Organisation for Economic Co-operation and Development

Unclassified **English - Or. English**

5 February 2025

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
PUBLIC GOVERNANCE DIRECTORATE
COMMITTEE ON DIGITAL ECONOMY POLICY
COMMITTEE FOR SCIENTIFIC AND TECHNOLOGICAL POLICY
PUBLIC GOVERNANCE COMMITTEE

Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence

Companion Document to the OECD Council Recommendation on Enhancing Access to and Sharing of Data

This document was approved and declassified by the Digital Policy Committee (DPC), the Committee for Scientific and Technological Policy (CSTP) and the Public Governance Committee (PGC), on 21 November 2024.

JT03559305

Foreword

The OECD Recommendation on Enhancing Access to and Sharing of Data (hereafter the "Recommendation") [OECD/LEGAL/0463] instructs the Digital Policy Committee (DPC), the Committee for Scientific and Technological Policy (CSTP) and the Public Governance Committee (PGC), (hereafter the "three partner committees") to develop and iterate further practical guidance on its implementation. This Companion Document provides practical guidance on the implementation of the Recommendation and is intended to complement and be read in conjunction with the Recommendation. The Companion Document was developed primarily based on a policy survey conducted via the DPC's Working Party on Data Governance and Privacy (WPDGP) from May 2022 to January 2024. This effort helped to cover the growing number of cross-sectoral data policy initiatives and legislative developments in OECD Members and partner economies as well as the European Union. Additionally, unique perspectives were included from policy initiatives in scientific research and open government data based on relevant work by the CSTP and the PGC respectively, along with the PGC's Working Party of Senior Digital Government Officials (E-Leaders).

The Companion Document also benefited from input from a joint steering group of experts nominated by the three partner committees to support the work. It also benefited from the Horizontal Project Going Digital Phase III on Data Governance for Growth and Well-Being. The Companion Document will serve as a foundation for the three partner committees' joint work in reviewing the implementation, dissemination, and continued relevance of the Recommendation, and reporting to the Council thereon in 2026, as instructed by the Council.

This document was prepared by Christian Reimsbach-Kounatze with contributions from Alain Paic, Andras Molnar, Barbara Ubaldi, Idil Uzun, and Jacob Arturo Rivera Perez, and under the supervision of Clarisse Girot. It was approved and declassified by the three partner committees on 21 November 2024 [COM/DSTI/CDEP/STP/GOV/PGC(2024)1]. The previous Note by the Secretariat has been replaced by a Foreword and minor changes to the formatting have been made.

Please cite this work as:

OECD (2025), "Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence: A Companion Document to the OECD Recommendation on Enhancing Access and Sharing of Data", https://one.oecd.org/document/COM/DSTI/CDEP/STP/GOV/PGC(2024)1/FINAL/en/pdf

Table of contents

Boxes

Executive summary

The *OECD Recommendation on Enhancing Access to and Sharing of Data* (hereafter the "Recommendation") adopted in October 2021 sets out general principles and policy guidance on how governments can maximise the benefits of enhancing data access and sharing arrangements while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives. It is the first internationally agreed upon set of principles on how to maximise the cross-sectoral benefits of all types of data with trust – public and private sector data included.

With artificial intelligence (AI) regulation on the rise, the access to and sharing of general AI models present both benefits and risks, while AI system developers face legal and other challenges in collecting sufficient high-quality data. In this context, enhancing access to and sharing of data (EASD) may be more crucial than ever for the future of AI.

As a collaborative initiative spearheaded by the Digital Policy Committee (DPC), the Committee for Scientific and Technological Policy (CSTP), and the Public Governance Committee (PGC), the Recommendation bridges different policy perspectives on data governance, and data access and sharing more specifically. This interdisciplinary approach helps to identify and address key considerations across policy domains and sectors when enhancing access to and sharing of data.

This Companion Document provides detailed and practical information on how Adherents (the OECD Members and partner economies having adhered to the Recommendation) can implement the provisions of the Recommendation particularly in view of recent developments in AI. In addition to governments, to whom the Recommendation is addressed directly, the Recommendation also encourages data holders, data producers, data intermediaries, and other relevant stakeholders in the data ecosystem to implement or, as appropriate according to their role, support and promote the implementation of this Recommendation. Consequently, this Companion Document can help both public and private sector actors, including "those who play an active role in the AI system lifecycle" (AI actors), implement the Recommendation in areas related to data access and sharing.

The Companion Document provide insights into three main areas: (i) the scope and structure of the Recommendation, (ii) the key concepts that are fundamental for its understanding and implementation, and (iii) specific examples in the public and private sector, highlighted in dedicated boxes, that can help inform the implementation of the Recommendation's principles.

Structure and scope of the Recommendation

The Recommendation provides guidance on *how* (e.g., voluntary or mandatory) public policies and arrangements on data access and sharing may be implemented to maximise their envisaged benefits while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives, alongside broader efforts to promote and enable a culture of responsibility for data governance throughout the data value cycle. The Recommendation focuses in particular on how to: (i) reinforce trust across the data ecosystem, (ii) stimulate investment in data, and incentivise data access and sharing, and (iii) foster effective and responsible data access, sharing and use across society. The Recommendation is not intended to address questions of *whether* or *when* to regulate access to data

(including data of public interest) although it does call on Adherents to "seek to maximise the benefits of measures for enhancing data access and sharing". This Companion Document however provides examples of policy initiatives that can help identify the conditions under which governments have regulated access to and sharing of data.

