Organisation for Economic Co-operation and Development

DSTI/CDEP(2022)6/FINAL

Unclassified

English - Or. English

19 January 2023

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
COMMITTEE ON DIGITAL ECONOMY POLICY

Data Stewardship, Access, Sharing and Control: A Going Digital III module synthesis report

Revised Draft

JT03510971

Foreword

This report demonstrates how approaches to data stewardship and control that are more balanced and differentiated can protect individuals' and organisations' rights, while enabling significant social and economic benefits – including addressing public health emergencies such as the COVID-19 pandemic and achieving the Sustainable Development Goals. It presents the mix of technical, organisational and legal approaches that characterises these more balanced and differentiated approaches, and how governments have implemented them.

This report was drafted by Christian Reimsbach-Kounatze (Directorate for Science, Technology and Innovation) with contributions from Tiago Cravo Oliveira Hashiguchi and Jillian Oderkirk (Directorate for Employment, Labour and Social Affairs) as well as Eleanor Carey and Ida McDonnell (Development Co-operation Directorate), Stéphan Vincent-Lacrin (Directorate for Education and Skills), Brigitte Acoca, Christian Biesmans, Nicholas McSpedden-Brown, Alan Paic and Jan Tscheke (Directorate for Science, Technology and Innovation), and Cecilia Emilsson, Jacob Arturo Rivera Perez and Barbara Ubaldi (Public Governance Directorate). Luke Slawomirski provided comments, and Mark Foss, Sebastian Ordelheide and Angela Gosmann provided editorial support.

This report is a contribution to IOR 1.3.1.2.3 of the 2021-2022 Programme of Work and Budget (PWB) of the Committee on Digital Economy Policy. It was approved and declassified by the Committee on Digital Economy Policy on 27 September 2022 and prepared for publication by the OECD Secretariat.

This publication is a contribution to Phase III of the OECD Going Digital project, which aims to provide policy makers with the tools they need to design and implement better data policies to promote growth and well-being.

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

#GoingDigital

Note to Delegations:

This document is also available on ilibrary as OECD (2022), "Responding to societal challenges with data: Access, sharing, stewardship and control", *OECD Digital Economy Papers*, No. 342, OECD Publishing, Paris, https://doi.org/10.1787/2182ce9f-en

Table of contents

Figures

Boxes

Executive summary

Overview

Data access, sharing and re-use ("data openness") can generate significant social and economic benefits. For example, they can empower individuals and enable data-driven innovation in the private and public sector to improve scientific and education outcomes. They can also manage public health emergencies such as the COVID-19 pandemic, and enhance energy resilience and the clean energy transition. Through all these benefits, data openness can contribute to achieving the Sustainable Development Goals.

However, data openness also comes with risks to individuals and organisations. These include risks to privacy and data protection, intellectual property rights, and digital and national security. Additionally, data openness raises ethical concerns. Data access, sharing and re-use may undermine ethical values and norms such as fairness, respect for human dignity, autonomy, self-determination. Consequently, data openness may lead to undue bias and discrimination. While rising granularity of personal data may provide for higher value, for example health care and scientific research, risks tend to increase correspondingly as even de-identification techniques cannot fully eliminate the risks of breaching the confidentiality of personal data.

This report explores societal challenges with data relating to access, sharing, stewardship and control. It examines the need to balance potential risks with benefits, and to develop a whole-of-government perspective in data strategies. Ultimately, it proposes a mix of technical, organisation and legal approaches for effective data stewardship and control.

Findings

A balance between potential risks and benefits is needed

Some restrictions to data openness can be necessary and justified to mitigate the risks they create. However, unjustified and unintended barriers can also disproportionately limit the social and economic potential of data. In so doing, they can create significant social and economic opportunity costs.

Stakeholders face a multitude of barriers to data access, sharing and re-use

Barriers to access, sharing, and re-use of data can take many forms: technical (lack of interoperable standards), legal (complex laws and regulations, and uncertainties about applicable regulatory frameworks), incentives (inability to recuperate sufficient returns on investments on data), skill and capacity gaps (inadequate data skills and poor access to data storage and processing infrastructures) and cultural (a poor culture of responsible data sharing and risk management).

Few countries have strategies to access and share data

Open data has been a priority for all OECD countries and partner economies to support business innovation, social value creation and government transparency. However, fewer countries have adopted broader national data strategies, including in the public sector. In 2018, up to 80% of OECD countries had a strategy for open government data, and 90% had requirements for public sector organisations to publish open data in a machine-readable format. In contrast, only 10% of OECD countries had a dedicated, comprehensive public sector data strategy covering a broader spectrum of data access and sharing arrangements, and their enablers.

Recommendations

Recognise vulnerabilities of consumers

When developing data governance frameworks based on transparency, consent and data control by individuals, policy makers should recognise two important factors. First, individuals can be affected by biases, including through information overload, which can impact their ability to take rational privacy and data-related decisions. Second, some business practices leveraging consumer data can create or exploit consumer vulnerabilities, including through so-called dark patterns on line. These practices are based on personalised advertising, pricing, ranking of offers or user interfaces.

Incorporate a whole-of-government perspective in data strategies

National and sectoral data strategies can facilitate more balanced and differentiated approaches and help address challenges in a comprehensive manner by incorporating a whole-of-government perspective. They can help create the conditions for effective data governance frameworks to better protect the rights and interests of individuals and organisations, while providing the flexibility needed for all to benefit from data openness.

Embrace more balanced and differentiated approaches to data stewardship and control

More balanced and differentiated approaches to data stewardship and control are needed to maximise the benefits of data. At the same time, these approaches should protect individuals' and organisations' rights and consider other legitimate interests and public policy objectives. This requires a mix of technical, organisation and legal approaches for effective data stewardship and control. These comprise three different types of measures. Technological measures include privacy-enhancing technologies such as tools for federated and distributed analytics. Organisational measures include developing trusted data intermediaries such as data trusts and personal data stores. Legal measures include for instance a right to data portability and legally binding and enforceable obligations to protect the rights of stakeholders. All these measures need to complement each other.

1 Introduction

The generation, collection and use of data are accelerating, driven in part by the digital transformation of key sectors and adoption of disruptive digital technologies such as artificial intelligence and the Internet of Things. The COVID-19 pandemic has been a major catalyst of these trends, highlighting the need for effective cross-sectoral and cross-border policies. This section introduces the societal challenges of data access, sharing, stewardship and control that are explored in this report.

1.1. The use of data emerge as a public policy priority

Two trends have accelerated the generation, collection and use of data. Key sectors such as education, energy, retail, transports, health, finance and government services are undergoing a digital transformation. Meanwhile, in a related development, disruptive digital technologies such as the Internet of Things (IoT) and artificial intelligence (AI) are being increasingly adopted.

The coronavirus (COVID-19) pandemic has been a major catalyst of these developments. In accelerating the digital transformation, the pandemic has revealed the critical need for cross-sectoral and cross-border access, sharing and re-use of data. In light of these developments, data access, sharing and re-use and the associated challenges to data stewardship and control increasingly become key public policy priorities for our societies.

The report is structured as follows:

- Section 2 introduces data access, sharing and re-use as key factors of "data openness." It presents their economic and social benefits for individuals, organisations and society at large, focusing on specific policy domains in dedicated subsections. These include the use of data for better consumer decision making, as well as enhancing business and consumer-driven innovation, health, science and research, education, government services, smart cities and development outcomes.
- Section 3 presents the risks and challenges that come with data openness, focusing on the risk of the violation of privacy and personal data protection rights, confidentiality and intellectual property rights (IPRs), and digital security. These risks are typically associated with the potential loss of control over data resulting from data openness and can lead to erosion of trust if left unaddressed.
- Section 4 discusses the importance of fostering trust while addressing unjustified barriers to data openness. Some restrictions to data openness can be necessary, and thus justified, to better control the risks of data openness. However, unjustified and unintended barriers can disproportionately limit the social and economic potential of data, thereby creating significant social and economic opportunity costs. The section addresses technical, legal, incentive, skill and cultural barriers and presents promising approaches to address them.
- Given the issues discussed in the previous sections, more balanced and differentiated approaches to data stewardship and control are needed. These should maximise the benefits of data, while protecting rights and considering other legitimate interests and public policy objectives. These approaches are discussed in Section 5 with a focus on national and sectoral data strategies.

Throughout the above-mentioned sections, the report draws on examples from OECD and non-OECD countries. In so doing, it provides insights that will help countries realise the full benefits of data openness. It also takes into account the digital realities of low- and middle-income countries and the particular risks their populations may face.

2 The social and economic benefits of data access, sharing and re-use

Data support and enable global and local activities on a large scale. This section examines the various types of economic and social benefits generated through access, sharing and re-use of data for individuals, organisations and society. It looks particularly at how use of both public and private sector data can empower users and enhance innovation in areas such as education, health, science and research, smart cities, government services and development outcomes.

Data have become the underlying foundation of today's knowledge economies and societies. Similar to transportation and communication systems, data are a "shared means to many ends" (Frischmann, 2012[1]) that support and enable global and local activities around the world and in a wide range of social and economic areas. In fact, the economic properties of data, in particular its nature as a non-rivalrous, general purpose[1] capital good,[2] suggest that data may be considered as an infrastructural resource (OECD, 2015[2]).[3] Access to this resource is crucial for growth and well-being in the 21st century.

Data access, sharing and re-use can be characterised by different levels of openness (data openness), ranging from full and unrestricted data access and use (e.g. open data) to arrangements that restrict access to and use of data to fewer or more specific users and/or purposes. Quantification of the overall benefits remains challenging. However, available evidence strongly suggests that greater data openness generates positive social and economic benefits for data providers (direct impact), their suppliers and data users (indirect impact) and for the wider economy (induced impact) (OECD, 2019[3]). The magnitude of the relative effects will vary depending on the sector (public vs. private sector) and the type of effect.[4]

Data openness generates benefits in six ways:

- contributing to greater efficiency, transparency and accountability across society
- providing support to address societal challenges and global emergencies such as the COVID-19 pandemic (see Box 2.1 on how data can help address the global food security challenge)
- boosting sustainable growth and enhancing social welfare and well-being
- improving evidence-based policy making, as well as public service design and delivery
- empowering users of digital goods and services, including enterprises, workers, citizens and consumers
- facilitating scientific discovery through enhanced opportunities for research, reproducibility of scientific results and cross-disciplinary co-operation (OECD, 2022[4]).

The following sections highlight these various types of economic and social benefits for individuals, organisations and society. They are related to the use of (public and private sector) data for empowering users and enhancing innovation, education, health, science and research, smart cities, government services and development outcomes.

Box 2.1. How the use and sharing of data can contribute to global food security

Urgent action is needed to address global food security[5] and to end hunger and malnutrition, which is among the greatest challenges humanity faces as recognised in the second UN Sustainable Development Goal. In 2020, between 720-811 million people were undernourished, around 118 million more people compared to 2019 (FAO, 2021[5]). Agricultural food production continues to face mounting pressures since the outbreak of the COVID-19 pandemic (OECD, 2020[6]). Climate change, labour shortages, and ecosystem degradation only add to the stress (McFadden et al., 2022[7]). Moreover, Russia's war against Ukraine,[6] supply chain disruptions and intensifying inflationary pressures are pushing food prices to all-time highs. Lower-income countries bear the brunt of this burden (OECD, 2022[8]; World Bank, 2022[9]; OECD, 2022[10]).

Against this background, the use and sharing of data can contribute to improving global food security by enhancing productivity, sustainability and resilience in the agricultural sector; and monitoring the state of global food security for better policy making and actions.

Data for enhancing productivity, sustainability and resilience in the agricultural sector

In agriculture, the use of data in combination with artificial intelligence (AI) holds significant promise to enhance productivity, sustainability and resilience. However, challenges remain, ranging from barriers to adoption of digital technologies and mistrust of technologies (OECD, 2017[11]; McFadden et al., 2022[7]). Optimising the supply and use of agriculture-related resources is key to achieving this objective. These include the efficient use of seed, fertiliser and irrigation, as well as farmers' savings in time. One estimate suggests an increase in net returns of 24% over operating with conventional machinery when including both input savings and a yield increase due to reduced compaction (Shockley, Dillon and Shearer, 2019[12]).

Data for monitoring the state of global food security to inform policy and decision making

The use of data can help stakeholders from policy makers to farmers better understand the conditions of farming and make appropriate decisions. For example, AI analysis of data collected by drones and satellite imaginary is used to help monitor and analyse field conditions to deliver precise agricultural interventions like fertilisers, nutrients and pesticides (Goedde et al., 2020[13]). This analysis can also guide policy interventions, including in lower-income countries. Uruguay, for instance, uses satellite imagery analytics to monitor the sustainability of soil use (Schroeder, Lampietti and Elabed, 2021[14]).

Data can also inform policy making at the international level: The Agricultural Market Information System (AMIS, n.d.[15]) is an inter-agency platform launched with the help of the OECD and the Food and Agriculture Organization of the United Nations in 2011. It aims to exchange timely and comparable market and policy information to enhance food market transparency and policy response. AMIS provides a platform to assess global food supplies and co-ordinate policy actions. This has helped prevent unexpected price hikes and strengthen global food security (OECD, 2020[16]).

As another example, the Executive Office of the UN Secretary-General launched Global Pulse to better monitor the impacts of socio-economic crises (UN, n.d.[17]). Since 2011, UN Global Pulse in collaboration with the World Food Programme (WFP) and the United Nations International Children's Emergency Fund (UNICEF) undertook joint projects to enhance their monitoring capacities. HungerMap, for example, is the main monitoring system of the WFP (n.d.[18]). It combines key metrics from various data sources such as food security, weather, population size, conflict, hazards, nutrition and macro-economic data to help monitor and predict the magnitude of hunger in near real time. The resulting analysis is displayed on an interactive map that helps the broader humanitarian community to make more informed and timely decisions relating to food security.

2.1. Enabling new business models and empowering consumers

2.1.1. Benefits for businesses

Data openness can create new business opportunities for smaller and larger firms. In manufacturing, for instance, sensors can provide data to monitor, analyse and optimise the efficiency of machine operations. These data can also be shared with suppliers and clients to enable after-sale services, including for preventative maintenance and production control optimisation (Wiebe, 2017[19]; Wang et al., 2020[20]). In agriculture, geocoded maps of fields and real-time monitoring of agricultural activities, from seeding to harvesting, are used to raise agricultural productivity. The same sensor data are also shared, re-used and linked with historical and real-time data on weather patterns, soil conditions, fertiliser usage and crop features to optimise and predict agricultural production (Bunge, 2014[21]; Sykuta, 2016[22]; Jouanjean et al., 2020[23]).

Better access to open government data, for instance, can allow entrepreneurs to develop new innovative commercial and social goods and services (OECD, 2018[24]) (see section 2.5). For example, Transport for London (TfL), a local government body responsible for the transport system in Greater London (United Kingdom), provided open access to its data. Entrepreneurs then developed apps that used these open data to provide real-time traffic information for more accurate navigation systems. Deloitte (2017[25]) estimates that users saved the equivalent of GBP 70-90 million in time savings and GBP 2-3 million compared to other means of information (such as SMS) thanks to TfL's open data. Evidence shows that data access and sharing enable business opportunities for data intermediaries such as data brokers, mobile application (apps) and Personal Information Management Systems (OECD, 2019[3]; OECD, 2021[26]).

2.1.2. Benefits for consumers

Some intermediaries can empower consumers and allow them to participate more actively in the data value creation process. Either they give consumers more control over their own data or they act on their behalf when interacting with potential data users and other market participants. This includes, for example, consumer facing (e.g. price comparison) platforms. These platforms use price and consumption data to better inform consumers that wish to switch service providers or access new services (OECD, 2020[27]; OECD, 2021[26]). For example, consumers that can access or share their mobile phone usage patterns with different network providers may find it easier to identify and switch to the most relevant offers in the market (OECD, 2020[28]). The same benefits are viable wherever business-to-consumer interactions are characterised by complex and potentially personalised contractual relationships (e.g. banking, insurance and electricity).

