to be stored. Data retention is also generally limited to a specified time period.

[23] For example, Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union prohibits localisation of non-personal data within the EU and contains provisions to help governments access data stored in other member states.

3. Another type of approach relates to those where there are ***no flow restrictions but foreign storage is not allowed***, implying that processing can occur abroad, but that post-processing, data must be returned to the home country for storage.

4. Finally, there is a category of approaches that require that ***data be stored locally with conditions attached to transferring and/or processing those data abroad***. These last two requirements can be related to a desire to encourage the development of domestic data storage and other data services industries and thus can be related to industrial policy objectives

Figure 7. Approaches to local storage requirements



Source: (OECD, 2019[5])

5. Approaches to cross-border data transfers

The multiplicity of applicable regimes may create uncertainty for governments, businesses and individuals with regard to which rules govern a given situation. While there are legitimate reasons for diversity in regulations, it is important to alleviate possible tensions and ensure that data can flow across borders with trust. In this context, governments and other stakeholders are increasingly using a range of approaches. These can help businesses find grounds for data transfers while ensuring that, upon crossing a border, data is granted the desired degree of protection. While there are many such approaches, broadly these can be categorised into 4 types: *plurilateral arrangements; trade agreements; unilateral instruments; and private sector and other initiatives*.

Each type of instrument tackles the issue of data transfers from a different perspective and the approaches are not exclusive to each other: countries can use different approaches with respect to different partners, types of data and in different situations. The review provided below is also not an exhaustive list of regulatory options, rather it is a mapping of existing approaches by broad type.

The scope of data that each approach covers also varies: for instance, rules on cross border data flow in trade agreements often cover all types of data. Meanwhile, existing plurilateral arrangements on cross border data transfers as well as some of the unilateral instruments focus mainly on issues around the protection of personal data, where there has been most activity in the context of emerging regulation.

5.1. Plurilateral Arrangements

Many plurilateral arrangements are developed in the context of personal data. These aim to generate consensus around privacy principles and foster cross-border data flows between participating economies.

They have often emerged under the auspices of regional organisations, but may be open to participation for adhesion by other countries as well (see Table 1).

There are many different approaches within this category, each with different levels of enforceability. On one side, there are non-binding plurilateral arrangements that rely on “soft law” to encourage parties to adopt data protection principles and promote interoperability between privacy protection regimes in order for data to be transferred between them seamlessly. An example of this is the 1980 OECD Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (“OECD Privacy Guidelines”), which were revised in 2013 and set out guiding principles to ensure the protection of privacy while avoiding restrictions on data flows that are disproportionate to the risks presented (see Box 3). The OECD Privacy guidelines were the first internationally agreed upon set of privacy principles on the protection of personal data whether in the public or private sector. They continue to be implemented by countries through legislation, enforcement and policy measures, and have influenced developments in privacy law, principle and practice in OECD countries and beyond. A growing number of countries have introduced privacy legislation in recent years, with many aligned with plurilateral arrangements such as the OECD Privacy Guidelines and the APEC Privacy Framework.

Box 3. The OECD Privacy Guidelines (2013 revisions)

Data flow governance has been a recurring focus of OECD work for over 40 years. Work in the 1970s led to the OECD's 1980 Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data ("OECD Privacy Guidelines"). The Guidelines are designed to ensure the protection of privacy whilst encouraging transborder flows of personal data with trust. They represent the first internationally agreed set of privacy principles that apply to the protection of personal data whether in the public or private sector. The Guidelines are drafted in technologically neutral language and are non-binding.

The 1980 Guidelines presumed that free transfers of personal data should generally be allowed, but recognised that they could be restricted when the receiving country "does not yet substantially observe the Guidelines or where the re-export of such data would circumvent its domestic privacy legislation" (paragraph 17 of the original Guidelines).