Key concepts

The following concepts are critical for the understanding of the provisions of the Recommendation, their interconnections, and thus their implementation, and they may also be foundational for future OECD legal instruments on data governance. The ambition is to adopt a common and robust conceptual framework of broad applicability that ensures legal and policy coherence, while accommodating the unique purposes of more specific definitions and opening new pathways for governance innovations:

- *Data, AI and other complementarities along the data value cycle*: The Recommendation takes a comprehensive and dynamic perspective on data, which it defines as "recorded information in structured or unstructured formats, including text, images, sound, and video". In the context of machine learning (ML) and AI, this includes both data used to train AI systems (AI input) and AI models, which encode information from AI input into their model parameters (e.g., weights) during the training process. Additionally, the Recommendation outlines a comprehensive data value cycle, encompassing stages from data creation and collection through to enrichment, processing and analysis, and eventually deletion. This dynamic perspective on data not only underscores the vital role played by complementary resources, such as other digital resources (e.g. algorithms and software) and human resources (e.g. skills). It also draws attention to the role of technological and organisational environments and methods for enhancing data access and sharing.
- *The data openness continuum*: The Recommendation promotes a differentiated approach to data access and sharing that leverages the "data openness continuum". This continuum covers a wide range of data access and sharing arrangements with variable degrees of openness that can be adjusted through technical, organisational and legal means so that data, as well as ML/AI models, can be as open as possible to maximise the risk-adjusted benefits and as closed as necessary to protect legitimate public and private interests.
- *The data ecosystem, the overlapping and conflicting rights, and interests of stakeholders*: The Recommendation recognises that data and the value derived from their use are often (co-) created as a result of the involvement and contributions of the relevant stakeholders within the data ecosystem, 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". It thereby emphasises "co-operation and trust between all stakeholders [as] crucial to shared value creation in the data ecosystem". However, these various stakeholders may have potentially conflicting rights and interests which policy makers need to balance in line with applicable laws and regulations.

EASD Principles and key findings from implementation examples

The main principles of the Recommendation are divided into three overall sections. Specific examples that can help inform the implementation of these principles were selected from a set of policies covering a total of 40 countries plus the European Union[1] as well as private sector initiatives.

Reinforcing trust across the data ecosystem

Trust plays an essential role in data access and sharing across organisations, sectors, and jurisdictions. To help reinforce trust across the data ecosystem, the Recommendation calls on Adherents to address the following three issues:

1. *Empowering and pro-actively engaging all relevant stakeholders alongside broader efforts to increase the trustworthiness of the data ecosystem*: Countries that engage with their relevant stakeholders to understand their respective interests and concerns do so via public consultations based on published documents, the establishment of working groups and standardisation processes, and the organisations of open events such as expert workshops and forums. The adoption of mechanisms to ensure the appropriate consideration of stakeholder input varies between countries. Data portability and the role of trusted data intermediaries (TDIs) are increasingly being considered as promising instruments for enhancing control and agency of users over data concerning them. Furthermore, in some countries, data-sharing public private partnerships (PPPs) foster collaboration between the public and private sector. Transparency is recognised as a means for enhancing trust and fairness within the data ecosystem, clearly evident in its essential role in privacy, data protection, and AI governance frameworks, despite conceptual differences in the latter.
2. *Adopting a strategic whole-of-government approach to ensure that data access and sharing arrangements effectively and efficiently meet specific societal, policy, and legal objectives that are in the public interest*: Some countries implement national or sectoral data strategies in line with a whole-of-government approach to data governance, increasingly in connection with their national AI strategies. The assessment of country examples suggests that sectoral data strategies, and in particular public sector data strategies appear most frequently. The establishment of inter-ministerial bodies and working groups that coordinate policy measures is often used as means to enable a whole-of-government approach without the need to implement a data strategy. Regulatory sandboxes are increasingly being explored to enable technology-neutral and agile legal and regulatory environments.
3. *Maximising the benefits of data access and sharing, while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives*: Most countries analysed refer to their national privacy and data protection legal frameworks when it comes to the protection of individual rights related to data access and sharing. Some countries are exploring new and complementary approaches to maximise data use while protecting privacy rights. In this context, governments and regulators are looking into the use of e.g. privacy enhancing technologies (PETs) and TDIs; as well as the adoption of new regulations and regulatory institutions that focus for instance on data governance and privacy implications in the context of AI.

Stimulating investment in data, and incentivising data access and sharing:

Substantial investments are often required over time to collect and manage data, including managing their associated risks, throughout the data value cycle. Such investments may be sustained by different business and financial models, including market-based approaches that rely on commercialisation of data and related contractual arrangements. The Recommendation calls on Adherents to provide coherent incentive mechanisms and promote conditions for the development and adoption of sustainable business models and markets for data access and sharing:

1. *Encouraging market-based approaches by fostering competitive markets for data and promoting, where appropriate, self- or co-regulation mechanisms*: Issues related to competition tend to be addressed primarily by competition enforcement authorities on a case-by-case basis, and competition authorities are stepping up their efforts to strengthen their enforcement capacities in

data- and AI-related cases. The establishment of guidance, codes of conduct or templates for data access and sharing are means used to encourage market-based approaches while reducing legal uncertainties and transaction costs, including in the area of data for AI. However, to address systematic competition issues related to data access and sharing, and observed limitations of self-regulation mechanisms, some policy makers have begun to consider complementary ex ante regulatory measures. These measures may include mandatory data access and portability arrangements focusing in particular on infrastructural sectors such as finance (open banking / finance), transportation and health care as well as online platform operators deemed as "gatekeepers". At the same time, the Recommendation recognises that such measures must ensure an adequate level of consumer, intellectual property, and privacy and personal data protection to foster competitive markets.