When the data exchange between consumers and businesses takes place routinely, consumers can benefit continuously. For example, smart banking apps can help consumers overcome behavioural biases (e.g. inertia) by nudging them to settle their credit card debt in time and to avoid expensive overdraft fees (OECD, 2020[29]). Interesting apps are also emerging in the smart home environment. For example, AI-enabled thermostats can provide consumers with automated suggestions on how to save energy and money by slightly adjusting their heating patterns (OECD, 2018[30]; IEA, 2019[31]). Importantly, if data openness is introduced across sectors, there is also potential for complementary data uses. Smart banking apps, for example, combine spending data with data on energy or telecommunication usage patterns. This can highlight switching opportunities in real time or generate personalised inflation forecasts based on actual spending patterns.

Consumers can also benefit from access to data concerning them in other ways to make more sustainable or healthier consumption choices (see also section 2.7). For example, start-ups like *Evocco* are developing apps that inform consumers about the carbon footprint of their recent grocery purchases and help them

switch towards more environmentally friendly alternatives. Other applications may warn consumers of products that are incompatible with their personal health conditions (e.g. known allergies). They might also provide suggestions for a healthier lifestyle and targeted training units based on their observable biodata (e.g. heartrate, glucose level, sleep patterns).

Finally, access to personal data has been at the core of personalisation. This includes customised products, like toys or clothes (OECD, 2017[32]). However, it also includes personalised online content and recommendations, which are often provided for "free" in exchange for personal data (OECD, 2019[33]; OECD, 2019[34]). Similar benefits can be perceived in the health sector. Providing consumers with better access to their own health data via portability rights, for example, can empower them (see section 2.2).

2.2. Monitoring health care systems and managing public health emergencies

Health data are necessary to improve the quality, safety and patient-centredness of health care services, to support scientific innovation, to enable the discovery and evaluation of new treatments and to redesign and evaluate new models of health service delivery (OECD, 2022[35]). The volume of personal health data in digital format continues to grow with technological progress. This includes electronic health and administrative records; behavioural and environmental monitoring devices and apps; and bio-banking and genomic technologies (OECD, 2022[35]).

2.2.1. Health data, big data analytics and artificial intelligence

Personal health data are increasingly linked and analysed via big data analytics and artificial intelligence (AI) to gain information and knowledge that can serve the health-related public interest. For example, AI can help improve diagnosis, particularly for rare diseases. It can identify optimal responders to treatment and personalising care for better patient outcomes. It can detect unsafe health care practices and treatments, and reward high quality and efficient health care practices. It can detect fraud and waste in the health care system and assess the long-term effects of medical treatments. Finally, it can discover and evaluate new health care treatments and practices (Oliveira Hashiguchi, Slawomirski and Oderkirk, 2021[36]; OECD, 2022[37]; OECD, 2022[35]).

2.2.2. COVID-19 sparks improvements in national personal health datasets

The COVID-19 pandemic shone a spotlight on the capacity of each country's health information systems to provide critical information for the public welfare. It also revealed aspects of data governance that created obstacles to responding to the pandemic in a timely way. Data sharing improved significantly within the public sector, sometimes through automated processes. Most OECD countries linked different data sources to monitor the COVID-19 pandemic and promoted open data policies. Timeliness of key national datasets was an area where countries almost universally drove data advancements as a result of the COVID-19 pandemic. Further, OECD countries widely reported improvements in the quality, coverage and completeness of national personal health datasets. These improvements were coupled with investments in health information systems and reporting tools (de Bienassis, 2022[38]). In particular, all countries expanded their capacity to provide timely statistics and information for policy decision making and public reporting.

2.2.3. Data are supporting a variety of health improvements

There were further improvements in access to health data for medical and health research. These included the development of health dataspaces or hubs that could approve access to health data, link health data and provide secure mechanisms for access to data. These innovations are numerous, including a Health

Data Hub in France, FinData in Finland, the Health Dataplace in Australia and OpenSAFELY in the United Kingdom (OECD, 2022[35]).

Furthermore, some countries have also explored means to leverage data portability to better empower patients. Gaining access to their health data in a machine-readable format can enhance consumers' agency and control to decide whether, with whom and how to share the data. Proposed changes to the Privacy Rule set out in the Health Insurance Portability and Accountability Act (HIPAA) of the United States – elaborated before the pandemic – would allow individuals to directly share their health care data between health care provider and health plans (HSS, 2020[39]; HSS, 2020[40]; Black, 2021[41]). This includes changes mandating the provision of electronic patient health information under specific circumstances to individuals at no charge (Black, 2021[41]).

An expansion in digitalisation and digital health services also occurred during the COVID-19 pandemic. The availability and volume of teleconsultations performed across OECD countries, driven by restrictions to mobility and social distancing, increased significantly. Meanwhile, online booking systems for COVID-19 testing and vaccination services also expanded (de Bienassis, 2022[38]). This rapid expansion, which produced new streams of health data, was made possible due to changes in reimbursement procedures, legislative reforms, relaxation of regulatory barriers and high-level political support.

2.3. Fostering new scientific methods, and the scrutiny and replication of scientific results

Data-intensive science can help address societal challenges such as climate, demographic changes and pandemics. Open science initiatives include efforts to foster unhindered access to scientific articles, access to data from public research and collaborative research enabled by open-source tools. Such initiatives are disrupting the way science is done, catalysing the creative process and removing the barriers to the diffusion of knowledge. This plays an essential role in accelerating needed scientific research and the innovation process itself. Open[7] sharing of scientific data is a recent phenomenon. It has gained momentum following adoption of the OECD Recommendation concerning Access to Research Data from Public Funding in 2006 (since revised). Some studies estimate the benefits from open research data at 0.15-0.4% of gross domestic product (Paic, 2021[42]).

2.3.1. Enhanced access to science data brings wide-ranging benefits

Enhanced access to data on science has wide-ranging benefits, including increased scientific discovery; improved reproducibility of scientific results and interdisciplinary co-operation; and transparency regarding disbursement of public funds, which in turn increases public trust in research in general. Moreover, access to open data promotes efficiency of resources in science, as well as economic growth and innovation (OECD, 2020[43]).

Digital technologies such as big data analytics and artificial intelligence (AI) offer new dimensions to data-intensive science. Often perceived as the fourth paradigm, digital technologies succeed the traditional empirical evidence paradigm, the model-based scientific theory and computational science. AI, for instance, enables simulations of theoretical predictions at unprecedented speed and scale. Indeed, as another indication of the increasing relevance of digital technologies to spur innovation and economic growth, the pharmaceutical industry expects AI to be the most important drug discovery tool by 2027. Increased adoption of big data analytics ultimately advances the predictability and deterministic nature of data in the social sciences. Some call it "social physics" given that increasingly "hard" data on human behaviour akin to the laws of physics will enable forecasting human responses to environmental changes. The underlying algorithms strongly depend on the accessibility of quality data management. Stewardship is also important as training algorithms requires large volumes of data (OECD, 2020[43]).

Enhanced access to data would also increase resource efficiency, alleviating publication bias and encouraging data re-use. As such, data resulting from scientific efforts deemed negative or non-significant would be still accessible. Consequently, they could prevent duplication of the same or similar scientific experiments. Furthermore, researchers or citizens could use available data to connect to other scientific endeavours, thereby enhancing cross-disciplinary scientific analyses. The data alone are seldom sufficient for reproducibility and need to be accompanied by relevant data analysis software.

Society at large stands to benefit from open access to data as research becomes more reliable, transparent and empowering of citizens and a greater contributor to resource efficiency, economic growth and innovation. Furthermore, enhanced access to data shows promises in addressing societal challenges such as climate change and biodiversity; greater monitoring capabilities of changing environments through improved data availability can inform policy making sooner and more effectively (Paic, 2021[44]). The recent revision of the OECD Recommendation concerning Access to Research Data from Public Funding, adopted by 41 countries, provides guidance in this regard (OECD, 2021[45]). Similarly, the immense significance of data sharing on a global scale was demonstrated through the combat against the COVID-19 pandemic. An integral part of the battle has been the quick and unencumbered access to data across countries and jurisdictions (OECD, 2020[46]).

2.4. Monitoring education systems and improving education processes

Like science, education has always been a data-rich activity. However, the scope for effective use of data for the improvement of learning and other educational outcomes such as equity or the personalisation of learning is undergoing a major and rapid boom. This is due to the emergence of new digital technologies and, in particular, AI- enabled applications (OECD, 2021[47]).

2.4.1. Artificial intelligence can transform teaching and learning in the classroom

In the classroom, intelligent tutoring systems assess students' knowledge, and diagnose what or how they should be learning before proposing a new task or unit to further develop their content or procedural knowledge (OECD, 2021[47]). Tutoring systems previously relied on traditional assessments of students' knowledge. With AI, systems are increasingly enriched to factor in students' engagement in learning, and metacognitive and other behavioural processes. To that end, they use sensors, cameras and sometimes analysis of how students go to the task. Such systems help teachers improve learning outcomes, and sometimes work especially well for students with lower prior academic achievement.

Classroom analytics usually monitor the entire classroom to give real time or delayed feedback to teachers. Real-time feedback helps teachers orchestrate their teaching. For example, it helps support them to time their transition from one task to the next, analysing whether students get disengaged or monitoring how many are finished with an individual task. Delayed feedback can give them feedback on their interactions with specific students, how they move in the classroom, how long they talk, etc. These systems give unique feedback that makes it visible to teachers how they behave in the classroom (rather than how they believe they behave). All the data collected in these types of apps are typically generated by learning apps/software, most often proprietary (OECD, 2021[47]).

2.4.2. Data can support early warning systems for school dropouts

At the system level, an increasing number of countries have developed systems that track students' learning over time, with a unique identifier for students and sometimes the possibility to link them to their teachers. Those information systems provide several opportunities to improve educational outcomes. For example, early warning systems use these data to try to identify when specific students are at risk of dropping out, notably from high school but also from higher education. Dropouts are a major problem for

OECD countries, which have too many young people not in education, employment or training. Typically, early warning systems identify students for which an intervention would be most necessary. More than that, they allow education stakeholders to redesign their policies.

In the United States, where the use of early warning systems has become relatively common, the availability of these educational data has also allowed a better understanding of the profiles of dropout students. While the usual profile of "jaded" dropouts (students who do not like school and have low and declining marks) represent about 40% of dropouts, the majority (50%) were described as "quiet" dropouts (students who like school with low and slowly increasing marks). Meanwhile, a small minority (10%) were "involved" dropouts (students who like school, have high grades, but sometimes for purely administrative reasons did not collect all their credits and decide to stop school) (Bowers, 2021[48]). In this case, better data allow for better understanding of the types of interventions that could help students to complete high school. These two examples show how using administrative data or data collected through private educational solutions can help improve learning outcomes, support teacher professionalism and performance or achieve better educational policy outcomes.

2.5. Enhancing public service delivery and addressing emerging societal needs

The digital transformation of the public sector and the associated growing adoption of digital technologies, including by governments, has coincided with an exponential increase in data generation and use (OECD, 2015[49]; Ubaldi et al., 2019[50]). This has expanded the possibilities for data access and sharing on many levels. This could involve co-creating public services in collaboration with external actors such as those from Govtech communities. It could improve and streamline core government functions through, for example, data interoperability and the sharing of basic data registers. It could also inform design and delivery of better policies and services and, where feasible and appropriate, automate decision making using algorithms to process these data at scale. At the same time, building blocks and tools such as digital identity and citizens' folders can help empower individuals by allowing them to know how data are handled within government, for what purpose and by whom (OECD, 2021[51]).

2.5.1. Strengthening base registries through open data

Within the public sector, data interoperability efforts facilitate data access and sharing among public bodies. One of the most important sources of data for public authorities are base registries. These are trusted and authoritative sources of information on data items such as people, companies, vehicles, licences, buildings, locations and roads (NIFO, 2022[52]). Base registries are used as the foundation for critical services such as legal identity, taxation, social benefits and business reporting.[8] How these data sources are governed is essential for the carrying out and delivery of public services but also for how governments manage personal data about citizens.

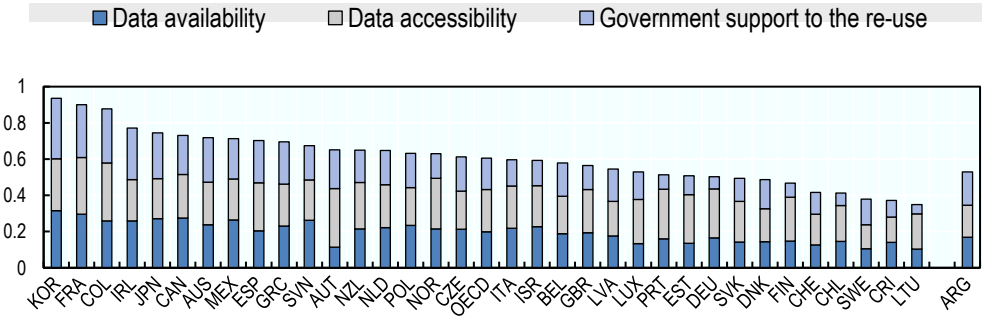
An important development in this context has been "open data." A growing number of OECD and non-OECD countries have scaled up adoption of "open by default" approaches to data access by including formal requirements in open data strategies, laws and regulations. This has contributed to stronger governance frameworks and the rise in data made available across the OECD (OECD, 2021[53]; Perez, Emilsson and Ubaldi, 2019[54]).

Since 2013, for example, Denmark's Basic Data Programme has provided free and open access to authoritative and trustworthy data sources from the public sector such as basic data registers. It has shown how opening up public sector data of high value can streamline data access and sharing processes for public bodies, including in terms of interoperability[9] (Agency for Data Supply and Infrastructure, n.d.[55]).

In general terms, evidence from OECD work on open government data points to steady progress among OECD member and partner countries in defining and implementing open government data strategies.

Results of the 2019 edition of the Open, Useful, Re-usable data (OURdata) Index (Figure 2.1) show that "the OECD average increased from 0.54 in 2017 to 0.60 in 2019, indicating a greater general maturity of open data policies at the central level" (OECD, 2020[56]).

Figure 2.1. 2019 OECD OURdata Index



Note: Data are not available for Hungary, Iceland, Türkiye and the United States.
Source: OECD Open Government Data Survey 2018. Data for Costa Rica was collected from the IDB-OECD Open Government Data Survey 2018.

2.5.2. Open government data supports the global response to COVID-19

Open government data has proven to be a foundation for the global response to the COVID-19 pandemic (OECD, 2020[46]; OECD, 2020[57]). Leading OECD countries in the area of open government data such as Korea[10] were quick to enable creation of citizen services from open data (including those released by private actors) (OECD and Govlab, 2021[58]). Evidence from OECD-Govlab (2021[58]) during the early stages of COVID-19 also showed the importance of publishing open data to facilitate situational awareness among the population during the pandemic and increase transparency of emergency public expenditure and procurement processes. Data visualisations supported with open government data unquestionably proved of great value during the pandemic in countries such as the Czech Republic,[11] Lithuania[12] and the United Kingdom.[13]

Yet open government data can also help improve trust in governments in more ordinary circumstances. Several international policy instruments have increasingly contributed to free data flows with trust, drawing upon open data policies as drivers of good governance and innovation. Early examples include the 2014 OECD Recommendation on Digital Government Strategies (OECD, 2014[59]) and the revised EU Directive on open data and the re-use of public sector information (European Union, 2019[60]). A more recent example is the proposal for a regulation at the European level on data governance (Data Governance Act) (European Commission, 2020[61]).

2.5.3. Strengthening public sector accountability and democracy through open data

The publication of standardised open data in areas such as public procurement, public budgeting and infrastructure can also contribute to public sector accountability. To that end, it can empower data-driven actors such as external auditors, journalists and civic watchdogs to monitor government activity and results. Not in vain, efforts to standardise open data have gained traction in recent years, including in the areas of open contracting data (Open Contracting Partnership, n.d.[62]), infrastructure (CoST Infrastructure Transparency Initiative, 2017[63]), beneficial ownership (openownership.org, n.d.[64]) and sponsorship (360Giving, n.d.[65]).

Open data is also a tool for reinforcing democracy (OECD, 2021[66]). On the one hand, the timely publication of open government data can help fill information gaps, supporting the fight against mis- and dis-

information. This nevertheless requires mainstreaming public communication efforts to channel information recipients to trusted data sources such as open government data portals. Moreover, access to and sharing of data about misinformation risks can help support "public-private access to and sharing of information and data on mis- and disinformation threats and risks … [through intermediaries such as] information sharing and analysis organisations (ISAO) or information sharing and analysis centres (ISAC)" (OECD, 2022[67]). Last, as discussed in section 2.1, open government data can support the creation of new business models and innovation in the private sector.