The 2013 revisions to the OECD Privacy Guidelines (OECD, 2013b) included important updates to the data flow governance provisions. With regard to free flow and legitimate restrictions, key principles are summarised in paragraphs 16 to 18 reproduced below:

> (16). A data controller remains accountable for personal data under its control without regard to the location of the data.
>
> (17). A Member country should refrain from restricting transborder flows of personal data between itself and another country where (a) the other country substantially observes these Guidelines or (b) sufficient safeguards exist, including effective enforcement mechanisms and appropriate measures put in place by the data controller, to ensure a continuing level of protection consistent with these Guidelines.
>
> (18). Any restrictions to trans-border 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 Guidelines also encourage states to co-operate on privacy matters and support the development of international arrangements that promote interoperability among privacy frameworks.

The Guidelines continue to be implemented by countries through legislation, enforcement and policy measures, and have influenced developments in privacy law, principle and practice even beyond OECD countries. For instance, the APEC Privacy Framework, which aims at promoting electronic commerce throughout the Asia Pacific region, is consistent with the core values of the OECD Guidelines, and reaffirms the value of privacy to individuals and to the information society.

A regional example of a non-binding plurilateral approach is the ASEAN Framework on Personal Data Protection (ASEAN PDP Framework), which sets out principles of personal data protection for ASEAN Member States to implement in their domestic laws. In 2018, building on the ASEAN PDP Framework, ASEAN endorsed the ASEAN Framework on Digital Data Governance[24] that sets out strategic priorities, principles and initiatives to guide ASEAN Member States in their policy and regulatory approaches towards digital data governance, including for cross border flows of all types of data (see Table 1 for participating economies).[25]

[24] https://asean.org/storage/2012/05/6B-ASEAN-Framework-on-Digital-Data-Governance_Endorsed.pdf

[25] Principles include facilitating cross-border data flows within ASEAN by developing unambiguous requirements that data can be transferred from one ASEAN Member State to another. In 2019, ASEAN commenced work to develop the various initiatives under the Framework for Digital Data Governance, including the ASEAN Cross Border Data Flows Mechanism that will see the creation of data transfer tools to facilitate intra-ASEAN data flows.

Another type of plurilateral arrangement is the 2014 African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention).[26] The Convention includes principles on personal data protection, which targets protecting privacy without prejudice to the principle of free flow of personal data. To date, 14 countries have signed the Convention and 5 countries have ratified it, while ratification of 15 countries is required for the Convention to enter into force (see Table 1 for participating economies).[27]

There are also binding plurilateral approaches with stronger enforcement mechanisms. For instance, the 1981 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, commonly referred to as Convention 108 of the Council of Europe, is a binding treaty protecting the right to privacy of individuals with respect to personal data which is automatically processed. To date, fifty-five states have committed to establish, under their own domestic law, sanctions and remedies for violations of the Convention's provisions (see Table 1 for participating economies). The 2018 Amending Protocol, when it enters into force, will update the provisions on the flow of personal data between signatories (creating what is commonly known as Convention 108+).

The APEC Cross-Border Privacy Rules (CBPR) System, in place since 2011, also has a binding element although it operates very differently.[28] The CBPR System is a government-backed data privacy certification framework that companies can join to demonstrate compliance with agreed privacy protection principles and enforcement mechanisms, allowing them to transfer data between CBPR participating economies with greater trust.[29] The CBPR System is not mandatory for APEC economies, and even when an economy adheres to it, companies can choose whether to seek certification under the System. However, once a company acquires the CBPR certification, it assumes liability under the CBPR framework.[30] To date, nine economies have participated to the APEC CBPR system and more than 30 companies have acquired the CBPR certifications – see Table 1 for participating economies.[31]

[26] https://au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf

[27] https://au.int/sites/default/files/treaties/29560-sl-AFRICAN%20UNION%20CONVENTION%20ON%20CYBER%20SECURITY%20AND%20PERSONAL%20DATA%20PROTECTION.pdf

[28] The current version of the Framework (2015) draws upon concepts introduced into the OECD Guidelines (2013) with due consideration for the different legal features and context of the APEC region.