2. *Promoting conditions for the development and adoption of sustainable business models and markets for data access and sharing*: Market-based approaches to data access and sharing, including commercialisation of data and freedom of contract, are essential for incentivising data access and sharing and related investments. The Recommendation also recognises that "there may be costs, risks, and limitations to these approaches' ability to fully meet demand for data". Policy makers are challenged by the tension between the requirements for meeting the demand for data, and the financial requirements for the establishment and operation of data sharing activities over the long term. This challenge is particularly, but not exclusively, pertinent in the context of open data arrangements. Some financing models for open data repositories that are being adopted in particular in the context of open science include structural funding, host or institutional funding, annual deposit-side contract, data deposit fees, and project-based funding. In recent years, public sector funders have increasingly required that a percentage of research grants be allocated to provide for open or publicly accessible data.
3. *Promoting appropriate incentive mechanisms*: Policies need to promote appropriate incentive mechanisms that enable the fair distribution of the benefits of data access and sharing arrangements and ensure that stakeholders are enabled, encouraged, recognised and rewarded for engaging in data access and sharing arrangements. The creation of guidance documents, codes of conduct, or templates, coupled with ex ante regulatory measures (as discussed above) are also often used to achieve a balance between rights and obligations related to data access and sharing, and their fair outcome. Standardised contractual terms for data sharing have attracted a lot of interest, with the open source community leading the way in their development. Innovative approaches to increase the recognition and rewards for engaging in data access and sharing arrangements can be observed in open science. For example, the dissemination of scientific data together with data citations are increasingly considered as factors for the performance evaluation of scientists and as a requirement for granting public funding. Under some legal frameworks, certain data access and sharing practices, including the release of AI models under open source licenses (open source AI models), are incentivised by linking them to less stringent regulatory obligations under certain conditions.

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

The availability and accessibility of data does not guarantee effective use and reuse across organisations, sectors and jurisdictions. In response to this challenge, the Recommendation guides Adherents to consider the following three issues when establishing policies for data access and sharing across society:

1. *Further improving conditions for cross-border data access and sharing with trust*: The approaches presented above on "Reinforcing trust across the data ecosystem", such as the appropriate use of e.g. PETs in accordance with national privacy and data governance frameworks, can help further improve conditions for cross-border access to and sharing of personal and non-personal data where they can enhance trust. Furthermore, such approaches should help (i) ensure that measures

that condition cross-border data access and sharing are non-discriminatory, transparent, necessary, and proportionate to the level of risk; and (ii) promote continued dialogue and international co-operation on ways to foster data access and sharing across jurisdictions. Most initiatives target the latter.

2. *Fostering the findability, accessibility, interoperability, and reusability of data ("FAIR data") across organisations, including within and across the public and private sectors*: The analysis of examples from various countries indicates that a significant number of government initiatives are specifically promoting FAIR data. This is particularly evident in the fields of public sector data as well as scientific research, where there is an increasing requirement for metadata to be machine-readable. Policy measures may also involve the establishment of dedicated data platforms, warehouses, and repositories that are used to validate, combine, and release public sector and research data. Some countries are also considering establishing or have established new institutions whose mandate would lie in ensuring FAIR data among other mandates, some of which involving extending the mandate of existing institutions such as national statistical offices to act as TDIs. FAIR data and data quality are becoming increasingly critical with the rise of generative AI models, which rely significantly on large amounts of data. The increasing release of AI models as "open source"[1], with various degrees of openness, also raises questions about the conditions necessary not only for FAIR data but also for the findability, accessibility, interoperability and reusability of the AI models themselves (FAIR AI models).
3. *Enhancing the capacity of all stakeholders to use data more effectively and responsibly:* The analysis of country examples suggests a significant number of government initiatives aimed at enhancing stakeholders' capacity to use data more effectively and responsibly, some of which targeting individuals or small and medium-sized enterprises (SMEs) in particular. Efforts range from raising awareness among individuals and businesses about the economic benefits of data access and sharing and about how to mitigate the risks through the use of technologies such as PETs, to fostering 'data literacy' skills including within higher education and research institutions. Promoting the adoption of data-related ICT infrastructures is also being considered although not as frequently as in the case of skills and awareness raising campaigns.

1. Introduction

With the increasing significance of data for digital transformation, access to and sharing of data are essential for economic and social activities in the private and public sectors. The benefits of data access and sharing may include i) contributing to greater efficiency, transparency and accountability across society; ii) providing support to address societal challenges and global emergencies; iii) boosting sustainable growth and enhance social welfare and well-being; iv) improving evidence-based policy making as well as public service design and delivery;[2] v) empowering users of digital goods and services, including enterprises, workers, citizens and consumers; and vi) facilitating scientific discovery through enhanced opportunities for research, reproducibility of scientific results, and cross-disciplinary co-operation[3]. Through all these benefits data access and sharing can contribute directly or indirectly to achieving the United Nations (UN) Sustainable Development Goals (SDGs) (OECD, 2022[1]).

In regard to its economic benefits, data access and sharing can help increase the value of data to data holders and may create 10 to 20 times more value to data users, and 20 to 50 times more value to the economy (OECD, 2019[2]). The monetisation of data contributes ten percent or more to the overall revenue of the 32 percent of high-performing businesses and nine percent to all other businesses (OECD, 2020[3]). In addition, businesses that use data have approximately five to ten percent faster productivity growth than those that do not (OECD, 2015[4]; OECD, 2020[3]). The secondary effects of data access and sharing are estimated to be between 0,5 percent to 1,2 percent of GDP, depending on the study (McKinsey Global Institute, 2013[5]; Deloitte, 2013[6]; Lateral Economics, 2014[7]).