2.6. Improving the performance and sustainability of smart cities and regions

2.6.1. Using data to boost efficiency in cities and regions

Data collection and data use are key drivers of smart cities, defined as "cities that leverage digitalisation and engage stakeholders to improve people's well-being and build more inclusive, sustainable and resilient societies" (OECD, 2021[68]). Data-based smart city solutions have traditionally aimed to increase efficiency in urban transport, energy, water and waste management systems. However, data infrastructure underpinning smart city initiatives has also been instrumental during the COVID-19 pandemic. It has helped cities monitor the spread of the virus and enforce physical distancing requirements, maintain the continuity of public services, support economic activity, foster more sustainable urban development (e.g. via smart homes and smart urban mobility) and engage residents on line in urban planning decisions.

Beyond providing an emergency response to the pandemic in the short term, smart city tools relying on data have also become an important component of national and local strategies to shape recovery and resilience in the longer term. This includes using data to enhance resilience in respect to energy supply and consumption, and to accelerate clean energy transitions. Examples of this range from Florence's ambition to offer full digitalisation of municipal services to Bratislava's goal of achieving full coverage of its street lighting network by smart LEDs by 2025 (OECD, 2020[69]). The capacity to leverage the benefits of data and digital innovation for all urban residents will therefore be critical to help cities accelerate their transition to a more sustainable, inclusive and resilient urban paradigm.

2.6.2. Using data to develop smart municipal services

Data are also essential to better measure smart city performance. The proposed OECD Smart City Measurement Framework aims to measure the inputs and outputs of smart cities (e.g. public investment in broadband infrastructure in cities, or share of urban households who own a smartphone). At the same time, it would capture the impact of digital innovation on well-being outcomes for residents across multiple sectors. It would assess whether smart city initiatives benefit everyone rather than selected population groups. It would consider stakeholders' engagement in shaping smart cities. Finally, it would monitor progress over time and across places in a comparable way (OECD, 2021[68]).

This framework requires developing a range of data and indicators to document the relationship between smart city tools and life outcomes. This includes telemedicine and health outcomes; car-pooling or bike-sharing applications and air quality; smart surveillance and crime rates; smart detection of water leakages and water consumption; and real-time transport applications and commuting times.

At the same time, the increasing amount of information that cities are collecting about their citizens is raising new governance challenges. Who can access the data? Who owns them? For long can they keep them and how can they use them? If left unaddressed, such concerns about data security and privacy might undermine citizens' trust in local governments and the contribution of smart cities to a better future.

Effective data governance is therefore fundamental to guide evidence-based decision making in smart cities and build stronger urban communities. Both national and local governments need appropriate

frameworks to implement responsible data collection, sharing and use. This should be developed within the overarching objective of improving residents' well-being and fostering sustainable, inclusive and resilient cities.

2.7. Enhancing global development progress towards the Sustainable Development Goals

Sharing and re-using data can contribute to, and help measure, development outcomes. Increasing attention and support began to emerge for open government data in low- and middle-income countries around the middle of the last decade. This trend coincided with the beginning of the Sustainable Development Goals and the need to align national strategies and development co-operation approaches to the associated indicator framework.

In recent years, the sharing and re-use of data generated through and held by private sector companies have gained prominence. These data can provide relevant insights for decision making or as inputs for context-specific digital tools and services that meet citizen and government needs. Such needs can differ significantly in developing countries compared to high-income economies. Using data from mobile phones to map population movements, for example, has helped predict and combat COVID-19 in developing countries with limited track and trace capacities (Benjamins, Vos and Verhulst, 2022[70]).

2.7.1. Open data can support greater inclusiveness in developing countries

Digitised payment records of informal merchants have been used to generate alternative credit scores, increasing access to credit for those financially excluded from the traditional banking system (OECD, 2021[71]). Satellite data have proven useful to develop new methodologies to measure illiteracy (Verhulst, 2021[72]), and to detect unofficial roads and other risk factors for deforestation (OECD, 2021[73]).

Other innovations include use of open data sources to deliver financial support more efficiently to populations in need. The government of Bangladesh, for example, is experimenting with real-time poverty ranking and pre-emptive benefits transfers using satellite and telecom data for faster and more accurate targeting (OECD, 2021[74]). Data openness and code sharing have also made possible the evolution and adoption of digital public goods that can offer developing countries control over their data systems. Such measures provide critical cost-saving and capacity-building opportunities. Local solutions providers can also use digital public goods as a foundation for new apps and services (OECD, 2021[75]).

2.7.2. Building enabling environments across public and private sources to strengthen development

Increasingly, low- and middle-income countries are cultivating enabling environments (legal, policy, regulatory and technical) for greater data openness across public and private sources, which can in turn underpin economic and social development. The COVID-19 pandemic had a similar catalysing effect on digital transformation and data generation and use in low- and middle-income economies as it had in advanced economies. However, across all policy dimensions of the OECD Going Digital Integrated Framework, developing countries lag behind (Carey and Mc Donnell, 2021[76]).

Managing cross-border data flows created by digital trade transactions raise new regulatory and policy issues for developing countries. However, only 33 of 54 countries in Africa have formal e-transaction legislation (OECD, 2021[77]). Meanwhile, capacity for data governance, protection and security is low (OECD, 2021[78]). Although value added tax provides nearly 30% of government revenue in developing regions, most African countries are losing out by not updating rules to take account of e-commerce (OECD, 2021[79]). Actors are focused on creating context-appropriate and regionally harmonised policy and legal

environments to enable data openness and other digital innovations, especially to bolster digital transformation as a pillar of COVID-19 economic recovery.

Low- and middle-income countries must be more meaningfully represented in global standard-setting forums to maximise the benefits of data openness for their countries. A stronger role would enable them to shape standards on data governance that consider their digital realities. Developing countries are often underrepresented in global forums or do not consider the content relevant for their situations. Greater effort among the international community to create more inclusive forums could remove key barriers to progress for low- and middle-income countries. This, in turn, could create powerful opportunities for digital transformation to enhance global development progress.

3 The risks of data access, sharing and re-use

With all its social and economic benefits, data openness (access, sharing and re-use) also comes with risks to individuals and organisations. This section discusses these risks with a focus on risks of confidentiality and privacy breaches; violation of intellectual property rights and risks to national security; and violation of other legitimate public and private interests, including threats to endangered species. It also examines ethical concerns, especially where data access, sharing and re-use may be legal but still undermine ethical values and norms.

3.1. The violation of privacy and personal data protection rights

Violations of privacy are often considered the biggest risks associated with data access, sharing and re-use, especially where personal data are involved. Individuals and organisations may agree on (and consent to) specific terms for data sharing and data re-use, including on reasons for that re-use. However, their data may still end up being used differently by a third party, for instance.

For example, many mobile application (app) providers share app events, sometimes containing highly sensitive user data, with multiple downstream third-party service providers for marketing or analytics purposes. This potentially multiplies risks for privacy violations [see (FTC, 2021[80]) for a recent case involving user data from a fertility-tracking app]. The violation of these terms may not always be the result of malicious intent. Transferring personal data from one context (e.g. health) to another (e.g. marketing) can make it increasingly challenging to ensure that rights and obligations are not undermined.

Moreover, while rising granularity of personal data may provide for higher value (e.g. for health care and scientific research), risks tend to increase correspondingly. Personal health data are particularly sensitive. Although personal de-identification techniques are usually applied, full anonymisation is rarely possible. In addition, the ability to associate data points with individual characteristics offers particularly high benefits. For instance, hospitals in the United Kingdom increased cancer patient survival rates by modifying medical treatments based on connections between hospital data and the cancer-data repository (OECD, 2020[43]). Therefore, balancing data access, sharing and re-use with protecting the privacy rights of individuals and protection of their personal data is among the most important data governance challenges for policy makers.

This section discusses the risks of privacy violation mostly in the context of privacy and data protection regulations in OECD countries. It acknowledges that some non-OECD countries have weaker regulations, making it challenging to address violation of privacy, including personal data breaches. In some of these countries, personal data have been used to target and persecute groups of individuals. In many countries, according to the UN Human Rights Council (2021[81]), groups that uphold human rights face threats and harassment. The Council also points to the use of technological tools developed by the "private surveillance industry" to hack into devices or otherwise violate individuals' right to privacy. On an individual level, privacy, safety and security concerns are key reasons why some groups in developing countries are not making greater use of digital technology, particularly related to women and girls.

Against this background, privacy and data protection regimes have been weak or ineffective in many of these countries. Since 2010, 64 countries – most of which are in Africa, Asia and Latin America and over 70% of which are categorised as lower middle-income countries – have enacted new data protection laws. This brings the total with such laws in place up to 146. However, in many cases, enforcement is weak; regulatory authorities lack independence; and resources and policies are poorly designed (OECD, 2021[78]).[14] Sections 4 and 5 will highlight promising approaches to privacy and data protection regulation and enforcement that are aligned with implementation capacity in low- and middle-income countries, alongside support capacity and resource improvement.

3.2. The violation of intellectual property rights and other business interests

Data openness can present risks to the intellectual property rights (IPRs) of organisations, their contractual agreements with third parties and the protection of their commercial interests. These risks can negatively affect incentives to invest and to innovate. This is true even in cases where these risks may be the unintended consequences of business decisions.

Small and medium-sized enterprises (SMEs) perceive identifying which data to share and defining the scope and conditions for access and re-use as a major challenge. Inappropriate sharing of data can lead to significant costs to the organisation. In addition to fines for privacy violations, inappropriate data sharing can lead to opportunity costs due to a lower ability to innovate. For example, premature data sharing can undermine the ability to obtain IPR (e.g. patent and trade secret) protection (OECD, 2019[3]).

The protection of data by IPR remains controversial (Determann, 2018[82]; Scassa, 2018[83]), but there is consensus that data under certain forms and conditions can be protected by copyright and/or trade secret (OECD, 2022[84]).

Copyright typically "protects and rewards literary, artistic and scientific works, whatever may be the mode or form of their expression, including those in the form of computer programs" (OECD, 2015[85]). To the extent that data include protectable works (e.g. electronic maps, photographs and text), those data will be protected by copyright. Their access, sharing and (secondary) re-use therefore needs to respect the copyrights of the original data holders. With the increasing use of application programming interfaces

(APIs), which are implemented via software code, copyrights have gained further in importance as legal means for controlling data access and re-use.

Trade secrets encompass "confidential business and technical information and know-how that a firm makes reasonable efforts to keep secret and that has economic value as a result" (OECD, 2015[85]). Not all data can be protected as trade secret, but even where they can, the data can only be disseminated to authorised persons (subject to confidentiality agreements).

IPR frameworks (including copyright and trade secrets) apply to data only under certain conditions. However, contract law has become the primary legal vehicle for determining rights related to data access, sharing and re-use, in particular in business-to-business (B2B) contexts. Individuals and organisations may agree on (and consent to) specific terms for data sharing and data re-use, including on the purposes for which the data should be re-used.

However, there remains a significant risk that a third party may end up using data differently, either intentionally or not. The case of Cambridge Analytica illustrates this risk: personal data of Facebook users were used for a commercially motivated political campaign rather than for academic purposes (to which some users had consented). This occurred although Facebook explicitly prohibits data to be sold or transferred "to any ad network, data broker or other advertising or monetisation-related service" (Granville, 2018[87]).[15]

3.3. Lack of transparency, loss of control over data and the challenged role of consent

Once data are shared, unless specific data stewardship and processing provisions are in place, they move out of the control of the original data holder (Sundareswaran, Squicciarini and Lin, 2012[86]; Henze et al., 2013[87]). The same can be said for individuals, who provide their data and give consent for their re-use and sharing. In both situations, data holders and individuals, respectively, may lose their capabilities to control how their data are re-used and to object to or (technically) oppose such uses. The risks of loss of control are multiplied where the data are further shared downstream across multiple tiers of (secondary) data users, especially when these tiers are in multiple jurisdictions (OECD, 2022[88]).

3.3.1. Lack of transparency can amplify loss of control

The lack of transparency that may exist during data collection, sharing and re-use can further exacerbate loss of control. Data and the value derived from their use are often (co-)created during interactions of various parties in the global data ecosystem,[16] in some cases even without their awareness. This could happen in two different types of cases. In one, data from multiple sources are linked across organisational borders. In the other, users (businesses and consumers) interact with a digital product (service or good), such as a digital government service, a portable smart health device or a social networking service.

The level of awareness will be particularly low where personal information can be inferred based on big data analytics and artificial intelligence (AI). The ubiquitous nature of these technologies, coupled with sensors and the Internet of Things (IoT), have made it increasingly easy to generate inferences about individuals from data collected in commercial or social contexts. This is true, even if these individuals never directly shared this information with anyone and in some cases even if the data may not be personal data.

3.3.2. The challenge of consent

As a result, trustworthy and transparent data governance within and across jurisdictions is needed to protect individuals' privacy rights such as the protection of their personal data, and to earn public trust in data sharing and use (OECD, 2022[89]; OECD, 2020[90]). Transparency is not only crucial with respect to

what, how and by whom data are collected, accessed and used but also with respect to how data are governed. This includes clarifying rules and policies that affect data along the entire value cycle – from data creation, collection, storage, use to protection, access, sharing and deletion. Moreover, it includes information on rights (e.g. for information and redress), responsibilities and respective liabilities in case of right violations.

User consent is another fundamental governance principle for the collection, sharing, use and re-use of data in many jurisdictions. However, challenges can arise when i) the consent decision is not deliberate or well-informed; and ii) when data collection, sharing, use and re-use of data go beyond the consent of data subjects (i.e. loss of control).

The first risk is especially pronounced in the context of individuals (e.g. consumers, citizens) providing consent regarding the use of their personal data. In contrast to institutional data holders, individuals are more often required to take ad hoc consent decisions regarding the collection, sharing, use and re-use of their data (e.g. when accessing media content on line or making an online purchase). Individuals' decisions regarding their privacy can then be affected by misperceptions of the costs and benefits of data collection, sharing and re-use, as well as social norms, emotions and heuristics (Acquisti, Brandimarte and Loewenstein, 2015[91]).

Individuals are also limited in the amount of information they can process (information overload). In addition, they can be victim to behavioural biases. For example, they tend to underestimate the possible future risks of a privacy choice compared to the immediate benefits (present bias). They may also revert to default options when a decision becomes too complex (default bias) (OECD, forthcoming[92]). Consumers are particularly susceptible to behavioural biases when they face information overload (Tversky and Kahneman, 1974[93]). This is critical in the context of privacy policies that often require consumers to read long disclosure statements full of technical vocabulary to take an informed decision (Degeling et al., 2019[94]).

Furthermore, consumers may not read privacy policies if their only choice is to agree to extensive personal data sharing (often through a clickwrap agreement bundling a wide range of consents) or completely forgo access to the desired content or service ("take-it-or-leave-it") (ACCC, 2019[95]; ConPolicy, 2020[96]). Such difficulties for consumers to engage with privacy policies and know the ultimate extent of their data sharing call into question the extent to which they can meaningfully give consent in such transactions. In addition, a consent-based regime may encounter further challenges where consumers lack trust to share data and disengage from markets. This could lead to a sub-optimal level of data sharing overall (see section 3.6). Such a situation could raise as a result of a high prevalence of dark patterns and misleading data practices.

Dark commercial patterns are understood as business practices that use elements of digital choice architecture, especially online user interfaces, to subvert or impair consumer autonomy, decision making or choice. Such practices often deceive, coerce or manipulate consumers. They are likely to cause direct or indirect consumer detriment in various ways, although it may be difficult or impossible to measure such detriment (OECD, 2022[97]). They are common on e-commerce websites, apps (including those of major online platforms) and cookie consent notices. Several dark patterns seek to get consumers to give up more personal data than desired, e.g. through defaults or by making it harder to opt out of privacy-intrusive settings.