[29] The APEC CBPR System requires participating businesses to implement data privacy policies consistent with the APEC Privacy Framework, a principles-based model for national privacy laws that encourages the development of appropriate information privacy protections and ensuring the free flow of information in the Asia Pacific region. The APEC Privacy Framework was first endorsed in 2005 and updated in 2015.

[30] Non-compliance may result in loss of CBPR certification, referral to the relevant government enforcement authority and penalties.

[31] See www.cbprs.com

Table 1. Examples of Plurilateral Arrangements

Non- binding plurilateral agreements	
OECD Privacy Guidelines	**ASEAN PDP Framework**
Australia, Austria, Belgium, **Canada**, Chile, Colombia, Czech Republic, Denmark, Estonia, Finland, **France, Germany**, Greece, Hungary, Iceland, Ireland, Israel, **Italy**, **Japan**, Republic of Korea, Latvia, Luxembourg, **Mexico**, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the **United Kingdom** and the **United States**.	Brunei, Cambodia, **Indonesia**, Lao, Malaysia, Myanmar, Philippines, Singapore, Thailand, Viet Nam.
Binding plurilateral agreements	
Malabo Convention **African Union Convention on Cyber Security and Personal Data Protection**	**Convention 108** **(Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data)**
The African Union Convention on Cyber Security and Personal Data Protection has not entered into force yet, the following are the ratifying countries as of latest available data published 28/06/2019: Ghana, Guinea, Mauritius, Namibia, Senegal.[32]	Albania; Andorra; Armenia; Austria; Azerbaijan; Belgium; Bosnia and Herzegovina; Bulgaria; Croatia; Cyprus[1]; Czech Republic; Denmark; Estonia; Finland; **France**; Georgia; **Germany**; Greece; Hungary; Iceland; Ireland; **Italy**; Latvia; Liechtenstein; Lithuania; Luxembourg; North Macedonia; Malta; Monaco; Montenegro; Norway; Netherlands; Poland; Portugal; Republic of Moldova; **Russian Federation**; Slovak Republic; Romania; San Marino; Serbia; Spain; Slovenia; Sweden; Switzerland; Turkey; Ukraine; United Kingdom; Argentina; Burkina Faso; Cabo Verde; Morocco; Mauritius; Mexico; Senegal; Tunisia; Uruguay.
APEC Privacy Framework	*2013 Additional Protocol to the Convention*
Australia; Brunei Darussalam; **Canada**; Chile; **China**; Hong Kong, China; **Indonesia**; Japan; Malaysia; **Mexico**; New Zealand; Papua New Guinea; Peru; The Philippines; the **Russian Federation**; Singapore; Republic of Korea; Chinese Taipei; Thailand; Viet Nam; and the **United States**.	Albania; Andorra; Armenia; Austria; Belgium; Bosnia and Herzegovina; Bulgaria; Croatia; Cyprus; Czech Republic; Denmark; Estonia; Finland; **France**; Georgia; **Germany**; Hungary; Ireland; Latvia; Liechtenstein; Lithuania; Luxembourg; North Macedonia; Monaco; Montenegro; Netherlands; Poland; Portugal; Republic of Moldova; the **Russian Federation**; Slovak Republic; Romania; Serbia; Spain; Sweden; Switzerland; Turkey; Ukraine; **Argentina**; Cabo Verde; Morocco; Mauritius; Senegal; Tunisia; Uruguay
APEC Cross-Border Privacy Rules (CBPR) System	*2018 Protocol amending the Convention*
The **United States**, **Mexico**, **Japan**, **Canada**, Singapore, **Republic of Korea, Australia**, Philippines and Chinese Taipei. with more expected to join soon	Bulgaria, Croatia, Lithuania

Note: G20 economies in bold. Data valid as of 16/03/2020

[1] Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

5.2. Trade agreements and digital trade partnerships

WTO agreements such as the General Agreement on Trade in Services (GATS) and the General Agreement on Tariffs and Trade (GATT) have bearing on data flows, as data measures may impact trade in goods, goods with embodied or embedded services and digitally enabled services. However, assessing the legality of measures on data can be complex (see OECD, 2019). While there are ongoing discussions at the WTO on specific e-commerce issues that include data flows, progress has been slow and cross border data flows are increasingly being addressed in regional trade agreements (RTAs).