Data access and sharing can also significantly help to enhance social welfare and resolve societal challenges through better public research, public services and public policies. For instance, during the COVID-19 pandemic, timely, secure and reliable data access and sharing were key to understanding COVID-19 and its spread, enhancing the effectiveness of government policies, and supporting global co-operation in the research, development and distribution of vaccines and treatments (OECD, 2020[8]; OECD, 2020[9]; OECD, 2021[10]). Lessons from previous epidemics have also underlined the importance of data concerning the spread of the virus (OECD-Harvard Global Health Institute, 2017[11]; OECD, 2020[3]; OECD, 2020[12])

The rise of Artificial Intelligence (AI), including in particular generative AI, has further contributed to the growing importance of data access and sharing. Open data, for instance, has been crucial to the remarkable advancements in AI, where data serve as inputs for machine learning (ML) algorithms. For example, AlexNet, the deep neural network that initiated the deep learning revolution ten years ago, was only made possible thanks to a human-annotated dataset of 3.2 million images (ImageNet), created and openly shared for scientific purposes by its creators (Deng et al., 2010[13]; Liesenfeld, Lopez and Dingemanse, 2023[14]). Likewise, recent breakthroughs in protein folding achieved by the AlphaFold deep learning system was only made possible thanks to open access to the Protein Data Bank, which was established almost half a century ago (Bernstein et al., 1977[15]; Liesenfeld, Lopez and Dingemanse, 2023[14]).

In an era where developers of generative AI models face a scarcity of input data despite extensive and increasingly controversial web scraping practices[4], enhancing access to and sharing of data (EASD) may be more crucial than ever before for the future of AI. In this context, the rapid uptake of open source AI models, i.e. the release of AI model parameters under open source licenses, warrants special attention.

(OECD, forthcoming[16]) By enabling innovative reuse and recombination of existing (foundation) models, these open approaches promise to democratise access to powerful AI models while reducing the need to build new models from scratch. Besides the economic benefits enabled by sharing AI models, the environmental benefits should not be underestimated given the significant volume of CO2 emissions associated with training a large language model (LLM) for instance. (OECD, 2022[17])

There is clear evidence of the growing need for data and of the economic and social benefits of data sharing. Yet, data sharing remains below its potential as individuals, businesses, and governments can face different sorts of legal, technical and economic restrictions, which may be compounded by reluctance to share, including within organisations and sectors, or the lack of data in common digital formats.

In particular, data access and sharing come with several risks that, if not properly addressed, can undermine the trustworthiness of both the data ecosystem and AI more specifically. These risks may include confidentiality and privacy breaches, violations of intellectual property rights (IPRs), and compromises to commercial and national security interests. Furthermore, data access and sharing can adversely impact other legitimate private and public interests, extending beyond the immediate data ecosystem to include concerns such as risks to endangered species (Paic, 2021[18]; OECD, 2007[19]). Where AI models are shared, further risks may be introduced such as the propagation of disinformation and illegal content. Additional risk factors may also include inadequate skills of data users, which lead to unexpected and undue biases and discrimination, the use of poor-quality data (including metadata), erroneous interpretation or misleading data analysis. Perceptions of inadequate treatment of these risks may lead to mistrust from the public and result in increased restrictions to data access and sharing.

1.1. Rationale: Improving the coherence of data governance frameworks for data access and sharing, including for AI

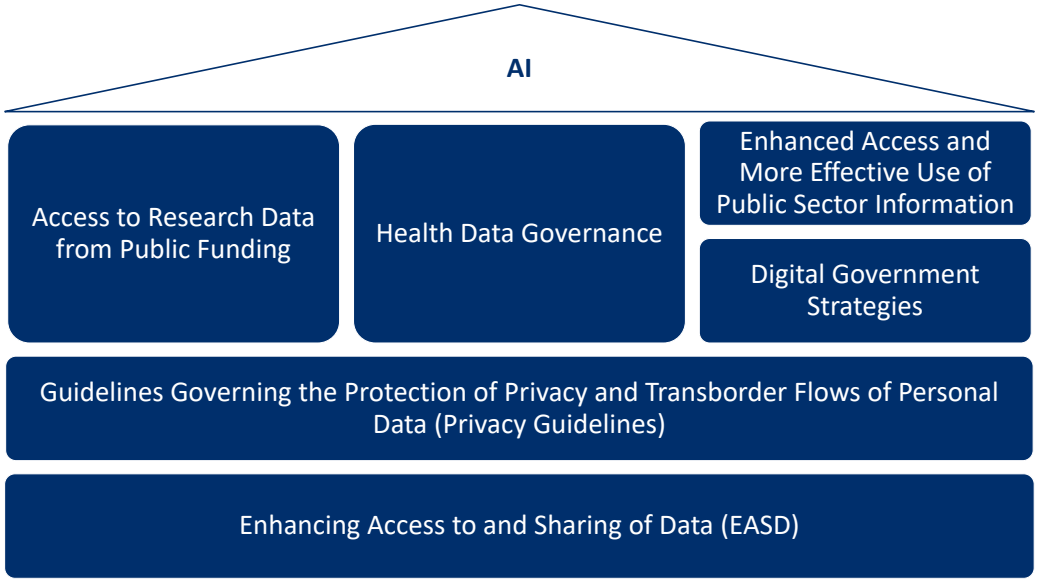
The OECD's work on data governance (OECD, 2018[20]; 2019[2]; 2019[21]; 2020[22]; 2021[23]) highlights the need for greater convergence between, and more coherent data governance frameworks as data access and sharing are increasingly taking place across sectors and jurisdictions. Many of the gains of data access and sharing are based on the fact that data created in one domain and sector can provide further insights when applied in another domain or sector. A clear example of this is open government data, where data sets used originally for administrative purposes have been re-used by various actors including entrepreneurs, academics, scientists, journalists, and civil society representatives to create services unforeseen when the data were originally created.