In addition, the data collected by businesses through online interaction with consumers increasingly allow them to build fine-grained consumer profiles. Businesses may increasingly be able to leverage information asymmetries from such data profiling to exploit consumer vulnerabilities at a highly granular level (though evidence does not indicate such practices are widespread). Consumer and data protection regulatory frameworks in OECD countries address many dark patterns and exploitative personalisation practices. However, new research has identified gaps and a range of new regulatory measures has been proposed or implemented in various jurisdictions (OECD, 2022[97]; forthcoming[98]).[17]

3.4. Ethical concerns associated with data openness

Some concerns and risks associated with data openness have been framed as ethical. This underscores the importance of issues such as fairness, respect for human dignity, autonomy, self-determination and the risk of bias and discrimination as a complement to regulatory actions. It includes issues related to the exploitative targeting of vulnerable socio-economic groups to encourage certain behaviours or consumer decisions.

Data ethics has been highlighted in cases where the collection, processing and sharing of data are legal but may generate moral, cultural and social concerns with potential direct or indirect adverse impacts on individuals or social groups. As discussed in the OECD Good Practice Principles for Data Ethics in the Public Sector, issues around data ethics sit on a broad spectrum. This ranges from how data are generated, their disaggregation and granularity (so they are representative of all population groups and "no one is hidden in the data") to how the data are selected, managed and governed when informing development of AI systems (OECD, 2021[51]).

In recent years, the representation of indigenous communities and their autonomy, and how these are reflected in the data and data governance arrangements, has gained traction. For instance, Mexico's National Institute of Statistics, Geography and Informatics collected data on individuals' self-identification as members of afro-Mexican communities in 2019 for the first time. Meanwhile Canada, (FNIG, n.d.[99]), New Zealand (Māori Data Sovereignty Network, n.d.[100]) and Australia (Bodkin-Andrews et al., 2019[101]) have been working on indigenous data sovereignty for some years already.

These ethical concerns are pronounced in low- and middle-income countries withsystemic digital divides. For instance, women in such countries are 16% less likely than men to use mobile Internet. At the same time, lack of connectivity, the high cost of mobile broadband and digital devices disproportionately impacts low-income individuals and households, especially in rural areas (Cruz and Tiel Groenestege, 2021[102]). Thus, those societal groups that are in most need of intervention can often be rendered invisible in even the most basic datasets.

As well, in low- and middle-income countries, data on which AI systems are trained are often drawn from and representative of the advanced economies in which AI tools were developed. The need for locally relevant training data and analytical approaches that reflect the lives of all social groups is only beginning to be addressed (Carey and Mc Donnell, 2021[76]). Other ethical risks of open data are the misuse of sensitive data, including of endangered species or rare materials by third parties (OECD, 2020[43]).

3.5. Digital security risks and confidentiality breaches

Data openness typically requires opening information systems so that data can be accessed, shared and re-used. This process may expose parts of an organisation to digital security threats. These, in turn, can lead to incidents that disrupt the availability, integrity or confidentiality of data and information systems on which economic and social activities rely.

Available evidence confirms the risk of digital security incidents is growing with the intensity of data use (OECD, 2017[103]). The actual proportion of the impact varies significantly, however, depending on the motivation and form of the incidents. Organised crime groups may target valuable assets they can sell on illegal markets, for example. As innovation is becoming more and more digital, industrial digital espionage is also likely to rise. In some cases, the motive may be political or the attacks may be designed to damage an organisation or an economy (OECD, 2017[103]).

Access to and sharing of personal data can increase risk of data breaches.[18] They can cause harm because of the privacy violation of the individuals whose personal data have been breached (see section 3.1). Moreover, they can also cause significant economic losses to the business affected (including loss of

competitiveness and reputation). In addition, further consumer detriment may result from a data breach, such as harm caused by identity theft.[19]

Data breaches are less frequently experienced than other types of digital security incidents such as malware, phishing and social engineering, or denial of service[20] attacks. Evidence from Privacy Rights Clearinghouse suggests the total number of identified incidents may be relatively small compared to other security incident types. However, it adds that their impact is increasing drastically. Large-scale data breaches, which involve more than 10 million records, are becoming more frequent. This is confirmed by available evidence suggesting that data breaches have increased with the collection, processing and sharing of large volumes of personal data (OECD, 2017[103]).

For many low- and middle-income countries, strengthening digital security became an area of focus only recently. Just 11 African countries have adopted substantive laws on cybercrime. This leaves most Africans exposed and undermines business confidence to operate in online economic space governed by low- and middle-income countries (Carey and Mc Donnell, 2021[76]).

4 Fostering trust while effectively addressing unjustified barriers

Countries need effective and trustworthy data governance to balance the benefits of digitalisation and data openness with the associated risks. However, some restrictions on the access, sharing and re-use of data can create significant economic and social opportunity costs, and even lead to unethical outcomes. In some cases, such barriers can also have negative effects on society, such as restrictions on health-relevant data to address pandemics like COVID-19. This section explores challenges identified by countries to use data in critical areas such as health care, including the need to overcome technical, legal, incentive, skill and cultural barriers.

4.1. Governance needs to address both benefits and risks of data openness

Given the benefits of digitalisation and data openness and the legitimate concerns and risks, countries need to build effective and trustworthy data governance arrangements. This includes regulations, policies and practices as part of data governance initiatives, including national data strategies. Without effective approaches, stakeholders can face significant barriers when trying to access, share and re-use data within and across organisations, sectors and jurisdictions.

Some of these barriers are needed to address risks of data openness such as violation of privacy, data protection and intellectual property rights. However, others may be unjustified and/or unintended barriers

that disproportionately limit the social and economic potential of data. This can create significant social and economic opportunity costs and, in some cases, unethical outcomes. Furthermore, these barriers can have negative effects on society. For instance, they could prevent access to and sharing and re-use of health-relevant data needed to address health crises such as the COVID-19 pandemic.

The OECD (2016[104]) Recommendation on Health Data Governance provides a roadmap towards more harmonised approaches to health data governance across countries. OECD (2022[35]) monitored the first five years of implementation of the Recommendation. It found only a small cluster of Adherents with policies, regulations and practices that fostered development, use, accessibility and sharing of key national health datasets for research and statistical purposes, while also having a high degree of recommended health data governance policies and practices. Adherents reporting the strongest national health data availability, maturity and use and health dataset governance policies and practices were Denmark, Finland and Korea. In general, however, the health sector remains significantly behind other economic sectors such as transportation, travel, banking and finance in the interoperability of data (OECD, 2022[35]).

Countries reported various challenges to develop health data and information. The following ranked highly: lack of human resources, funding shortfalls, poor data standardisation and quality, lack of policy planning or strategy to support digitalisation, and legislative barriers. Further, the rapid expansion in health data sharing and access during the pandemic raises data governance uncertainties. It is unclear whether mechanisms to collect, analyse, anonymise and share personal health data are sufficiently secure and effective (OECD, 2020[105]; OECD, 2020[106]). This section highlights barriers typically faced by stakeholders, focusing on technical, legal, incentive, skill and cultural barriers.

4.2. Identifying and overcoming technical barriers

4.2.1. The need for technical interoperability

Various studies have recognised inconsistent data formats as an impediment to the creation of certain datasets. This includes longitudinal datasets, where changes in measurement and collection practices would make it hard to compare and aggregate data (OECD, 2019[3]; OECD, 2019[107]; OECD, 2022[35]; OECD, 2018[24]). These inconsistencies can be a significant barrier to interoperability, although these data often need to be shared and re-used across information systems or different sectoral applications.

Even when commonly used machine-readable formats are used for accessibility, interoperability is sometimes not guaranteed. These common formats may enable *syntactic* interoperability, i.e. the transfer of “data from a source system to a target system using data formats that can be decoded on the target system” (and thus accessibility) (OECD, 2019[3]). However, they do not guarantee *semantic* interoperability, which is defined as “transferring data to a target such that the meaning of the data model is understood” (OECD, 2019[3]).

4.2.2. Building interoperable specifications including common technical standards

Governments and international organisations can play an important role in fostering creation of common standards, including across sectoral barriers, and ensuring their implementation. This is particularly the case if the potential benefits of interoperability will only slowly emerge over time. Business incentives to invest in the development of common standards, for example, may remain low when incumbent firms and data holders benefit from limited interoperability. In this respect, the development of common standards can be seen as similar to an infrastructure investment.[21] Once a common standard has been defined, it needs to be adopted. Ensuring wide stakeholder participation in the standard-setting process can be critical in this regard.

Additionally, an independent oversight body can help ensure that application programming interfaces (APIs) of all participants, such as with open data initiatives, are compatible with adopted standards. Such a body could also arbitrate any conflicts between different data holders. Some governments have also taken a more active role facilitating the technical implementation of standards, providing on-boarding support, testing tools or collecting performance measures (e.g. the Conformance Test Suit of Australia's CDR initiative) (OAIC, 2020[108]).

Data quality can also become a technical barrier to data access, sharing and re-use. This is because the information that can be extracted from data depends on their quality. Poor-quality data will almost always lead to poor data analysis and results. Therefore, data cleaning is often an important step before data can be analysed. This, in turn, involves significant costs. Data cleaning can account for 50-80% of a data analyst's time together with the actual data collection (Lohr, 2014[109]).

However, data quality may not only affect the ability and the cost to re-use data. It can also prevent stakeholders from participating in data-sharing arrangements. According to some studies, uncertainties about data quality may explain why open data repositories are used at far lower rates than most scholars and practising data curators would expect.[22]

In addition, "many datasets are not of requisite quality, are not adequately documented or organised, or are of insufficient (or no) interest for use by others" (OECD, 2017[110]). The lack of a common understanding about what quality means in the context of data has also been a major source of uncertainty among organisations. Some authors have therefore argued that data quality should be considered a key determinant of trust for data sharing (Wallis et al., 2007[111]; Federer et al., 2015[112]; Sposito, 2017[113]).

4.3. Legal frameworks for trustworthy and effective data access, sharing and re-use

The legal landscape is critical for trust and therefore considered a key enabler of data openness. However, the legal landscape surrounding data openness is also highly complex. It involves multiple overlapping legal and regulatory frameworks, which introduce legal uncertainties when considered in isolation. Indeed, they introduce even more uncertainties regarding their applicability, in particular where multiple jurisdictions are involved.

The positive potential of data re-use to enhance development outcomes could be maximised through new governance mechanisms and structures to guide relevant parties. They could encompass risk assessments, technical protocols to promote data responsibility by design, model data-sharing agreements, ethic review boards, a global governance framework that smooths cross-border data flows and enhanced public engagement (Verhulst, 2021[72]).

4.3.1. An intricate net of legal frameworks

The need for additional legal frameworks is particularly apparent for personal data and their access, sharing and re-use. Privacy and data protection frameworks largely govern the collection, use, access and sharing of personal data. However, additional legal frameworks can also be pertinent. For example, consumer protection regulations often restrict the use of deceptive or unfair commercial practices. This may become relevant, for example, in the context of some non-monetary transactions such as access to "free" services in exchange for consumer data. It could also be pertinent with respect to a deceptive or misleading representation of privacy choices (OECD, 2019[33]). Furthermore, IPRs, in particular copyright and trade secrets, can be applicable under certain conditions. In certain jurisdictions, cyber-criminal law may effectively confer data control rights to data holders. Meanwhile, competition law can regulate questions related to data access emerging between firms (Clarke, 2016[114]).

In terms of open data, instruments such as Freedom of Information and Transparency laws are commonly used as the main legal foundations to support open government data efforts. Nevertheless, some OECD countries have gone further to align with the digitalisation of the economy and society. Relevant examples include Korea's Act on the Promotion, Provision and Use of Public Data (2013), France's Digital Republic Law (Loi pour une République Numerique) (2016), Germany's Law for the Promotion of Electronic Government (E-Government Law) (2017) and the United States' Foundations for Evidence-based Policymaking Act (2018) (OECD, 2018[24]).

On top of all of this, some sector-specific regulatory frameworks may also apply, including open data provisions, when public sector data are concerned. In 2018, Australia introduced complementary legal reforms that entail the Consumer Data Right giving Australians better control of their own data. To improve accountability around data management, the country appointed a National Data Commissioner. Meanwhile, a dedicated Council advises on international best practices, ethics and industry developments. Finally, the Data Sharing and Release Act fosters usability and re-use of data, as well as privacy protection of sensitive data (OECD, 2020[43]).

As a consequence, data will often be governed by an "intricate net of existing legal frameworks" (Determann, 2018[82]) that reflects these various laws and regulations related to data governance. These legal and regulatory frameworks might apply differently to stakeholders, within and across countries. This can be challenging in many low- and middle-income countries that adopted frameworks developed with advanced economies in mind. Such frameworks are unsuitable for their contexts and implementation capacity (OECD, 2021[115]).

The overlapping legal and regulatory frameworks combined with the involvement of multiple parties in the creation of data (and their value) can explain legal uncertainties related to data governance. In identifying the main challenges to transborder data flows, for instance, countries responding to the OECD 2019 Privacy Guidelines Questionnaire most often noted "uncertainty regarding legal privacy regimes", followed by "incompatibility of legal regimes." Other popular responses were time and resources required to enable transborder data flows, and recent trends in favour of data localisation.

4.4. Effective incentives to encourage responsible data access, sharing and re-use

The marginal costs of transmitting, copying and processing data are typically close to zero. However, substantial investments are often required to collect data and to enable data sharing and re-use. These include addressing other barriers (legal, skills and technical) and mitigating risks (through development and adoption of digital technology solutions to protect and manage data). Given these significant investments, data holders may not necessarily have the incentive to share their data. This is especially the case if the costs (risks) of data access and sharing are perceived to be higher than the expected (private) benefits.

A similar argument can be made for individuals: they are more likely to share their personal data if they expect benefits from it. In other words, organisations and individuals need to recuperate a sufficient level of the return on their data-related investments. This could take the form of revenues arising from granting data access against licence fees or added-value services. Otherwise, data sharing may not occur at a sufficient level across society.

4.4.1. The evolution of incentives to promote electronic health records

The health sector's efforts to improve the quality and exchange of clinical data captured in electronic health records (EHRs) provides an interesting case study of the role of incentives. A decade or more ago, many countries took one of several approaches to provide incentives around EHRs. Some introduced legislation

requiring that health care providers adopt EHRs. Others did the same, and also offered financial incentives to adopt electronic record keeping and participate in information exchange. Some only offered the financial incentives (OECD, 2013[116]). These early efforts did encourage electronic record keeping but were inadequate regarding the coverage and quality of the data exchanged.

In 2021, many countries had a more sophisticated approach to incentivising "verifiable interoperability." This included multidisciplinary governance supporting development of national data standards that meet the needs of different stakeholders in the health information system (14 countries); laws requiring that health care providers' EHRs meet national standards for data interoperability (18 countries); certification of EHR software vendors to ensure their products meet national interoperability standards (13 countries); financial incentives or penalties to health care providers to meet national requirements for EHR interoperability (11 countries); and auditing the EHRs of health care providers for the quality of the data (13 countries) (OECD, 2022[35]).

Overall, the countries where EHR data are already contributing to national statistics and health research employ most of these policy levers to succeed. Further, countries seeking to introduce incentives to improve data interoperability must first identify disincentives that may exist from legacy legislations; siloed approaches to funding data and research activities; and public mistrust of data governance resulting from prior data breaches or misuses (OECD, 2022[4]; OECD, 2022[117]).

4.4.2. The need for coherent incentives for sharing science and research data

The need for recognition and reward systems for data authors has also been recognised as key for incentivising access to publicly funded data for science, technology and innovation. OECD (2020[43]) notes that "data sharing entails cultural change among researchers in many scientific fields. Appropriate acknowledgement and reward systems need to counterbalance the perceived barriers and risks of enhancing access to data." The report further highlights that

> (r)esearchers have incentives to publish (preferably positive) scientific results. Incentives to publish data are less developed, and usually seen as a constraint imposed by funding agencies and/or publishers. Data citation has not been widely implemented. Although the prerequisites for achieving this (e.g. standard formats and citation metrics) already exist, they are not being broadly adopted.

4.5. Recognising and addressing skill and capacity gaps and the data divide

Lack of sufficient data-related skills and competences and poor access to computation and data storage capacities can become bottlenecks. They may thus prevent the effective re-use of data, even where data are made available through access and sharing. Insufficient investments in skills and competences and information and communication technologies (ICTs) therefore can constitute a major barrier to reaping the benefits of data openness.