[32] According to the most resent accessible official document online.

So far, the following trade agreements and digital trade partnerships include provisions on cross border data flows (members of G20 are highlighted in bold)[33]:

- CPTPP (**Australia**, Brunei, **Canada**, Chile, **Japan**, Malaysia, **Mexico**, New Zealand, Peru, Singapore, and Vietnam)
- USMCA (**Canada, Mexico and the United States)**
- **Korea**-**US** FTA
- **Australia -** Singapore FTA
- Chinese Taipei - Nicaragua FTA
- **Canada** - Peru FTA
- Caribbean Forum - EC EPA
- Cameroon - EC Interim EPA
- Hong Kong - New Zealand FTA
- **Korea** - Peru FTA
- Colombia - Costa Rica FTA
- Pacific Alliance Additional Protocol (PAAP)
- **Mexico** - Panama FTA
- **Canada** - **Korea** FTA
- **Japan** - Mongolia FTA
- **Korea** - Vietnam FTA
- Chile - Uruguay FTA
- **Argentina** - Chile FTA
- **Australia** - Peru FTA
- **EU**[34] - **Mexico** Modernised Global Agreement
- **Brazil** - Chile FTA
- **Indonesia** - **Australia** CEPA
- **Japan** - **US** Digital Trade Agreement
- (Digital Economy Partnership Agreement between Chile, New Zealand and Singapore (DEPA))35
- (**Australia** - Singapore Digital Economy Agreement)[36]

These agreements could be regarded as binding approaches with enforcement mechanisms, however, the depth and density of rules vary from one agreement to another. For instance, the recent Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) and the United States-Mexico-Canada Agreement (USMCA) include substantive rules on cross-border data transfers, local storage requirements and the protection of personal information. The agreements stipulate that cross border data transfers shall not be restricted, while allowing parties to maintain measures that achieve

[33] Came into force as of 30/11/2019. This part relied on Trade Agreements Provisions on Electronic-commerce and Data dataset, which is a part of National Research Programme (https://www.unilu.ch/en/faculties/faculty-of-law/professorships/managing-director-internationalisation/research/taped/).

[34] Although the EU is a member of the G20, each country constituting the EU is not highlighted here individually except France, Germany, Italy and the United Kingdom.

[35] Has not come into force yet.

[36] Has not come into force yet.

legitimate public policy objectives (provided measures are non-discriminatory nor unnecessarily trade restrictive). They also stipulate that use of domestic computing facilities shall not be required as a condition for conducting business.[37] These agreements also require parties to adopt a framework for the protection of personal information and encourage the development of mechanisms to promote compatibility between different privacy protection regimes. They also require parties to publish information on personal information protection.[38] These provisions are subject to dispute settlement mechanisms under each agreement unless otherwise provided.[39]

At the same time, new digital trade partnership agreements that include provisions on data flows are emerging. For instance, the Japan-US Digital Trade Agreement and the Digital Economy Partnership Agreement between Chile, New Zealand and Singapore (DEPA), include similar principles on cross-border data flows, computer facilities and the protection of personal information as those detailed above.

In sum, trade agreements and digital partnership agreements are increasingly incorporating provisions that refer to the flow of data and promote interoperability between data protection regimes. These often reference well-established principles requiring that approaches be transparent, non-discriminatory, not unnecessarily trade restrictive to meet desired policy-objectives and that are interoperable to promote cross border data flow. They also recognise the importance of personal data protection.