Furthermore, data governance is becoming an increasingly important issue across various policy domains, ranging from agriculture, health, finance, and transportation to competition, public governance, science, and trade. The latest policy domain high on policy makers' agenda is AI, which raises questions about the access, sharing, and re-use of data and AI models. All these policy domains share a common set of challenges related to data access and sharing (e.g. trust, privacy and data protection, IPRs, transparency, etc) that need to be addressed coherently. (OECD, 2022[24]; OECD, 2022[25])

To support governments develop coherent data governance frameworks that maximise the cross-sectoral benefits of data whilst protecting the rights of individuals and organisations, the OECD Council, on 6 October 2021, adopted the Recommendation on Enhancing Access to and Sharing of Data [OECD/LEGAL/0463] (hereafter "the Recommendation"). The Recommendation was developed by three partner committees, the Digital Policy Committee (DPC) via its Working Party on Data Governance and Privacy (DGP), the Committee for Scientific and Technological Policy (CSTP), and the Public Governance Committee (PGC) via its Working Party of Senior Digital Government Officials (E-Leaders). The development was informed by an open and inclusive multi-stakeholder process.[5]

The Recommendation aims to reinforce trust across the data ecosystem, stimulate investment in data and incentivise data access and sharing, and foster effective and responsible data re-use across sectors and jurisdictions. It is part of a broader body of Recommendations, guidance documents, and analytical work by the OECD on digital policy, public sector governance, and science and technology policy. Figure 1 provides an overview of the OECD Recommendations relating to data governance.

Figure 1. OECD Recommendations relating to data governance



Source: Based on OECD (2022[25]), *Going Digital Guide to Data Governance Policy Making*, https://10.1787/40d53904-en, and online Compendium of OECD legal instruments, https://legalinstruments.oecd.org (accessed 1 June 2024).

Further, the Recommendation seeks to provide a common reference for existing and new OECD legal instruments that provide guidance for policy making in relation to the governance of data in areas such as research, health, and digital government.[6] In so doing, the Recommendation helps foster and ensure coherence across OECD guidance in line with the objectives of the Standard-Setting Action Plans of the DPC, CSTP, and PGC, as well as help ensure coherence with other international reference documents including G7 and G20 ministerial declarations.

1.2. Scope of the Recommendation

OECD Recommendations are adopted by the OECD Council and are not legally binding. They represent a political commitment to the principles they contain and entail an expectation that Adherents will do their best to implement them. (OECD, n.d.[26]) Non-Adherents are invited to take due account of and adhere to the Recommendation. Governments beyond OECD Membership can use the Recommendation to inform the development of their policies, including their national strategies. In addition, data holders, data producers, data intermediaries, and other relevant stakeholders in the data ecosystem are encouraged to implement or, as appropriate according to their role, support and promote the implementation of the Recommendation.

The Recommendation sets out general principles and policy guidance on how governments can maximise the benefits of enhancing data access and sharing arrangements while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives. These arrangements may include data access and sharing arrangements based on voluntary and mutually agreed commercial

or non-commercial terms (e.g. open data initiatives in the private sector) as well as data access and sharing arrangements that are regulated by law or are mandatory (e.g. some open government data, data portability and research data initiatives).

The Recommendation recognises "that data access and sharing arrangements, including government access to proprietary and personal data held by the private sector, may involve activities governed by specific national and international legal frameworks that need to be taken into account in such arrangements". In this context, the *Declaration on Government Access to Personal Data Held by Private Sector Entities* [OECD/LEGAL/0487], adopted in 2022, is pertinent. The Declaration seeks to improve trust in cross-border data flows by clarifying how national security and law enforcement agencies can access personal data under existing legal frameworks.

The Recommendation is not intended to address questions of *whether* or *when* to regulate access to data (including data of public interest), given that the answer is context-dependent and thus will differ on a case-by-case basis. The Recommendation rather provides guidance on *how* (e.g., voluntary or mandatory) data access and sharing arrangements may be implemented to assure that their envisaged benefits are maximised while individuals' and organisations' rights are protected. That said, this Companion Document provides examples of policy initiatives that can help identify the conditions under which governments have regulated access to and sharing of data.

Primarily, the general principles and policy guidance of the Recommendation are aimed at data in digital formats. The Recommendation thereby recognises "the importance of data-driven innovation, including AI and the Internet of Things (IoT), the growing demand for data across society, including on the part of both public and private sector organisations and individuals, and the enhanced ability to collect, access, share and use data as it is increasingly stored in digital formats".

The Recommendation is pertinent to data access and sharing in the context of AI during both the build (pre-deployment) and use (post-deployment) phases. This Companion Document looks at access to and sharing of both, data used to train AI systems (AI input); and AI models, which encode information from AI input into their model parameters (e.g. weights) during the training process. The focus is on AI systems[7] that are built via machine learning techniques, as other approaches such as symbolic or knowledge-based AI systems may rely less on big data collection and use.

1.3. Structure of the Recommendation

The Recommendation starts with a preamble (e.g., "Having regard", "Recognising", etc.), followed by the purpose of the Recommendation ("I. Agrees…") and clarification about terminology ("II. Agrees…"). The latter includes definitions of terms which should be a reference to future OECD standards on data governance. Harmonised based on terms used across other OECD Recommendations, the definitions are intended to allow for maximum flexibility and applicability, so that they can serve as a common framework while still allowing for further specification to meet the requirements of more specific policies.