Better data-related skills are needed along the whole value cycle of data to ensure the effective sharing and (re-)use of data – from data collection and storage management to analysis and sharing. Data holders therefore need data management and data curation capabilities to ensure the long-term quality and availability of data. Data users, on the other hand, need adequate digital and data analytics skills to re-use data effectively. Evidence also shows that skills and competences can improve awareness of the actual risks of data access and sharing. Lack of data-related skills is therefore an issue across all sectors. It may prevent the effective re-use of data, even if made available via open access (OECD, 2019[3]). Available evidence from open government data initiatives, for instance, shows that "open data literacy programmes" are essential to engage all stakeholders.

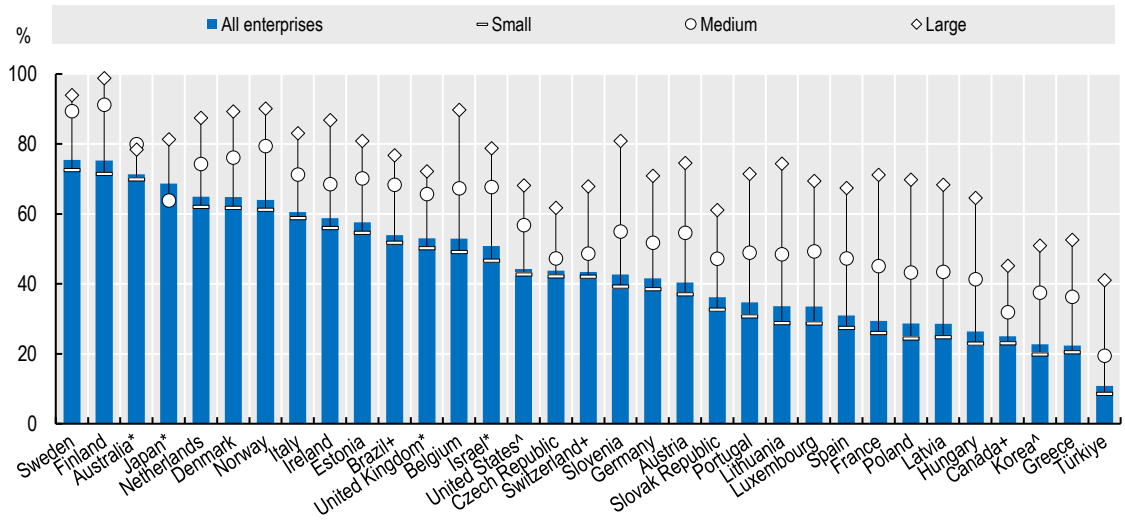
4.5.1. Government initiatives to promote access to data-related infrastructures

Access to digital infrastructure is a major condition for the effective re-use of data across society. Therefore governments may also need to support development of the infrastructure needed for data storage, processing and analytic. This is particularly critical for small and medium-sized enterprises (SMEs). However, it applies equally to individuals, including scientists, as it involves the cost of operating, maintaining and scaling data infrastructure.

The diffusion of cloud computing has been a major catalyst for the re-use of data and big data in particular (OECD, 2019[3]). However, the adoption of cloud computing by firms remains much below expectations. In 2021, the share of businesses that used cloud computing services ranged from over 75% in Sweden and Finland down to 11% in Türkiye. In most countries, uptake is higher among large businesses (close to 70%) compared to SMEs, which record around 40% and 60%, respectively (Figure 4.1).

Figure 4.1. Enterprises using cloud computing services, by firm size, 2021

As a percentage of enterprises in each employment size class



Note: Data refer to manufacturing and non-financial market services enterprises with ten or more persons employed, unless otherwise stated. Size classes are defined as: small (10-49 persons employed), medium (50-249) and large (250 and more).
* Data refer to 2020 for Australia, Israel, Japan and the United Kingdom. For Australia, data refer to the fiscal year 2019/20 ending 30 June. For Japan, data refer to businesses with 100 or more employees. Medium-sized enterprises have 100-299 employees. Large enterprises have 300 or more employees.
+ Data refer to 2019 for Brazil, Canada and Switzerland. For Canada, data refer to enterprises that have made expenditures on software as a service. Medium-sized enterprises have 50-299 employees. Large enterprises have 300 or more employees. For Switzerland, data refer to firms with five or more employees.
^ Data refers to 2018 for Korea and the United States.
Source: OECD, ICT Access and Use Database (businesses), http://oe.cd/bus (last updated in June 2021).

4.5.2. Government initiatives to strengthen data-related skills

Governments have recognised that availability of data-related skills and competences can be a critical bottleneck for the effective re-use as well as provision of data in both the private and public sectors. Some have established dedicated initiatives to support development of data-related skills and infrastructures. The United Kingdom has a number of initiatives to support skill development in the private and public sectors. The Digital Skills Partnership, for instance, brings together public, private and charity sector

organisations to boost skills for a world-leading, inclusive digital economy. Other complementary initiatives are the UK Government Data Ethics Framework and the Centre for Data Ethics and Innovation. These aim to ensure that public servants from across disciplines understand insights from data and emerging technologies and use data-informed insight responsibly (Department for Digital, Culture, Media & Sport, 2018[118]).

4.5.3. Closing skill and capacity gaps in developing countries

Skill and capacity gaps have been challenging for developing countries. Despite advances in recent years, these countries remain disadvantaged in terms of access to connectivity. Just 57% of individuals in developing countries used the Internet in 2021 versus 90% in developed countries (Amadou-Garba, 2021[119]). In addition, computation capacity and data storage facilities are only now being developed (OECD, 2021[120]).

Closing basic connectivity gaps, particularly in Africa, remains crucial. Addressing computational and in particular data storage and connectivity capacity gaps, however, has been viewed within the context of ongoing debates around "data sovereignty" (Susan Ariel Aaronson, 2021[121]). The latter could result in the rise of data localisation measures. As a consequence, it could hamper cross-border flows, undermine interoperability and data portability initiatives, and put local businesses at a competitive disadvantage (OECD, 2022[88]).

Developing countries also lag in data-related skills such as AI. This hampers development of local solutions to unlock new economic opportunities. Skills investments must be carefully tailored to context to match labour market needs, education levels and likely future opportunities in low- and middle-income countries (Broecke, 2021[122]). As noted elsewhere, regulatory skills and capacities are also being challenged in low- and middle-income countries and bespoke support is required.

4.6. Towards a culture of responsible data access, sharing and re-use

Legislation, technologies and applications need to be matched by a culture of responsible data stewardship in the data ecosystem that supports a presumption of the responsible sharing and use of data. Raising capacity as discussed above is a necessary, but not sufficient, condition for establishing such a culture. Active stakeholder participation through which the views, values and interests of all relevant stakeholders can be reflected is an important element of such culture.

4.6.1. Engaging stakeholders to build trust

Evidence shows that engaging with communities of stakeholders to understand their respective interests and concerns is a major success factor for building trust. Trust, in turn, aligns with ethical values and norms such as fairness, human dignity, autonomy and self-determination. In addition, active community engagement can help better allocate responsibilities and define acceptable risk levels. In science, for example, an inclusive and cross-organisational stakeholder consultation process involving communities, data infrastructures and funders has shown to be critical to establishing a common culture of open data based on trust. In the Netherlands, the National Plan for Open Science interconnects open access, open data and reward systems in the wider framework of open science (OECD, 2020[43]).

In the public sector, legislation, IT platforms and applications also need to be matched by a culture within the public service that supports a presumption to publish, release and share data (OECD, 2019[107]; OECD, 2015[123]). Raising capacity relevant to open government data and awareness of civil servants, citizens, civil society organisations and the private sector with regard to their rights is important for society as a whole to fully capture the benefits of public sector data.

Government departments, in partnership with civil society groups, can create awareness of legislation and policies that empower citizens around data, such as the Access to Information or Freedom of Information Acts. Additionally, undertaking research to establish users' information needs and barriers to information use and re-use, or seeking public-private partnerships is also relevant. This can encourage data use to foster innovation, lead to ventures for the worthwhile re-use and redistribution of and universal participation in open government data. Examples include development of applications and provision of e-government services.

4.6.2. Risk-based approaches gain prominence

A risk-based approach can be seen as another important element of a culture of responsible data access, sharing and re-use. Such an approach requires a cultural shift away from exclusive protection of an asset or an environment from threats to the optimisation of benefits. In this way, it recognises that risk is not a binary concept; a certain level of risk has to be accepted. Indeed, there is always some risk with carrying out data-related activities, including data access, sharing and re-use (OECD, 2015[124]).

A risk-based approach offers additional degrees of freedom. It allows adjustment of the restrictedness of control and protection measures to the acceptable level of risk based on stakeholder engagement and participation. Nevertheless, a risk-based approach remains challenging to implement for organisations and policy makers. This is especially the case where the rights of third parties are involved, such as privacy rights of individuals and the IPRs of organisations and individuals (OECD, 2019[3]).

Risk management requires setting the acceptable level of risk, and treating the risk accordingly on the basis of a full risk assessment. Complex questions remain in the context of privacy protection, such as how to allocate responsibility and how to define the acceptable level of risk. Further, the risk management framework requires establishing a full and ongoing risk management cycle. In such a cycle, awareness, skills, responsibility and co-operation play key roles. Moreover, risk assessment and treatment are continuous to consider the dynamic nature of activities and the environment. Finally, risks to the organisation and the individual need to be separated. Third-party accreditation may be useful in some situations to validate internal processes for risk-based approaches.

5 Data strategies for trustworthy and effective stewardship and control

Governments play a pivotal role in developing the environment within which societies can become digitalised and data can be more open, while maintaining public trust and protecting individuals' privacy rights. This section highlights the potential of strategic whole-of-government approaches that support a dynamic and innovative social and economic environment for data openness. It reviews the concept of data control and stewardship via the various technical, organisational, legal measures to enhance control and stewardship. Subsequently, it addresses how a strategic approach to data control and stewardship can help maximise the benefits of data openness, while managing the associated risks. The section benefits from policy examples on national data strategies based on results of recent OECD surveys on data governance.

5.1. Addressing the tension between data opennes, public trust and privacy

Governments play a pivotal role in developing the environment within which societies can become digitalised and data can be more open, while maintaining public trust and protecting individuals' privacy rights. Open data generate significant economic and social benefits with unjustified barriers creating major opportunity costs. Yet the rights and interests of stakeholders must also be protected. Given these tensions, more balanced and differentiated approaches to data stewardship and control are needed to

maximise the benefits of data, while protecting the rights and other legitimate interests of both individuals and organisations.

The COVID-19 pandemic underlined the pivotal role of governments for creating such an environment. For instance, the pandemic raised awareness of the importance of a mature health data infrastructure and governance for enhancing health care and research on COVID-19 and for greater resilience during future public health emergencies. This is important because most OECD countries are still implementing national health data governance frameworks. Indeed, many low- and middle-income countries do not have mature infrastructure and governance models (OECD, 2022[35]; Barker, 2020[125]). In a 2021 OECD survey, 15 of 24 countries indicated making legal, regulatory or policy reforms in 2020 and 2021 to improve health data availability, accessibility or sharing. Meanwhile, 9 of 24 countries made reforms to improve data privacy or security protections (de Bienassis, 2022[38]).

Data availability, information needs, analytical techniques and data protection and privacy, as well as digital security risks, change over time. Therefore, improvements to data governance frameworks in response to the COVID-19 pandemic also require ongoing support and adaptation. Building integrated health information systems to meet the needs of the digital age and support governments in times of crisis will require sustained commitment to a national (health) data governance framework (OECD, 2022[4]; OECD, forthcoming[126]). Encouragingly, countries reported in 2021 that almost all changes and improvements to health data and governance as a result of the pandemic were likely to be kept in the long term (de Bienassis, 2022[38]).

5.2. Enhancing data access, sharing and re-use with trust

As highlighted above, a common cause of challenges discussed in Section 3 is the loss of control over data, which is rooted in the partial excludability[23] of data. Technological, organisational and legal means for re-establishing control over data and information can thus help address the above challenge. However, they must avoid creating unjustified barriers as discussed in Section 4. Technological measures would include privacy-enhancing technologies (PETs), application programming interfaces (APIs), as well as data sandboxes; legal measures such as contractual agreements; and organisational measures such as trusted third parties, including data fiduciaries and trusts, and participatory data stewardship (OECD, 2021[78]). The latter also includes lawful alternatives to consent and the data governance mechanisms needed to earn public trust in sharing data through legislation.

5.2.1. Technological measures

Technological measures include, most prominently, PETs and other related technologies to enhance confidentiality of data. According to OECD (2002[127]), “privacy enhancing technologies (PETs) commonly refer to a wide range of technologies that help protect personal privacy. Ranging from tools that provide anonymity to those that allow a user to choose if, when and under what circumstances personal information is disclosed, the use of privacy enhancing technologies helps users make informed choices about privacy protection.” They are typically not stand-alone tools but can be viewed as new functionalities that can be added to data governance frameworks used by organisations.

OECD (forthcoming[126]) differentiates between the following three classes of PETs:

- Data accountability tools provide enhanced control to the sources over how data can be gathered and used, or provide greater transparency and immutability into related transactions. These include accountable software systems that manage the use and sharing of data by controlling and tracking how data can be collected, how they are processed and when they can be used. As a key goal of accountable system design, data access can be granted with limitations attached to the data. In other words, the sphere of control follows the data.

- Data obfuscation tools, like encrypted data processing tools (see bullet below), reduce the need for sensitive information to leave a data source's sphere of control, where the underlying data are kept and processed. Unlike encrypted data processing tools, however, obfuscation tools require access to the underlying (unencrypted) raw data that need to be either processed locally on the data subject's device, or altered by adding "noise" or by removing identifying details. Data obfuscation tools commonly include anonymisation and pseudonymisation techniques.[24] It also include federated learning, a technique gaining increased attention, where raw data are pre-processed at the data source. In this way, only the summary statistics/results are transferred to those executing the tasks.[25]
- Encrypted data processing tools allow running computations over encrypted data that are never disclosed. In contrast to data obfuscation tools, the underlying data remain intact but hidden by encryption. The most prominent example includes homomorphic encryption. In this technique, a data processor can perform simple (but increasingly complex) calculations over the encrypted data. It extracts an encrypted result that can only be unlocked with the original data source's cryptographic key.

The health sector has expanded use of federated learning solutions for secure and privacy-protective access to health data for both public- and private-sector research. These include the EU Health Data and Evidence Network project, the European Medicines Agency Darwin project and the global Observational and Health Data Sciences and Informatics project (OECD, 2022). In a federated learning model, software and statistical analysis programmes travel to where data are located, rather than data flowing to a central data lake for analysis. Further, these approaches do not permit researchers to visualise or directly access data. Such approaches succeed because all data to be used by researchers are first coded to a common data model, such as the Observational Medical Outcomes Partnership model. Federated learning models are a solution in countries with limitations on data sharing or with data localisation policies because the data always remain with their custodians and under national data protection laws.

Another important technical measure includes the use of APIs. As applications increasingly rely on data without human intervention, APIs are underpinning most data transfers. APIs essentially enable service providers to make their digital resources (e.g. data and software) available over the Internet. This enables the smooth interoperability of the different actors, their technologies and services, particularly through cloud computing.[26] As a key advantage, an API enables a software application (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 ensure 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 or personal information.

5.2.2. Organisational measures

Organisational measures refer to institutional arrangements that may involve contracts – often in combination with technical measures – to help manage control over data. Data sandboxes, for example, are isolated environments, through which data are accessed *and* analysed. Analytic results are only exported, if at all, when they are non-sensitive. Data sandboxes typically require execution of the analytical code at the same physical location as the data. These sandboxes can be realised through technical means of various complexity. Isolated virtual machines, for example, cannot be connected to an external network to federated learning solutions. However, they can also have a physical on-site presence within the facilities of the data holder (where the data are located).

Data intermediaries can also help manage control over data. The OECD (2021[128]) Recommendation on Enhancing Access to and Sharing of Data defines data intermediaries as "service providers that facilitate data access and sharing under commercial or non-commercial agreements between data holders, data producers, and/or users." As the Recommendation notes, the "intermediaries" can in fact be both the data

holders themselves (e.g. where they come together to an arrangement to facilitate data sharing, such as through data spaces), as well as trusted or certified third parties.

In *data trusts,* for example, a trusted third party (an informed person or organisation) takes on a fiduciary duty to steward/govern data use or sharing on behalf of its members in relation to third parties. This aims to increase access and sharing of the data while safeguarding the rights and interests of the data holders (Hardinges, 10 July 2018[129]; Ruhaak, 2021[130]). This arrangement can also include personal data stores or Personal Information Management Systems (PIMS). PIMS are service providers that use data accountability tools to enhance individual's control over their personal data. Individuals can then choose where and how they want their data stored, accessed or processed. This enables them to manage their data at a more granular level (Royal Society, 2021).