5.3. Unilateral instruments

Where cross-border personal data transfers depend on ex-ante conditions (see section 3.2), different unilateral instruments exist to facilitate the transfer of data with trust. One such instrument is an adequacy or equivalence determination, which is a unilateral recognition certifying that the data protection regime of another country meets certain privacy requirements and so data can be transferred unimpeded to this country. Adequacy or equivalence is frequently determined by a public body, such as the data protection authority (DPA), although it can also be evaluated by a data exporter in some jurisdictions. Examples of adequacy include the EU's determination that Israel provides an adequate degree of privacy protection, and the unilateral recognition by the EU and Japan that the other's data protection systems met their adequacy requirements. The Privacy Shield Framework between the US and the EU is another example of an adequacy decision. It gives companies operating in the United States the choice of adhering to privacy protection compatible with EU regulation, making it enforceable in the United States. This enables personal data to move between the two economies. The Privacy Shield also contains a number of commitments regarding the conditions under which the US authorities can access data transferred from the EU.

Even where an adequacy determination has not been made, alternative legal bases for data transfers unilaterally established by each country can enable cross-border transfers. These unilateral instruments range from (pre-) approved contractual safeguards to binding corporate rules (BCR) or standard exceptions (e.g. legitimate interest, data subject consent, public interest). Standard contractual clauses (SCCs) are ready-made rules that provide for data transfers to third-parties located in other countries. The clauses, designed to be incorporated into private-sector contracts, are developed by DPAs and, as such, are automatically considered to provide sufficient safeguards for transferring data, even to countries that

[37] The CPTPP allows parties to maintain measures inconsistent with this requirement to achieve legitimate public policy objectives (provided measures are non- discriminatory nor unnecessarily trade restrictive), while the USMCA does not include such an exception.

[38] These agreements include other principles that could promote data flow. For instance, the USMCA recognizes the importance of facilitating public access to and use of government information.

[39] Similar rules on cross border data transfer and local storage requirement can be found in the Financial Services Chapter of USMCA. This is an example of sector-specific data policies beyond the traditional digital trade chapters.

do not enjoy an equivalence or adequacy recognition. BCRs bind the affiliates of a multinational company located in different countries to apply effective rights and legal remedies for the protection of data. These rules also enable data to move between affiliates located in different countries, even when these are in countries that do not recognise each other's data protection systems. Transfers are, however, restricted to affiliates within the group. By taking advantage of these mechanisms, companies may still be able to transfer data to a country that does not benefit from an adequacy or equivalence decision although this will require some additional action on the part of the firm to activate the legal base.

5.4. Private sector and other initiatives

Other actors are also increasingly active in endeavours that could help promote cross border data flows. For instance, the International Organization for Standardization (ISO), an international standard-setting body composed of representatives from various national standards organizations, has developed a privacy protection ISO (ISO/IEC 27701:2019), which specifies requirements and provides guidance for establishing, implementing, maintaining and continually improving a Privacy Information Management System. Industry associations could also play a role in this area. For instance, the Business Software Alliance has developed a Privacy Framework as a guide for policymakers to draft privacy legislation, which includes a recommendation that governments create tools to bridge gaps among different domestic privacy regimes in ways that both protect privacy and facilitate the free flow of data. While further work is needed to analyse these and other private sector initiatives, standard setting by private sectors could promote cross-border data flows with trust that is built upon compliance with those standards.

Other approaches may also provide pathways toward trusted cross-border data flows between different organisations across different countries. For instance, data sandboxes that offer strong levels of control and protection on data, could technically be leveraged towards enabling cross-border access in the case of specific types of data. *Data sharing partnerships,* including public-private partnership, could also be agreed internationally.

6. Making ends meet: data, trust and data regulation

Data is a critical resource that creates benefits for our economies and societies. Understanding how data creates value and how it supports economic activity and identifying the challenges that data raises is key to making the most out of the digital transformation.