Thereafter, the Recommendation includes numbered recommendations, which are addressed directly to Adherent governments. The Recommendation also encourages data holders, data producers, data intermediaries, and other relevant stakeholders in the data ecosystem to implement or, as appropriate according to their role, support and promote the implementation of this Recommendation; without however specifying how this could be undertaken.

The preamble to the Recommendation recognises key factors and background considerations that policy makers need to take into account when establishing data access and sharing arrangements. These include the significance of data and the importance of data sharing across society including for data-driven innovation and the wide range of benefits that data access and sharing can generate. The preamble also recognises the need to foster trustworthiness and safeguard against risks such as potential breaches of

confidentiality or privacy, unethical uses of data or the violation of other legitimate private or public interests, including IPRs and national security interests. In addition, the preamble highlights key considerations such as the heterogeneity of the data ecosystem and its complementarities, and the data openness continuum, which are discussed in more detail in the next section on "Understanding key concepts".

The seven main principles of the Recommendation are divided into three overall sections, which are discussed in more detail in the section on "Principles to enhance access to and sharing of data, and implementation examples" with examples of government and private sector initiatives. These three sections and seven principles are:

Section 1: Reinforcing Trust across the Data Ecosystem

1. Empowering and pro-actively engaging all relevant stakeholders alongside broader efforts to increase the trustworthiness of the data ecosystem;
2. Adopting a strategic whole-of-government approach to data access and sharing;
3. Maximising the benefits of data access and sharing, while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives alongside broader efforts to promote and enable a culture of responsibility for data governance;

Section 2: Stimulating Investment in Data and Incentivising Data Access and Sharing

4. Providing coherent incentive mechanisms and promoting conditions for the development and adoption of sustainable business models and markets for data access and sharing;

Section 3: Fostering Effective and Responsible Data Access, Sharing, and Use across Society

5. Further improving conditions for cross-border data access and sharing with trust;
6. Fostering the findability, accessibility, interoperability and reusability of data across organisations, including within and across the public and private sectors; and
7. Enhancing the capacity of all stakeholders to effectively use data responsibly along the data value cycle.

2. Understanding key concepts

This section introduces the key concepts that are fundamental for understanding and implementing the principles of the Recommendation. Some of these key concepts are included in the definition section of the Recommendation, while others are introduced in its preamble. This section does not explore each individual term defined or used in the Recommendation. Instead, it focuses on concepts that are considered "key", i.e. critical for guidance on implementation and for an understanding of the interconnections between the various principles of the Recommendation.

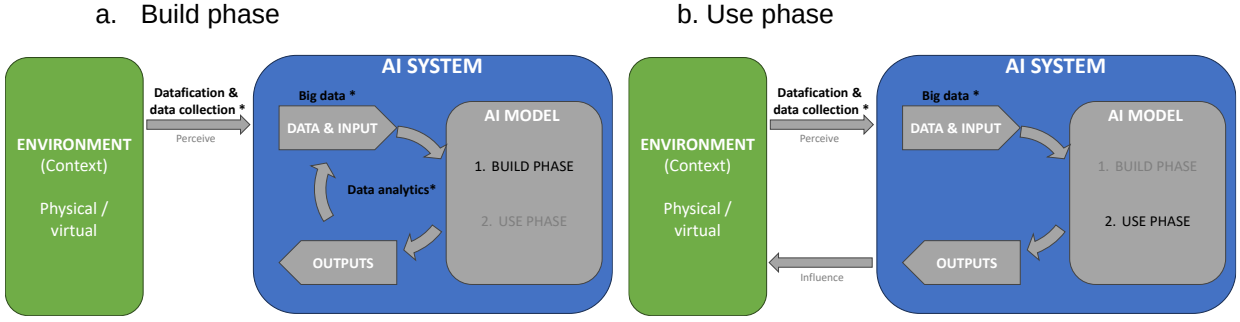
2.1. Data, AI and other complementarities along the data value cycle[8]

Data are increasingly a key resource that can enhance value creation and foster new industries, processes, and products, which is commonly referred to as data-driven innovation (OECD, 2015[4]). The use of data thereby involved a series of "data-related processes through which value is created with data, including, but not limited to, data creation, collection, validation, verification, storage, curation, enrichment, processing and analysis, access and sharing, and deletion", which the Recommendation refers to as "data value cycle". This process is not a linear value chain but rather a dynamic value cycle that encompasses feedback loops at numerous phases of the value creation process, through which information and insights are extracted from data to make data-driven, human or machine decisions. These decisions might result in more or different data generated and might lead to a new data value cycle (OECD, 2015[4]).

Figure 2 presents the data value cycle applied during both the build (pre-deployment) and use (post-deployment) phases of an AI system. The latter (AI system) is defined as follows:

> An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment. *OECD Recommendation on AI* [OECD/LEGAL/0449][9]

Figure 2. Illustrative, simplified overview of an AI system emphasising the data value cycle



Note: This figure presents only one possible relationship between the development and deployment phases. In many cases, the design and training of the system may continue in downstream uses. For example, deployers of AI systems may fine-tune or continuously train models during operation, which can have significant impacts on the performance and behaviour of the system.
Source: Based on OECD (2024[27]), "Explanatory memorandum on the updated OECD definition of an AI system".