In *trusted data-sharing platforms*, major data holders come together and either designate an existing trusted organisation or create a new trusted organisation and platform to share data with third parties. The Health Care Cost Institute (HCCI), for example, is a non-profit organisation designated by health care and health insurance companies in the United States (e.g. Aetna, Humana, Kaiser Permanente and United Healthcare) to share information about health care use and costs in the United States with selected research institutions. HCCI removes information about which company has provided the data before sharing them (OECD, 2019[3]).

Based on their literature review, Paprica et al. (2020[131]) suggest that institutions should comply with the following minimal requirements to be able to operate as data trusts:

- Legal: The data trust must fulfil all legal requirements, including the authority to collect, share and hold data.
- Governance: The data trust must have a stated purpose, be transparent in its activities, have an accountable governing body and be adaptive.
- Management: There must be well-defined policies and processes for the collection, storage, use and disclosure of data, which must include data protection safeguards, be reviewed and updated regularly, and be complemented by an ongoing process to identify, assess and manage risks.
- Data user requirements: All data users must complete training before they access data, and must agree to a data user agreement that acknowledges that data use will be monitored and includes consequences for non-compliance.
- Public and stakeholder engagement: There must be early and ongoing engagement with stakeholders including members of the public, including direct engagement tailored for subpopulations or groups, where there is a reasonable expectation that these subpopulations or groups would have a particular interest in, or be affected by, an activity of the data trust.

In some cases, governments can act as, or create, a trusted third party. In certain countries, national statistical offices have acted as such a trusted third party. In Australia, the government invested AUD 131 million over three years to maximise the use and value of its data assets from 2017 to 2020, an initiative known as the Data Integration Partnership for Australia (DIPA) (Australian Government, 2017[132]). Agencies in social services, health, education, finance and other government agencies would provide data for linking and integration. Subsequently, "sectoral hubs of expertise, independent entities that are funded by the Commonwealth" and denominated accredited integrating authorities (AIAs), would enable integration of longitudinal data assets – "housed in a secure environment, using privacy preserving linking methods and best practice statistics to link social policy and business data" (Productivity Commission, 2017[133]).[27] The Australian Bureau of Statistics, Australia's national statistical office, was the first institution to be recognised as an AIA.[28]

5.2.3. Legal measures

In areas where data control is likely to remain governed through consent decisions, applied consent mechanisms should remain user friendly and account for the prevalence of behavioural biases. On the one hand, this involves providing users with sufficient information and options to flexibly grant or revoke permissions to make (real time) use of or share their data. On the other, it must ensure these options and choices remain sufficiently easy to digest for consumers. Mechanisms involving differentiated consent options mixed with *privacy by default* settings may be useful in this regard and seem to be well aligned with consumer preferences (ConPolicy, 2020[96]). However, data control through consent naturally extends to use by third parties, where enforcement of standards often remains challenging.

Data portability has also become an essential tool to grant users better agency and control over "their" data. This empowers them to play a more active role in the re-use of these data across digital services and platforms. Data portability enables the data subject to download the data. However, in some cases, it can also enable transfer of data to third parties. In so doing, it enhances access to and sharing of data across digital services and platforms, while strengthening the control rights of individuals and firms (OECD, 2021[26]). Under certain conditions, data portability helps foster competition between digital services and platforms (OECD, 2021[134]; OECD, 2022[88]). Thus, significant interest in data portability has arisen in the competition policy community.

Data portability initiatives and arrangements differ significantly along five key dimensions (OECD, 2021[26]):

- Sectoral scope, including whether they are sector-specific or horizontal and thus directed potentially at all data holders regardless of the sector.
- Beneficiaries, including whether only natural persons (individuals) or also legal persons (i.e. businesses) have a right to data portability.
- The type of data subject to data portability arrangements, including whether data portability is limited to personal data and whether it includes volunteered, observed or derived data.
- Legal obligations, especially the extent to which data portability is voluntary or mandatory and if the latter, how it is enforced.
- The modalities of data transfer, meaning the extent to which data transfers are limited to or include ad hoc (one-time) downloads of data in machine-readable formats (regarded as "data portability 1.0"), ad hoc direct transfers of data to another data holder ("data portability 2.0") or real-time (continuous) data transfers between data holders that enable interoperability between their digital services ("data portability 3.0"). Another important modality is whether third-party data recipients need to be accredited to participate in data portability arrangements.

Governments have established rules, institutional and regulatory frameworks, standards, guidelines, infrastructure, architecture and processes that support data delivers value with trust. This applies in particular to the handling and processing of personal data by public sector authorities. In this light, as an evolution from standard data protection and privacy regulation in OECD countries, governments are also looking into giving citizens greater visibility and control over how their personal data are shared and used across government.

OECD work on digital government has underlined that "the practical steps to unlock the potential of digital identity should be built on the existing efforts of international organisations and standard-setting bodies to guide the conversation around digital identity in terms of authentication, personal data protection, and security risks" (OECD, 2021[135]). On the Luxembourg government's digital public service portal MyGuichet.lu, for example, citizens can consult personal data recorded about them in government authentic sources, such as the National Register of Natural Persons. They can also see what organisations have accessed the data and ask for their modification in case of error.

The growing discussions about decentralised or hybrid approaches to digital identity is an important development in this context. These approaches focus on helping citizens take more control over the sharing of their personal identity attributes and credentials as they do, for example, in Finland.[29] Also, in 2021, at the European Union level, the European Commission proposed a framework for a European Digital Identity (European Commission, 2021[136]). This would replace the Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market. The European Digital Identity wallets will enable people to choose which aspects of their identity, data and certificates they share with third parties, and to keep track of such sharing. User control ensures that only information that needs to be shared will be shared.

These advanced digital identification systems developed by governments across the world are, as with traditional analogue systems, only possible through registration, access to and sharing of sensitive personal attributes. In some cases, such sensitive data are also held by non-governmental organisations. Humanitarian aid and development organisations, for example, use data to establish refugee registers or identification systems where proofs of legal identity are lacking (Carey and Mc Donnell, 2021[76]). The level of control and visibility of data subjects over these data depend on the type of digital identity system in place. Digital identity systems need to be designed to ensure trust between the different parties, including between providers of digital identity, the source of attributes and the data subject. Approaches that move beyond consent as the primary basis for protecting personal data could help address some of these issues.

5.3. Achieving greater policy coherence through national and sectoral data strategies

Some countries have developed or are developing national data strategies or sectoral data strategies to assure the coherence and flexibility of their national and sectoral data governance frameworks. These strategies could help address many of the policy issues in a comprehensive manner by incorporating a whole-of-government perspective. In particular, sectoral and national data strategies could be instrumental. On the one hand, they could create the conditions for effective data governance frameworks to better protect the rights and interests of individuals and organisations. On the other, they could provide the flexibility needed for all to benefit from data openness.

COVID-19 accelerated demand for support for digital transformation among low- and middle-income countries, focused on regulatory and policy guidance. The United Nations Development Programme, for example, fielded requests from over 100 countries (OECD, 2021[137]). In this context, national data strategies also provide a strategic opportunity to achieve policy coherence.[30] Many developing countries have data strategies relating to government statistics (PARIS21, 2017[138]) and open government data (Open Data for Development Network, n.d.[139]). However, a growing number are now developing digital economy and digital transformation strategies (OECD, 2021[140]). Critical success factors for strategies in low- and middle-income countries are emerging. These include leadership from a central body, clear vision, prioritising voices of the disadvantaged and taking a whole-of–government approach to identifying relevant sectoral areas of focus and long-term goals.

5.3.1. National data strategies

National data strategies are cross-sectoral by nature. In many instances, they are designed explicitly to help reach higher level objectives, such as gross domestic product growth, productivity, well-being and/or combating climate change and fostering sustainable development. To achieve these strategic objectives, some national data strategies build on specific strengths of the country, often relating to pre-existing national strategies. These might include national digital economy strategies, e-government strategies and national AI strategies, which national data strategies often complement.

Approaches to governing national strategies vary across countries, but commonalities can be found. Development, co-ordination and monitoring of sectoral data strategies, for example, fall essentially under the responsibility of the ministry in charge of the sector. For national data strategies, a few countries (though still exceptions so far) have appointed a high-level government official to lead development, co-ordination and monitoring. Several have tasked a ministry, ministerial position or other body dedicated to digital affairs. Not surprisingly, therefore, several ministries, bodies or institutions implement national data strategies in most countries. In some cases, multiple stakeholders are also involved. In almost all countries, multiple private and public stakeholders and bodies contribute input to development of national and sectoral data strategies.

Countries such as the United States, the Netherlands, Germany and the United Kingdom have developed specific national data strategies. These take a holistic approach to data-specific policies such as open data, privacy and data protection, interoperability and business innovation (OECD, 2019[107]). Raising. However, evidence from the *OECD Digital Government Index* shows that "progress towards a comprehensive and dedicated approach that addresses data as a strategic asset seems to be lacking" (OECD, 2020[141]). Only 12% of respondents of the OECD Digital Government Index survey confirmed they had a single dedicated data policy (or strategy) for the central or federal government.

Some government initiatives may not be explicitly referred to as national data strategies, although they could be considered as such. National open data initiatives, such as Australia's data sharing and release legislation, can be considered national data strategies with a focus on public sector data. For its part, the United States adopted its Federal Data Strategy to provide a co-ordinated and integrated approach to using government data to deliver on mission, serve the public and steward resources, while respecting privacy and confidentiality (United States Government, n.d.[142]).

5.3.2. Sectoral data strategies

Some national data strategies are cross-sectoral by nature. However, they may also include sector-specific elements or be complemented by sectoral data strategies that focus on selected sectors such as health, transportation, energy or the public sector.

In the health sector, for example, the COVID-19 pandemic has been a catalyst for development of sector-specific national health data strategies including in Australia, Canada, Finland, France, the United Kingdom and the United States (OECD, 2022[35]). Significantly, the pandemic has also been a catalyst for the first multi-country health data strategy. In May 2022, the European Union announced a proposed regulation and funding to create the European health dataspace (EHDS) (European Commission, 2022[143]) (see Box 5.1). EHDS would become the first common European data space that establishes the rules, common standards and practices, infrastructures and a governance framework for use of electronic health data by patients, health care providers and for research, health care improvement, statistics, innovation and regulation. Other sector-specific dataspaces are expected within the overall EU 2020 data strategy (European Commission, 2020[144]).

Key components of the proposed EHDS include strengthening health data interoperability and exchange within countries and across borders. It would also provide access to health data for secondary uses within a trusted and secure framework (Box 5.1). By fostering secure multi-country data-driven health policy and research, this regulation would have significant influence on non-EU countries as they determine their capacity to participate with EU countries in secure multi-country research and innovation projects. The European Commission has made an economic case for the EHDS, estimating the project will save EUR 11 billion over ten years. Savings are expected from better access to and exchange of health data for patients and health care providers and better use of data for research, innovation and policy making. The planned investment for EHDS includes EUR 12 billion for digital health under the Recovery and Resilience Facility, EUR 810 million from the European Commission, EUR 280 million under the EU4Health Programme and further investments from other programmes (European Commission, 2022[143]).

In the domain of science and research, as another example, many countries have adopted national strategies, governance arrangements, policies and regulations. Countries like Korea, Norway, Finland and Spain have adopted top-down strategies (the Korean Strategy to Promote Sharing and Use of Research Data for Innovative Growth, the Norwegian national strategy on access to and sharing of research data, and the Finnish Open Science and Research Initiative). Others have preferred bottom-up initiatives driven by the institutions, such as the Netherlands National Plan on Open Science and the UK Concordat on Open Research data (OECD, 2020[43]).[31]

Box 5.1. The legislative proposal to create the European Health Data Space

The European Health Data Space (EHDS) would support within-country and cross-border secure access to and use of health data for purposes such as health care and research and innovation within the private and public sectors. Key elements of this framework follow:

Primary data uses

- Patients will have access to their electronic health data, and a cross-border digital infrastructure for primary use will connect EU member states and allow patients to share their health data for primary use.
- EU member states will be required to adhere to a common European electronic health record exchange format for priority data, such as patient summaries, e-prescriptions, e-dispensations, medical images and image reports, laboratory results and discharge reports.
- Health professionals must be given access to electronic health records (EHRs) and update the electronic health data of the patients they treat. Mandatory requirements for interoperability, security, safety and privacy will be introduced, as well as mandatory self-certification of EHR interoperability and security.
- EU member states will be required to set up a digital health authority to ensure rights for individuals are implemented. A transitional period is provided to implement requirements. A pilot project will support patients having access to their data on a mobile device in the language of the country of destination.
- All member states are required to participate in cross-border digital infrastructure for the exchange of health data for health care delivery (MyHealth@EU).

Secondary data uses

- An application process for a permit from a health data access body will set out how the data may be used and for what purpose.
- Data access would only be available through a closed secure environment provided by health data access bodies that conforms to standards for cybersecurity.
- Only anonymous data can be extracted by the user who applied for the permit from the secure processing environment and, where needed, only pseudonymised microdata may be analysed.
- Data users are prohibited from re-identifying data subjects or using data to take decisions that are detrimental to individuals.
- Transparency through public information is required for data access applications, and data users are required to make their research results public and inform the health data access body about their findings.
- Researchers from non-EU countries can access data for secondary use under the same conditions and requirements as EU member states.
- EU member states will be required to participate in the EU-infrastructure for secondary use (HealthData@EU) to facilitate cross-border studies. This infrastructure will begin to be piloted in 2022.

Source: (European Commission, 2022[143]).

References

360Giving (n.d.), *360Giving*, website, https://www.threesixtygiving.org/ (accessed on 7 October 2022). [65]

ACCC (2019), *Digital Platforms Inquiry – Final Report*, Australian Competition and Consumer Commission, Canberra, https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report. [95]

Acquisti, A., L. Brandimarte and G. Loewenstein (2015), "Privacy and human behavior in the age of information", *Science*, Vol. 347/6221, pp. 509-514, https://doi.org/10.1126/science.aaa1465. [91]

Agency for Data Supply and Infrastructure (n.d.), "Basic Public Data", webpage, https://eng.sdfi.dk/data-creates-value/basic-public-data (accessed on 2 September 2022). [55]

Amadou-Garba, A. (2021), "Practical solutions to connect the last mile", in *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/3f8054bf-en. [119]

AMIS (n.d.), "Agricultural Market Information System: About", webpage, http://www.amis-outlook.org/amis-about/en/ (accessed on 7 October 2022). [15]

Australian Government (2017), *Information about the Data Integration Partnership for Australia*, Department of the Prime Minister and Cabinet, Australian Government, http://www.pmc.gov.au/sites/default/files/publications/DIPA-information.pdf. [132]

Australian Government (n.d.), "Data Integration – Accredited Integrating Authorities", webpage, https://toolkit.data.gov.au/Data_Integration_-_Accredited_Integrating_Authorities (accessed on 7 October 2022). [147]

Australian Government - Treasury (2022), "Consumer Data Right Sectoral Assessment: An explainer for anyone engaging in a Consumer Data Right (CDR) consultation process for the first time", (fact sheet), Australian Government – Treasury. [148]

Barker, A. (2020), "Consumer data and competition: A new balancing act for online markets?", *Going Digital Toolkit Policy Note*, No. 5, OECD, Paris, https://goingdigital.oecd.org/data/notes/No5_ToolkitNote_ConsumerData.pdf. [125]

Benjamins, R., J. Vos and S. Verhulst (2022), "Mobile big data in the fight against COVID-19", *Data & Policy*, Cambridge University Press, https://www.cambridge.org/core/journals/data-and-policy/article/mobile-big-data-in-the-fight-against-covid19/A6DF0E9C2FB55E1E11DF62AD36FFB123. [70]

Black, K. (2021), "Cheers to heightened health (privacy) in 2021", *National Law Review*, Vol. 19 January, https://www.natlawreview.com/article/cheers-to-heightened-health-privacy-2021. [41]

Bodkin-Andrews, G. et al. (2019), *Delivering Indigenous Data Sovereignty*, 2 July, AIATSIS, https://aiatsis.gov.au/publication/116530. [101]