Data is different, it cannot be depleted and can be shared and re-used by many different users and for many different purposes. This means that data sharing has the potential to give rise to huge economies of scale and scope. But we sometimes treat data as a monolithic entity, forgetting that different data raise different issue and that there are overlapping data domains, each subject to different data governance frameworks.

Moreover, as more and more data crosses borders, new challenges emerge. Data and digital activity are inherently borderless but regulations are not. This raises concerns about ensuring privacy and security, protecting intellectual property, economic development and maintaining the reach of regulatory and audit bodies as data flows across jurisdictions. In response to this, there is an increasing number of policies that affect the movement of data across borders. These have been introduced to meet various policy objectives and affect different types of data.

However, the multiplicity of applicable regimes, and the additional conditions introduced, create uncertainty for governments, businesses and individuals. This means that there is value in seeking to find grounds for data transfers that create an environment of trust and ensure that, upon crossing a border, data is granted the desired degree of protection.

As has been shown, a number of policy approaches have emerged to enable data to flow across borders while creating trust and mitigating potential risks. For instance, trade agreements increasingly contain provisions on the free movement of all types of data across borders but provide grounds for exceptions. Plurilateral arrangements and unilateral approaches have been developed to overcome restrictions on data flows in the context of privacy protection. Finally, other approaches have established mechanisms through which data-sharing can take place in the context of trusted environments, whether through sandboxes or privacy enhancing technologies.

Going forward, understanding the diversity of data and the evolving regulatory environment on cross-border data flows, is an important first step in helping identify and advance approaches for greater interoperability and /or convergence between existing domestic regimes, helping economies meet the dual goal of ensuring that data can flow across borders with trust.

To date, it has proven challenging to get international discussion on ways to bridge the very different positions on this sensitive issue. Hence, building up from areas where there is comfort, G20 commitment to this dialogue would help move the debate forward. In this context, there seems to be several areas which might deserve further consideration in this space.

First, there might be scope for further dialogue and international cooperation between G20 economies with a view to promoting greater sharing of experiences in the area of data policy, including in particular interoperability and transfer mechanisms and identifying commonalities between existing approaches and instruments used to enable data to flow across borders with trust. There is also value in reaffirming the importance of the interface between trade and the digital economy, including in discussions under the WTO. Finally, there is also scope for encouraging the exploration of using technological solutions, such as privacy enhancing technologies, to meet challenges arising from technological transformation.

While, as highlighted, the concerns raised by the digital transformation of an increasingly global economy are multiple, challenges related to regulatory differences across countries are not new. Overcoming these differences will require engaging in a process of international dialogue to identify appropriate and proportionate policy responses that address the different concerns that such regulatory differences raise, allowing to deliver an environment of data flows with trust.

References

Abrams, M. (2014), *The Origins of Personal Data and its Implications for Governance*, http://dx.doi.org/10.2139/ssrn.2510927. [25]

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. [55]

Banham, R. (2014), "Who Owns Farmers' Big Data?", *Forbes* EMC Voice, https://www.forbes.com/sites/emc/2014/07/08/who-owns-farmers-big-data/. [43]

Belissent, J. (2017), "Insights Services Drive Data Commercialization", *Insights*, https://go.forrester.com/blogs/17-03-08-insights_services_drive_data_commercialization/. [32]

Brynjolfsson, E. and K. McElheran (2019), "Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing", *Rotman School of Management Working Paper No. 3422397*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397. [6]

Cadwalladr, C. and E. Graham-Harrison (2018), "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach", *The Guardian*, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. [31]

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. [47]

Choi, H. and H. Varian (2009), "Predicting the present with Google trends", *Google Research Blog*, http://dx.doi.org/10.2139/ssrn.1659302. [45]

Chui, M., J. Manyika and S. Kuiken (2014), *What executives should know about open data*, http://www.mckinsey.com/industries/high-tech/our-insights/what-executives-should-know-about-open-data (accessed on 11 February 2019). [42]

CoE (1981), *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data*, ETS No. 108. [77]