The Figure illustrates how 'datafication'[10] and data collection, including via data access and sharing, enable the AI to perceive its "physical or virtual environments". The collected data, often "big data" depending on the data volume, undergoes preliminary processing (e.g. validation, verification, storage, cleaning, curation, and enrichment including e.g. tagging). This pre-processed data is then used to train AI models, using machine learning and other data analytic techniques to encode information from the input data into the AI model parameters (e.g. weights). The training enables the AI system to generate the aforementioned outputs ("predictions, content, recommendations, or decisions") through inferences based on the AI model parameters. These outputs are applied during the use phase to affect environments, while they are reintegrated into the AI system through feedback mechanisms to refine the AI model parameters during the build phase.

2.1.1. Data, information, and the role of algorithms and software

The relationship between data and information is reflected in the Recommendation's definition of data, which refers to data as "recorded information in structured or unstructured formats, including text, images, sound, and video." The information can be stored on a physical device (i.e., data at rest) or be moving from one location to another such as across the Internet, through a private network, or through an information system (i.e., data in transit). The medium on which data are stored is thereby irrelevant for the definition. In theory, the definition captures also data stored on paper. However, as clarified in the Recommendation: "These general principles and policy guidance are principally aimed at data in digital formats."

As highlighted above, the extraction of information from data[11] relies on its processing and analysis and thus on the access and use of digital resources, including data processing algorithms and software. The use and correct interpretation of data also depends on the possibility to identify the source of data (data provenance), the methodology including the criteria of its selection and collection, as well as the various steps of curation. Furthermore, complex datasets can frequently be correctly interpreted and used only if suitable data processing algorithms and software[12] are accessible and used with the data.

All these requirements imply that any initiative for enhancing access to and sharing of data should also take into consideration the role of algorithms and software. This is reflected in the preamble of the Recommendation, which recognises that: "data management, including creating, collecting, storing, curating, enriching, deleting, providing access to, and sharing data, as well as using data and managing the associated risks, … may involve a wide range of complementary digital resources, including algorithms, software, hardware, and other foundational infrastructures from multiple parties".

The Recommendation then calls on Adherents to "Strive to ensure that data are provided together with any required meta-data, documentation, data models and algorithms in a transparent and timely manner" under its Section 3 on "Fostering effective and responsible data access, sharing, and use across society". In the context of AI, for instance, an important development in this respect has been the introduction of data statements/sheets (McMillan-Major, Bender and Friedman, 2024[28]) and model cards (Mitchell et al., 2019[29]) to report essential information about the characteristics of AI input data and AI models (see section on "Fostering the findability, accessibility, interoperability and reusability of data across organisations, including within and across the public and private sectors").

2.1.2. Data taxonomy

The Recommendation also recognises that there are different types of data; in other words, that data is not a monolithic entity, but rather a heterogeneous asset. The Recommendation explicitly defines "personal data" as "information relating to an identified or identifiable individual (data subject)" in alignment with the *Recommendation concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* [OECD/LEGAL/0188] (the "OECD Privacy Guidelines"). The Recommendation also defines

"metadata" as "recorded structural or descriptive information about the primary data. Metadata can include personal data".

This differentiation is critical since the most appropriate data access and sharing arrangements will depend on the risks associated with the various types of data (see section 2.2. below on "the data openness continuum and its relationship between control, risk and trust").[13] Other types of data that can help further identify appropriate data access and sharing arrangements, and enable addressing data governance in a more differentiated manner, are covered by the data taxonomy including the following three major dimensions (OECD, 2019[2]):

1. Personal data and the degrees of identifiability, noting that a higher degree of identifiability might be associated with higher risks and hence would require data access and sharing arrangements that are appropriately conditioned (defined in the Recommendation as "conditioned data access and sharing arrangements");
2. The different domains of the data, whether personal, public or proprietary (private), can overlap and create complexities, although they are typically subject to different data governance frameworks that can affect each domain differently; and
3. The manner data originates, which reflects the level of awareness and control that data subjects can have about data collected about or from them. This dimension differentiates whether data are *observed,*[14] *volunteered,*[15] *derived,*[16] or *acquired.*[17] Some assert that derived data would for example include AI models (see Box 1).

For instance, from a privacy and data protection angle, personal data usually requires more restrictive access than non-personal, non-proprietary, data in the public domain. On the other hand, industrial data will in many cases be proprietary data which may require more restrictive access than certain non-proprietary, non-personal public-sector data, which can be shared through open data arrangements (as noted above) (OECD, 2019[2]).

Furthermore, what data should be made available to customers or the public may depend on the way that data originated and whether the data is regarded as personal, proprietary, or both. In certain cases, access to highly sensitive (identified) personal or proprietary data might be granted, however only within a restricted digital and/or physical environment to trusted users (e.g., data sandboxes). Furthermore, if appropriately anonymised and aggregated, personal, non-proprietary data might also be provided to the public (OECD, 2019[2]).

The manner data originates can also have a significant impact on the certain rights and obligations. For example, the "Right to Data Portability" (Art. 20) of the European Union (2016[30]) General Data Protection Regulation (GDPR), for instance, only applies to personal data "provided by" the data subject with consent or under contract that is electronically processed. This includes both observed and volunteered data. However, there is consensus that inferred data cannot be subject to this right. ([Former] Article 29 Data Protection Working Party, 2017[31]) Consequentially, as they are inferred data, AI models cannot be subject to Art. 20 of the GDPR.

The most appropriate data access and sharing arrangements may also depend on the complexity of the data, including the degree to which data is unstructured (e.g. text, images, and video without tags), semi-structured (e.g. using tags), or structured (e.g. tables and relational databases). Structured data can vary further in their degree of complexity, often reflected in the data's structural and dimensional properties. Simple data points, such as individual numbers or categorical labels, represent the most basic form, containing only a single piece of information each. As complexity increases, data structures evolve into multidimensional arrays, such as vectors and matrices, with implications on data accessibility, shareability and usability.