Bowers, A. (2021), "Early warning systems and indicators of dropping out of upper secondary school: the emerging role of digital technologies", in *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, OECD Publishing, Paris, https://dx.doi.org/10.1787/c8e57e15-en. [48]

Broecke, S. (2021), "Case study: Planning for the future of work", in *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ff4b8bd3-en. [122]

Bunge, J. (2014), "Agricultural firms, farm groups strike deal on crop data", 13 November, The Wall Street Journal, https://www.wsj.com/articles/agricultural-firms-farm-groups-strike-deal-on-crop-data-1415854870. [21]

Carey, E. and I. Mc Donnell (2021), "Overview: Powering an inclusive digital future", in *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/9daec555-en. [76]

Clarke, O. (2016), *Legal Study on Ownership and Access to Data*, European Commission, Brussels, https://data.europa.eu/doi/10.2759/299944. [114]

ConPolicy (2020), *Innovatives Datenschutz-Einwilligungsmanagement – Abschlussbericht*, [Unofficial English Translation], ConPolicy – Institut für Verbraucherpolitik, https://www.bmj.de/SharedDocs/Downloads/DE/Service/Fachpublikationen/090620_Datenschutz_Einwilligung.pdf?__blob=publicationFile&v=3. [96]

CoST Infrastructure Transparency Initiative (2017), *CoST Infrastructure Data Standard*, CoST Infrastructure Transparency Initiative, https://infrastructuretransparency.org/resource/cost-infrastructure-data-standard/ (accessed on 7 October 2022). [63]

Cruz, G. and M. Tiel Groenestege (2021), "Tackling digital disadvantage with people-centred policies", in *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/9ce8f762-en. [102]

de Bienassis, K. (2022), "Health data and governance developments in the wake of COVID-19: How OECD countries are adapting health data systems for the new normal", *OECD Health Working Papers*, No. 138, OECD Publishing, Paris, https://doi.org/10.1787/aec7c409-en. [38]

Degeling, M. et al. (2019), "We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy", *Proceedings 2019 Network and Distributed System Security Symposium*, https://doi.org/10.14722/ndss.2019.23378. [94]

Deloitte (2017), *Assessing the Value of TfL's Open Data and Digital Partnerships*, Deloitte, http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf (accessed on 2 March 2018). [25]

Department for Digital, Culture, Media & Sport (2018), "Guidance Data Ethics Framework", webpage, http://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework (accessed on 1 October 2018). [118]

Determann, L. (2018), "No one owns data", *UC Hastings Research Paper*, Vol. 265, https://doi.org/10.2139/ssrn.3123957. [82]

European Commission (2022), "Communication from the Commission to the European Parliament and the Council – A European health data space: Harnessing the power of health data for people, patients and innovation", No. COM/2022/196 final, European Commission, Brussels, https://ec.europa.eu/health/publications/. [143]

European Commission (2021), "Commission proposes a trusted and secure digital Identity", 3 June, Press Release, European Commission, Brussels, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2663. [136]

European Commission (2020), "European Data Strategy", webpage, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en (accessed on 8 November 2022). [144]

European Commission (2020), "Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act"*, EUR-Lex*, European Commission, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 (accessed on 17 May 2022). [61]

European Commission (n.d.), "The European Interoperability Framework in Detail", webpage, https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail (accessed on 7 October 2022). [158]

European Union (2019), *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information*, European Union, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024 (accessed on 18 October 2019). [60]

FAO (2021), *The State of Food Security and Nutrition in the World 2021*, Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/3/cb4474en/online/cb4474en.html. [5]

FAO (1996), *Report of the World Food Summit*, Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/3/w3548e/w3548e00.htm. [146]

Federer, L. et al. (2015), "Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff", *PLOS One*, https://doi.org/10.1371/journal.pone.0129506. [112]

FNIG (n.d.), *The First Nations Information Governance Centre*, website, https://fnigc.ca/ (accessed on 7 October 2022). [99]

Frischmann, B. (2012), *Infrastructure: The Social Value of Shared Resources*, Oxford University Press, Oxford. [1]

FTC (2021), "Developer of popular women's fertility-tracking app settles FTC allegations that It misled consumers about the disclosure of their health data", 13 January, Press Release, Federal Trade Commission, Washington, DC, https://www.ftc.gov/news-events/news/press-releases/2021/01/developer-popular-womens-fertility-tracking-app-settles-ftc-allegations-it-misled-consumers-about (accessed on 27 April 2022). [80]

Goedde, L. et al. (2020), "Agriculture's connected future: How technology can yield new growth", 9 October, McKinsey, http://www.mckinsey.com/industries/agriculture/our-insights/agricultures-connected-future-how-technology-can-yield-new-growth. [13]

Granville, K. (2018), "Facebook and Cambridge Analytica: What you need to know as fallout widens", 19 March, The New York Times, https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html. [149]

Hardinges, J. (10 July 2018), "A data trust provides independent, fiduciary stewardship of data", ODI blog, https://theodi.org/article/what-is-a-data-trust/. [129]

Henze, M. et al. (2013), "Maintaining user control while storing and processing sensor data in the cloud", *International Journal of Grid and High Performance Computing*, Vol. 5/4, pp. 97-112, https://doi.org/10.4018/ijghpc.2013100107. [87]

HSS (2020), "HHS proposes modifications to the HIPAA privacy rule to empower patients, improve coordinated care, and reduce regulatory burdens", (fact sheet), 10 December, Department of Health and Human Services, Office for Civil Rights, Washington, DC, http://www.hhs.gov/about/news/2020/12/10/hhs-proposes-modifications-hipaa-privacy-rule-empower-patients-improve-coordinated-care-reduce-regulatory-burdens.html. [39]

HSS (2020), *Proposed Modifications to the HIPAA Privacy Rule to Support, and Remove Barriers to, Coordinated Care and Individual Engagement*, 45 CFR Parts 160 and 164, Department of Health and Human Services, Office for Civil Rights, Washington, DC, http://www.hhs.gov/sites/default/files/hhs-ocr-hipaa-nprm.pdf. [40]

IEA (2019), *Energy efficiency and digitalisation*, IEA, Paris, https://www.iea.org/articles/energy-efficiency-and-digitalisation. [31]

Jouanjean, M. et al. (2020), "Issues around data governance in the digital transformation of agriculture: The farmers' perspective", *OECD Food, Agriculture and Fisheries Papers*, No. 146, OECD Publishing, Paris, https://dx.doi.org/10.1787/53ecf2ab-en. [23]

Lohr, S. (2014), "For big-data scientists, 'janitor work' is key hurdle to insights", 18 August, New York Times, http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html. [109]

Māori Data Sovereignty Network (n.d.), *Te Mana Raraunga*, website, https://www.temanararaunga.maori.nz/ (accessed on 7 October 2022). [100]

McFadden, J. et al. (2022), "The digitalisation of agriculture: A literature review and emerging policy issues", *OECD Food, Agriculture and Fisheries Papers*, No. 176, OECD Publishing, Paris, https://doi.org/10.1787/285cc27d-en. [7]

Merriam-Webster (n.d.), "Infrastructure", *Merriam-Webster.com Dictionary*, webpage, https://www.merriam-webster.com/dictionary/infrastructure (accessed on 10 May 2020). [150]

Ministry of Finance [Finland] (n.d.), "Digital Identity Development Project – Valtiovarainministeriö", webpage, https://vm.fi/en/digital-identity (accessed on 7 October 2022). [151]

National Research Council (1987), *Infrastructure for the 21st Century: Framework for a Research Agenda*, National Research Council, Committee on Infrastructure Innovation, Washington, DC, https://doi.org/10.17226/798. [152]

NIFO (2022), "Glossary - Base Registries", webpage, https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/base-registries#:~:text=A%20base%20registry%20is%20a,updating%20and%20preservation%20of%20information. (accessed on 4 November 2022). [52]

OAIC (2020), "ACCC/OAIC Compliance and Enforcement Policy for the Consumer Data Right", webpage, https://www.oaic.gov.au/consumer-data-right/compliance-and-enforcement-policy (accessed on 4 November 2022). [108]

Observatory of Public Sector Innovation (n.d.), "Coronavirus Tracker in the United Kingdom", webpage, https://oecd-opsi.org/covid-response/coronavirus-tracker-in-the-united-kingdom/ (accessed on 7 October 2022). [153]

Observatory of Public Sector Innovation (n.d.), "Data Visualisation Dashboard with Open Government Data about COVID-19 in the Czech Republic", webpage, https://oecd-opsi.org/covid-response/data-visualisation-dashboard-with-open-government-data-about-covid-19-in-the-czech-republic/ (accessed on 7 October 2022). [154]

Observatory of Public Sector Innovation (n.d.), "Data Visualisation Dashboard with Real-time Tracking of COVID-19 Cases in Lithuania", webpage, https://oecd-opsi.org/covid-response/data-visualisation-dashboard-with-real-time-tracking-of-covid-19-cases-in-lithuania/ (accessed on 7 October 2022). [155]

Observatory of Public Sector Innovation (n.d.), "Release of Open API Dataset for the Availability of Public Masks at Designated Stores", webpage, https://oecd-opsi.org/covid-response/release-of-open-api-dataset-for-the-availability-of-public-masks-at-designated-stores/ (accessed on 7 October 2022). [157]

OECD (2022), "Dark commercial patterns", *OECD Digital Economy Papers*, No. 336, OECD Publishing, Paris, https://doi.org/10.1787/44f5e846-en. [97]

OECD (2022), *Draft Principles of Good Practice for Public Communication Responses to Mis- and Disinformation*, https://www.oecd.org/governance/reinforcing-democracy/. (accessed on 17 May 2022). [67]

OECD (2022), *Expert Workshop on Data Ethics: Balancing Ethical and Innovative Uses of Data: Summary of Main Points*, OECD, Paris. [89]

OECD (2022), "Fostering cross-border data flows with trust", *OECD Digital Economy Papers*, No. 343, OECD Publishing, Paris, https://doi.org/10.1787/139b32ad-en. [88]

OECD (2022), *Going Digital Guide to Data Governance Policy Making*, OECD Publishing, Paris, https://doi.org/10.1787/40d53904-en. [84]

OECD (2022), *Health Data Governance for the Digital Age: Implementing the OECD Recommendation on Health Data Governance*, OECD Publishing, Paris, https://doi.org/10.1787/68b60796-en. [35]

OECD (2022), *OECD Economic Outlook, Volume 2022 Issue 1*, OECD Publishing, Paris, http://10.1787/62d0ca31-en. [8]

OECD (2022), "Supporting Health Innovation with Fair Information Practice Principles", *Key issues emerging from the OECD-Israel Workshop, 19-20 January*, OECD Publishing, Paris, https://www.oecd.org/health/OECD-Israel-Health-Data-Governance-Workshop-Report.pdf. [37]

OECD (2022), "The impacts and policy implications of Russia's aggression against Ukraine on agricultural markets", *OECD Policy Responses on the Impacts of the War in Ukraine*, https://www.oecd.org/ukraine-hub/policy-responses/the-impacts-and-policy-implications-of-russia-s-aggression-against-ukraine-on-agricultural-markets-0030a4cd/. [10]

OECD (2022), *Towards an Integrated Health Information System in Korea*, OECD Publishing, Paris, https://doi.org/10.1787/c4e6c88d-en. [117]

OECD (2022), *Towards and Integrated Health Information System in the Netherlands*, OECD Publishing, Paris, https://doi.org/10.1787/a1568975-en. [4]

OECD (2021), "Data portability, interoperability and digital platform competition", *OECD Competition Committee Discussion Paper*, OECD, Paris, https://www.oecd.org/daf/competition/data-portability-interoperability-and-digital-platform-competition-2021.pdf. [134]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [115]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [77]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [71]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [73]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [75]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [140]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [79]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [120]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [78]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [74]

OECD (2021), *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://doi.org/10.1787/ce08832f-en. [137]

OECD (2021), *G20 Collection of Digital Identity Practices: Report for the G20 Digital Economy Taskforce*, OECD Publishing, Paris, https://assets.innovazione.gov.it/1628073752-g20detfoecddigitalid.pdf. [135]

OECD (2021), *Good Practice Principles for Data Ethics in the Public Sector*, OECD, Paris, http://www.oecd.org/gov/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.pdf (accessed on 14 April 2021). [51]

OECD (2021), "Mapping data portability initiatives, opportunities and challenges", *OECD Digital Economy Papers*, No. 321, OECD Publishing, Paris, https://dx.doi.org/10.1787/a6edfab2-en. [26]

OECD (2021), "Measuring smart city performance in COVID-19 times: Lessons from Korea and OECD countries: Proceedings from the 2nd OECD Roundtable on Smart Cities and Inclusive Growth", *OECD Regional Development Papers*, No. 19, OECD Publishing, Paris, https://doi.org/10.1787/72a4e7db-en. [68]

OECD (2021), *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, OECD Publishing, Paris, https://dx.doi.org/10.1787/589b283f-en. [47]

OECD (2021), *Recommendation of the Council concerning Access to Research Data from Public Funding*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347. [45]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463. [128]

OECD (2021), "Reinforcing Democracy: Addressing the Main Governance Challenges – OECD", webpage, https://www.oecd.org/governance/reinforcing-democracy/ (accessed on 17 May 2022). [66]

OECD (2021), *The E-Leaders Handbook on the Governance of Digital Government*, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/ac7f2531-en. [53]

OECD (2020), "Cities policy responses", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/fd1053ff-en. [69]

OECD (2020), *Consumer data rights and competition*, OECD, Paris, https://one.oecd.org/document/DAF/COMP(2020)1/en/pdf. [28]

OECD (2020), *Development Co-operation Report 2020: Learning from Crises, Building Resilience*, OECD Publishing, Paris, https://doi.org/10.1787/f6d42aa5-en. [27]

OECD (2020), "Digital Government Index: 2019 results", *OECD Public Governance Policy Papers*, No. 03, OECD Publishing, Paris, https://doi.org/10.1787/4de9f5bb-en. [141]

OECD (2020), *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, OECD Publishing, Paris, https://dx.doi.org/10.1787/947717bc-en. [43]

OECD (2020), "Ensuring data privacy as we battle COVID-19", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/36c2f31e-en. [105]

OECD (2020), *Financial Consumer Protection Policy Approaches in the Digital Age: Protecting consumers' assets, data and privacy*, https://www.oecd.org/finance/Financial-Consumer-Protection-Policy-Approaches-in-the-Digital-Age.pdf. [29]

OECD (2020), "Food Supply Chains and COVID-19: Impacts and Policy Lessons", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/71b57aea-en. [6]

OECD (2020), "Open, Useful and Re-usable data (OURdata) Index: 2019", *OECD Policy Papers on Public Governance*, No. 1, OECD, Paris, https://www.oecd.org/gov/digital-government/policy-paper-ourdata-index-2019.htm (accessed on 12 May 2021). [56]

OECD (2020), "Report on the Implementation of the OECD Recommendation concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data", OECD, Paris, https://one.oecd.org/document/C(2021)42/en/pdf. [90]

OECD (2020), "The COVID-19 crisis: A catalyst for government transformation?", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/1d0c0788-en. [57]

OECD (2020), "The role of transparency in avoiding a COVID-19 induced food crisis", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/d6a37aeb-en. [16]

OECD (2020), "Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/8f394636-en. [106]

OECD (2020), "Why open science is critical to combatting COVID-19", *OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/cd6ab2f9-en. [46]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, https://dx.doi.org/10.1787/276aaca8-en. [3]

OECD (2019), "Good practice guide on consumer data", *OECD Digital Economy Papers*, No. 290, OECD Publishing, Paris, https://dx.doi.org/10.1787/e0040128-en. [33]

OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/059814a7-en. [107]

OECD (2019), *Unpacking E-commerce: Business Models, Trends and Policies*, OECD Publishing, Paris, https://dx.doi.org/10.1787/23561431-en. [34]

OECD (2018), "Consumer policy and the smart home", *OECD Digital Economy Papers*, No. 268, OECD Publishing, Paris, https://dx.doi.org/10.1787/e124c34a-en. [30]

OECD (2018), *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/9789264305847-en. [24]

OECD (2017), "Benefits and challenges of digitalising production", in *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, http://10.1787/9789264271036-6-en. [11]

OECD (2017), "Business models for sustainable research data repositories", *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris, https://doi.org/10.1787/302b12bb-en. [110]

OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264276284-en. [103]

OECD (2017), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264271036-en. [32]

OECD (2016), *Recommendation of the Council on Health Data Governance*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433. [104]

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229358-en. [49]

OECD (2015), *Digital Security Risk Management for Economic and Social Prosperity*, https://www.oecd.org/digital/ieconomy/digital-security-risk-management.pdf. [124]

OECD (2015), "Drawing value from data as an infrastructure", in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264229358-8-en. [2]

OECD (2015), *Enquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris, https://www.oecd.org/sti/ieconomy/KBC2-IP.Final.pdf. [85]

OECD (2015), "Governments leading by example with public sector data", in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264229358-14-en. [123]

OECD (2014), *Recommendation of the Council on Digital Government Strategies*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0406. [59]

OECD (2013), *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/9789264193505-en. [116]

OECD (2002), *Inventory of Privacy-Enhancing Technologies (PETs)*, OECD, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=dsti/iccp/reg%282001%291/final. [127]

OECD (n.d.), "7 main challenges addressed", *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, webpage, https://community.oecd.org/community/cstp/enhanced-data-access/overview (accessed on 6 October 2022). [145]

OECD (forthcoming), "Consumer Vulnerability in the Digital Age", *OECD Digital Economy Policy Papers*, OECD Publishing, Paris. [98]

OECD (forthcoming), "Emerging privacy enhancing technologies: Maturity, opportunities and challenges", *OECD Digital Economy Papers*, OECD Publishing, Paris. [126]

OECD (forthcoming), "Enhancing online disclosure effectiveness", *OECD Digital Economy Papers*, OECD Publishing, Paris. [92]

OECD (n.d.), "War in Ukraine: Tackling the policy challenges", webpage, https://www.oecd.org/ukraine-hub/en/ (accessed on 7 October 2022). [156]

OECD and Govlab (2021), *Open data in action: Initiatives during the initial stage of the COVID-19 pandemic*, OECD Publishing, Paris, https://www.oecd.org/fr/gouvernance/gouvernement-numerique/use-of-open-government-data-to-address-covid19-outbreak.htm. [58]

Oliveira Hashiguchi, T., L. Slawomirski and J. Oderkirk (2021), "Laying the foundations for artificial intelligence in health", *OECD Health Working Papers*, No. 128, OECD Publishing, Paris, https://dx.doi.org/10.1787/3f62817d-en. [36]

Open Contracting Partnership (n.d.), "Open Contracting Data Standard 1.1.5 documentation", webpage, https://standard.open-contracting.org/latest/en/ (accessed on 7 October 2022). [62]

Open Data for Development Network (n.d.), *Open Data for Development Network*, website, https://www.od4d.net/ (accessed on 4 November 2022). [139]

openownership.org (n.d.), "Beneficial Ownership Data Standard", webpage, https://www.openownership.org/en/topics/beneficial-ownership-data-standard/ (accessed on 7 October 2022). [64]

Paic, A. (2021), "Making data for science as open as possible to address global challenges", *OECD Innovation Blog*, https://express.adobe.com/page/wt6Xz7XVVX91k/ (accessed on 10 November 2022). [44]

Paic, A. (2021), "Open Science – Enabling Discovery in the Digital Age", *Going Digital Toolkit Note*, No. 13, https://goingdigital.oecd.org/data/notes/No13_ToolkitNote_OpenScience.pdf. [42]

Paprica, P. et al. (2020), "Essential requirements for establishing and operating data trusts: Practical guidance based on a working meeting of fifteen Canadian organizations and Initiatives", *International Journal of Population Data Science*, Vol. 5/1, https://ijpds.org/article/view/1353. [131]

PARIS21 (2017), "National Strategies for the Development of Statistics", Paris21, https://paris21.org/national-strategy-development-statistics-nsds. [138]

Perez, J., C. Emilsson and B. Ubaldi (2019), "The OECD 2019 Open Useful Reusable Data (OURdata) Index", *OECD Policy Papers on Public Governance*, No. 1, OECD Publishing, Paris, https://www.oecd.org/gov/digital-government/policy-paper-ourdata-index-2019.htm. [54]

Productivity Commission (2017), "Data availability and use", *Productivity Commission Inquiry Report*, No. 82, Productivity Commission, Canberra, https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf. [133]

Ruhaak, A. (2021), "How data trusts can protect privacy", *MIT Technology Review*, http://www.technologyreview.com/2021/02/24/1017801/data-trust-cybersecurity-big-tech-privacy/. [130]

Scassa, T. (2018), "Data ownership", *CIGI Papers*, Vol. 187, https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf. [83]

Schroeder, K., J. Lampietti and G. Elabed (2021), *What's Cooking: Digital Transformation of the Agrifood System*, World Bank Group, Washington, DC, https://openknowledge.worldbank.org/bitstream/handle/10986/35216/9781464816574.pdf. [14]

Shockley, J., C. Dillon and S. Shearer (2019), "An economic feasibility assessment of autonomous field machinery in grain crop production", *Precision Agriculture*, Vol. 20/2019, pp. 1068-1085, https://doi.org/10.1007/s11119-019-09638-w. [12]

Sposito, F. (2017), "What do data curators care about? Data quality, user trust, and the data reuse plan", presented to IFLI WILIC 2017, http://library.ifla.org/1797/1/S06-2017-sposito-en.pdf. [113]

Sundareswaran, S., A. Squicciarini and D. Lin (2012), "Ensuring distributed accountability for data sharing in the cloud", *IEEE Transactions on Dependable and Secure Computing*, Vol. 9/4, pp. 556-568, https://doi.org/10.1109/TDSC.2012.26. [86]

Susan Ariel Aaronson (2021), "Data is disruptive: How data sovereignty is challenging data governance", 3 August, Hinrich Foundation, Scotsdale, AZ, https://www.hinrichfoundation.com/research/article/digital/data-is-disruptive-how-data-sovereignty-is-challenging-data-governance/. [121]

Sykuta, M. (2016), "Big data in agriculture: Property rights, privacy and competition in ag data services", *International Food and Agribusiness Management Review*, Vol. 19/A, https://www.ifama.org/resources/documents/v19ia/320150137.pdf. [22]

Tversky, A. and D. Kahneman (1974), "Judgment under uncertainty: Heuristics and biases", *Science*, Vol. 185/4157, pp. 1124-1131, https://doi.org/10.1126/science.185.4157.1124. [93]

Ubaldi, B. et al. (2019), "State of the art in the use of emerging technologies in the public sector", *OECD Working Papers on Public Governance*, No. 31, OECD Publishing, Paris, http://www.oecd.org/gov/digital-government/working-paper-the-use-of-emerging-technologies-in-the-public-sector.htm. [50]

UN (n.d.), "UN Global Pulse", webpage, https://www.unglobalpulse.org/about/ (accessed on 7 October 2022). [17]

UN Human Rights Council (2021), *Right to Privacy in the Digital Age*, UN Human Rights Council, Geneva, https://undocs.org/A/HRC/RES/48/4. [81]

United States Government (n.d.), *Federal Data Strategy*, website, https://strategy.data.gov/ (accessed on 2 September 2022). [142]

Verhulst, S. (2021), "Reusing data responsibly to achieve development goals", in *Development Co-operation Report 2021: Shaping a Just Digital Transformation*, OECD Publishing, Paris, https://www.oecd-ilibrary.org/sites/ce08832f-en/1/3/4/10/index.html?itemId=/content/publication/ce08832f-en&_csp_=17c2a7153f8f3e72e475ec60ee15c40c&itemIGO=oecd&itemContentType=book#. [72]

Wallis, J. et al. (2007), "Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries", *Center for Embedded Network Sensing* 4675, https://doi.org/10.1007/978-3-540-74851-9_32. [111]

Wang, Y. et al. (2020), "Use of big data tools and industrial Internet of Things: An overview", *Scientific Programming*, Vol. 2020/8810634, https://doi.org/10.1155/2020/8810634. [20]

WFP (n.d.), *HungerMap LIVE*, website, https://hungermap.wfp.org/ (accessed on 7 October 2022). [18]

Wiebe, A. (2017), “Protection of industrial data – A new property right for the digital economy?”, *Journal of Intellectual Property Law & Practice*, Vol. 12/1, pp. 62-71, https://doi.org/10.1093/jiplp/jpw175. [19]

World Bank (2022), “Food Security Update”, webpage, http://www.worldbank.org/en/topic/agriculture/brief/food-security-update (accessed on 2 October 2022). [9]

Endnotes

[1] A general-purpose resource can be used as an input into a wide range of goods and services, thus providing "basic, multipurpose functionality." How data are used will typically depend on the initial purpose for which they have been collected. For example, at the outset, agricultural data will primarily be used for agricultural goods and services. However, in theory, there are no limits with regard to how data can be re-used. Many of the benefits stemming from their re-use are based on the fact that data created in one domain can provide further insights in another domain. Use of anonymised mobile call data records (CDRs) of telecommunications services providers, for example, have been re-used to monitor and control the spread of pandemics such as COVID-19.

[2] The capital good property describes the fact that social demand for the resource is driven primarily by downstream productive activities that require the resource as an input. Data, in most cases, are used as an input for goods or services; this is especially true of large volumes of data (i.e. big data), which are means rather than ends in themselves.

[3] This may sound counterintuitive since traditionally infrastructures typically refer to large-scale physical facilities provided for public consumption. The classic examples are transportation systems, communication systems, and basic services and facilities such as buildings, and sewage and water systems. However, as recognised by the US National Research Council (1987[152]), the notion of infrastructure also refers to non-physical facilities. This can include education systems and governance systems (e.g. court system). This is also in line with Merriam-Webster (n.d.[150]), which defines infrastructures as "the resources (such as personnel, buildings, or equipment) required for an activity" and "the underlying foundation or basic framework (as of a system or organization)." According to Frischmann (2012[1]), infrastructural resource satisfy the following three criteria: (i) the resource may be consumed in a non-rivalrous fashion for some appreciable range of demand; (ii) social demand for the resource is driven primarily by downstream productive activities that require the resource as an input; and (iii) the resource may be used as an input into a wide range of goods and services, which may include private goods, public goods, and social goods.

[4] Studies discussed in OECD (2019[3]) show that, while data openness can increase the value of data to holders (direct impact), it can help create 10-20 times more value to data users (indirect impact), and 20-50 times more value for the wider economy (induced impact). In some cases, however, data openness, and in particular open data, may also reduce the producer surplus of data holders. This is the cause of the incentive problem discussed in Section 3. Overall, these studies suggest that data openness can help generate social and economic benefits worth between 0.1-1.5% of GDP in the case of public sector data, and between 1-2.5% of GDP when also including private sector data.

[5] "Food security exists when all people, at all times, have physical, social, and economic access to sufficient, safe, and nutritious food that meets their food preferences and dietary needs for an active and healthy life" (FAO, 1996[146]).

[6] See OECD (n.d.[156]) on the OECD's latest insights, analysis and data to shed light on the policy challenges ahead due to Russia's war against Ukraine.

[7] This term is often used to describe a movement that promotes greater transparency in the scientific methodology used and data collected; advocates the public availability and reusability of data, tools and materials; and argues for broadly communicating research (particularly when publicly funded) and its results.

[8] Base registries include population registries, tax registries, property registers and company registers.

[9] See, for instance, the European Interoperability Framework (European Commission, n.d.[158]).

[10] See Observatory of Public Sector Innovation (n.d.[157]).

[11] See Observatory of Public Sector Innovation (n.d.[154]).

[12] See Observatory of Public Sector Innovation (n.d.[155])

[13] See Observatory of Public Sector Innovation (n.d.[153])

[14] The median budget for data protection authorities in non-OECD countries in 2018 was USD 500 000 while in OECD countries it was USD 6 Million. These constraints can lead to under-regulation, over-regulation or regulating the wrong things in the wrong way. For example, it could lead to too much emphasis on protecting against individual harms and not enough on collective harms (OECD, 2021[78]).

[15] Researchers in collaboration with Cambridge Analytica had asked users to take a personality survey via an app. It collected personal data not only of the 270 000 Facebook users, who had given their consent to participate in what was believed to be a research-related activity, but also personal data of their "friends." This led to the collection of 50 million profiles, which were used by Cambridge Analytica (Granville, 2018[149]).

[16] The EASD Recommendation defines the data ecosystem as "the integration of and interaction between different relevant stakeholders including data holders, data producers, data intermediaries and data subjects, that are involved in, or affected by, related data access and sharing arrangements, according to their different roles, responsibilities and rights, technologies, and business models."

[17] Beyond privacy harms and loss of autonomy, other harms from dark patterns at the individual level include financial loss and psychological detriment, with disproportionate impacts on certain subsets of consumers such as less educated consumers or children. At the collective level, dark patterns may also weaken or distort competition (e.g. because they may bestow a competitive advantage on businesses that use them) and sow consumer distrust and disengagement. Reductions in autonomy through online manipulation can also lead to collective harms beyond the consumer realm, such as threats to democracy and freedom of expression. While evidence does not suggest they are widespread, personalisation practices that leverage personal data to exploit vulnerabilities at the individual level may also present similar harms. Moreover, concrete evidence of harms resulting from such practices is lacking in many cases, however. This means further research is needed to appropriately guide policy and enforcement responses (OECD, 2022[97]; OECD, forthcoming[98]).

[18] A data breach is "a loss, unauthorised access to or disclosure of personal data as a result of a failure of the organisation to effectively safeguard the data" (OECD, 2017[103]).

[19] The severity and impact of data breaches have also increased (OECD, 2017[103]). According to a study released in 2018 by the Ponemon Institute (the data security research organisation) (2014[1]), the total average cost of a data breach is now USD 3.9 million, compared to USD 3.5 million in 2014**.**

[20] DoS incidents affect an organisation by flooding its online service or bandwidth with spam requests, knocking it off line for hours or days (OECD, 2017[103]).

[21] For example, the Australian Consumer Data Right (CDR) is sometimes considered as a "core *infrastructure* to support a data-driven economy" [emphasis added] (Australian Government - Treasury, 2022[148]).

[22] Through a survey of 190 clinical and basic science researchers in the Intramural Research Program at the National Institutes of Health in the United States, Federer et al. (2015[112]) show that while 71% of respondents reported sharing data directly with colleagues, only 39% reported ever having used a data repository service for data sharing.

[23] For non-excludable goods, the cost for one person to prevent another from consuming the resource must be significant. The marginal costs of transmitting, copying and processing data can be close to zero, making it easy for others to reproduce and use it. However, ICTs including user access control and encryption, have also dramatically reduced the costs of exclusion. Thus, where data are kept within a controlled environment the cost of exclusion will be typically low enough to prevent others from using them. Therefore, data should be considered at least a partially excludable good.

[24] Anonymisation is the process of removing identifying elements from data that would allow them to be linked to an individual or organisation. Anonymised data, in theory, should not be linkable back to an individual even when combined with outside/additional data sets.

[25] . Federated learning is increasingly applied in the health sector and was particularly important for accelerating the pace of global COVID-19 research (OECD, 2022[35]).

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

[27] The Australian (Productivity Commission, 2017[133]) Data Availability and Use Inquiry Report recommends to the (i) "government [to establish] a new Office of the National Data Custodian", (ii) "(predominantly existing) public bodies [to] be accredited as sector-based, national data release authorities" and that (iii) "a small number of nationally beneficial data sets be designated as National Interest Datasets."

[28] It was followed by the Australian Institute of Health and Welfare (AIHW), the Australian Institute of Family Studies (AIFS), the Department of Social Services (DSS), the Queensland Government Statistician's Office (QGSO), the Centre for Victorian Data Linkage (CVDL), and South Australia Northern Territory DataLink (SA NT DataLink). (Australian Government, n.d.[147]).

[29] "Promote the rights of citizens to manage their personal data in accordance with the MyData principle." is one objective of the Finnish government's 2022-2023 (Ministry of Finance [Finland], n.d.[151]).

[30] Government bodies in OECD and non-OECD countries responsible for development co-operation are also beginning to consider policy coherence issues raised by the role played by other parts of their domestic administrations in the creation and enforcement of norms and standards for digital space that can have spill-over effects on developing countries (e.g. adequacy criteria that disproportionately impacts developing countries).

[31] A few national case studies are available on the OECD online community on "Enhanced Access to Publicly Funded Data for Science, Technology and Innovation" (OECD, n.d.[145]).