Council of the European Union (2014), "Interinstitutional File: 2012/0011 (COD), 5879/14". [61]

Department for Business Innovation & Skills (2012), "midata: Government response to 2012 consultation", http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/34700/12-1283-midata-government-response-to-2012-consultation.pdf. [64]

Department for Business Innovation & Skills (2012), "midata: Impact assessment for midata", http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32689/12-944-midata-impact-assessment.pdf. [63]

Department for Business Innovation & Skills (2011), "Better Choices: Better Deals – Consumers Powering Growth". [38]

EU (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 (General Da*, O.J. (L 119) 32. [78]

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. [65]

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. [66]

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. [29]

France (2016), *Loi pour une République numérique*, http://www.senat.fr/leg/pjl15-744.html. [69]

Frischmann, B., M. Madison and K. Strandburg (eds.) (2014), *Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo*, Oxford Univerty Press. [28]

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). [74]

G20 (2017), *G20 Digital Economy Ministerial Conference*, https://www.bmwi.de/Redaktion/DE/Downloads/G/g20-digital-economy-ministerial-declaration-english-version.pdf?__blob=publicationFile&v=12 (accessed on 12 October 2017). [76]

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). [52]

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. [62]

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. [30]

Grzywaczewski, A. (2017), *Training AI for Self-Driving Vehicles: the Challenge of Scale*, https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/. [12]

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/. [46]

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. [40]

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. [70]

ISO/IEC (2017), *Information technology -- Cloud computing -- Interoperability and portability*, http://www.iso.org/standard/66639.html. [22]

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. [59]

Konsynski, B. and F. McFarlan (1990), "Information Partnerships—Shared Data, Shared Scale", https://hbr.org/1990/09/information-partnerships-shared-data-shared-scale. [41]

Konsynski, B. and F. McFarlan (1990), "Information Partnerships—Shared Data, Shared Scale", *Harvard Business Review*, https://hbr.org/1990/09/information-partnerships-shared-data-shared-scale. [68]

Lynskey, O. (2015), *The Foundations of EU Data Protection Law*, Oxford University Press. [60]

Lyons, S. (2006), "Measuring the Benefits of Mobile Number Portability", http://www.tcd.ie/Economics/TEP/2006_papers/TEP9.pdf. [67]

MGI (2016), *Digital globalisation: the new ear of global flows*, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20globalization%20The%20new%20era%20of%20global%20flows/MGI-Digital-globalization-Full-report.ashx. [11]

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. [19]

National Board of Trade (2014), *No Transfer, No Trade – the Importance of Cross-Border Data Transfers for*. [81]

National Board of Trade (2014), *No Transfer, No Trade – the Importance of Cross-Border Data Transfers for for Companies Based in Sweden*, https://unctad.org/meetings/en/Contribution/dtl_ict4d2016c01_Kommerskollegium_en.pdf. [8]

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. [13]

ODI (2017), *What is 'open data' and why should we care?*, https://theodi.org/article/what-is-open-data-and-why-should-we-care/. [71]

OECD (2019), *Digital Opportunities for Trade in the Agriculture and Food Sectors*, https://doi.org/10.1787/91c40e07-en. [82]

OECD (2019), *Digital Opportunities for Trade in the Agriculture and Food Sectors*, https://doi.org/10.1787/91c40e07-en. [10]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, http://dx.doi.org/10.1787/276aaca8-en. [4]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, http://dx.doi.org/10.1787/276aaca8-en. [9]

OECD (2019), *Trade and Cross-Border Data Flows*, https://doi.org/10.1787/b2023a47-en. [5]

OECD (2018), *Digital Trade and Market openness*, https://doi.org/10.1787/1bd89c9a-en. [49]

OECD (2018), *Digital Trade and Market Openness*, p. 61, https://doi.org/10.1787/1bd89c9a-en. [3]

OECD (2017), *Digital Trade: developing a framework for analysis*, https://doi.org/10.1787/524c8c83-en. [80]

OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264276284-en. [15]

OECD (2017), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264271036-en. [2]

OECD (2016), *Declaration on the Digital Economy: Innovation, Growth and Social Prosperity (Cancún Declaration)*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0426. [73]

OECD (2016), *Ministerial Declaration on the Digital Economy (Cancún Declaration)*, https://www.oecd.org/internet/Digital-Economy-Ministerial-Declaration-2016.pdf (accessed on 12 October 2017). [75]

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. [18]

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), *Summary of OECD Expert Roundtable Discussion on "Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking"*, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg%282014%293&doclanguage=en. [26]

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. [16]

OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013 - C(2013)79, http://webnet.oecd.org/OECDACTS/Instruments/ShowInstrumentView.aspx?InstrumentID=114. [53]

OECD (2013), "The App Economy", *OECD Digital Economy Papers*, No. 230, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3ttftlv95k-en. [39]

OECD (2011), *Recommendation of the Council on Principles for Internet Policy Making*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0387. [54]

OECD (2011), *Thiry Years after the OECD Privacy Guidelines*, http://www.oecd.org/sti/ieconomy/49710223.pdf. [21]

OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, C(2008)36, http://webnet.oecd.org/OECDACTS/Instruments/ShowInstrumentView.aspx?InstrumentID=122. [56]

OECD (2006), *Recommendation of the Council concerning Access to Research Data from Public Funding*, C(2006)184, http://webnet.oecd.org/OECDACTS/Instruments/ShowInstrumentView.aspx?InstrumentID=159. [34]

OECD (2005), *Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, http://www.oecd.org/sti/sci-tech/38500813.pdf. [35]

OECD (2001), *Private sector*, https://stats.oecd.org/glossary/detail.asp?ID=2130. [57]

OECD (1985), *Declaration on Transborder Data Flows*, C(85)139, http://webnet.oecd.org/OECDACTS/Instruments/ShowInstrumentView.aspx?InstrumentID=108. [51]

OECD (1980), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013 - C(2013)79, http://webnet.oecd.org/OECDACTS/Instruments/ShowInstrumentView.aspx?InstrumentID=114. [58]

OECD Publishing, P. (ed.) (2014), *Measuring the Digital Economy: A New Perspective*, OECD, https://doi.org/10.1787/9789264221796-en. [14]

Ohm, P. (2009), "The rise and fall of invasive ISP surveillance", *University of Illinois Law Review* 1417. [20]

Open Knowledge International (n.d.), *Open Data*, http://opendatahandbook.org/glossary/en/terms/open-data/ (accessed on 5 February 2019). [72]

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). [27]

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. [44]

Schimmelpfennig, D. and R. Ebel (2016), "Sequential adoption and cost savings from precision agriculture", *Journal of Agricultural and Resource Economics*, Vol. 41/1, pp. 97-115, http://www.waeaonline.org/UserFiles/file/JAREJanuary20166Schimmelpfennigpp97-115.pdf. [7]

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). [23]

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). [17]

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. [33]

United Nations Global Pulse (2012), *Big data for development: Opportunities & challenges*, http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf. [48]

United States Department of Energy (n.d.), *Green Button: Open Energy Data*, https://www.energy.gov/data/green-button (accessed on 6 September 2018). [37]

US, International Trade Administration (n.d.), *Privacy Shield Overview*, https://www.privacyshield.gov/Program-Overview. [79]

Waters (2015), "IBM's latest deal is a new test case for the big data economy", *Financial Times*, http://www.ft.com/content/0fe3ac2e-7e22-11e5-a1fe-567b37f80b64. [50]

WEF (2014), *Rethinking Personal Data: A New Lens for Strengthening Trust*, http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf. [24]

www.oecd.org/innovation
www.oecd.org/trade

@OECDinnovation
@OECDtrade

STI.contact@oecd.org
TAD.contact@oecd.org

OECD
BETTER POLICIES FOR BETTER LIVES