Relational databases, for example, reflect more complex attributes and relationships and may be shared through direct downloads of files (using SQL, CSV, or JSON formats), or more commonly via Application

programming interfaces (APIs) as data complexity increases (see Section 2.2.1 on "Factors determining the degree of openness"). At the higher end of the data complexity spectrum lie, for example, data models stored via so called *(data) tensors*, which are multi-dimensional arrays representing multilinear relationships between abstract entities. These more complex data structures can store complex information by capturing their intricate patterns and relationships that simpler data structures cannot. AI models are typically stored in tensors of popular deep learning frameworks like TensorFlow, Keras, PyTorch (Box 1).

Box 1. AI models as data: a data policy perspective

"An AI model is a computational representation of all or part of the external environment of an AI system encompassing, for example, processes, objects, ideas, people and/or interactions that take place in that environment." (OECD, 2022[32]) AI models, particularly those using deep learning techniques like large language models (LLMs), are constructed and refined through a process that encodes information from the input data into their model parameters, thereby learning to perform tasks such as translation, content generation, or image recognition. (OECD, 2023[33]; Lorenz, Perset and Berryhill, 2023[34]).

Model parameters, particularly model weights, are numerical values inferred through training. They represent the model's learned and recorded information about key characteristics from the input data. These parameters are tailored to the chosen model architecture design. For instance, recurrent neural networks (RNNs) or transformers, which may vary in layer types and numbers, are typically used for processing sequential data like text or speech. (OECD, 2023[33]; IWGDPT, 2024[35])

Consequently, sharing e.g. a LLM requires disseminating both its model weights and metadata about its architecture, including its layers and their interconnections. While, there is no consensus on which model components need to be shared for an AI model to be considered "open source" (OECD, forthcoming[16]), frameworks like the *Model Openness Framework* therefore require at a minimum that "the final model parameters and optimizer state (when applicable) must be distributed in an acceptable format whether compressed or uncompressed for usage with popular deep learning frameworks like TensorFlow, Keras, PyTorch or in the framework independent ONNX file format" (White et al., 2024[28]; see also Open Source Initiative, n.d.[29]). These frameworks ensure that the model's structural and operational details are preserved, facilitating ease of use, portability, or explicit control depending on the framework used. (Janapa Reddi, n.d.[36]; Hauser, 2023[37])

Furthermore, since model parameters can be considered a form of data, it has been argued that they are not explicitly covered by standard software licenses. (White et al., 2024[38]; Open Source Initiative, n.d.[39]). This raises critical legal questions about the suitability of open source licenses, such as Apache 2.0, commonly used by producers of foundational models to release their AI models as "open source AI" (Benhamou, 2024[40]). These model parameters may be more appropriately governed by open-data licenses like CDLA-Permissive-2.0 or content licenses such as CC-BY-4.0, which have provisions that take into account specific data governance considerations like privacy[18] and database rights (see Section 3.2.1.2).

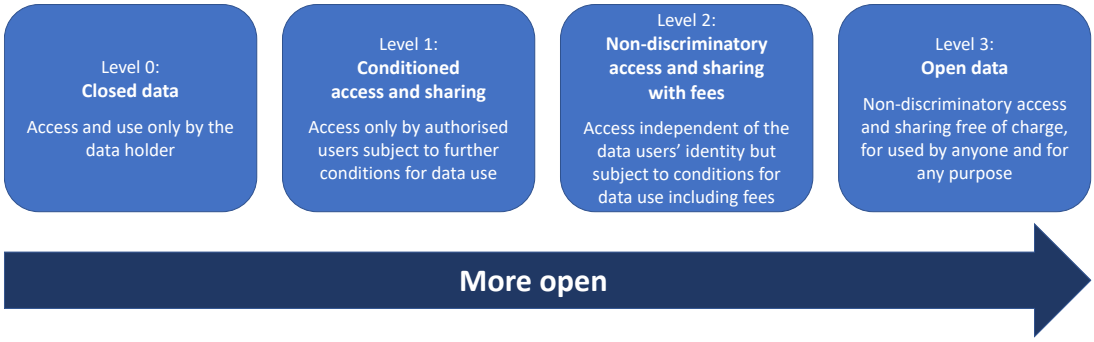
2.2. The data openness continuum and its relation to control, risk and trust

A key characteristic of the Recommendation is its differentiated approach to data access and sharing that leverages the 'data openness continuum', the most important fundamental concept of the Recommendation. Data openness is not a 'binary concept' opposing close to open access to data (open

data). It is rather a continuum of various degrees of openness, ranging from closed access and use only by the data holder (Level 0 in Figure 3), to:

- Level 1: conditioned access and sharing arrangements where data access and sharing *are* subject to "terms that include limitations on the users authorised to access the data (discriminatory arrangements), conditions for data use including the purposes for which the data can be used, and requirements on data access control mechanisms through which data access is granted".
- Level 2: non-discriminatory data access and sharing arrangements, where data can be accessed and shared for fees, but "based on terms that are independent of the data users' identities"; and
- Level 3: open data (arrangements) as the extreme form of data openness, which are "non-discriminatory data access and sharing arrangements, where data is machine readable and can be accessed and shared, free of charge, and used by anyone for any purpose subject, at most, to requirements that preserve integrity, provenance, attribution, and openness". (OECD, 2021[41])

Figure 3. The degrees of data openness



Source: Based on OECD (2015[4]), *Data-Driven Innovation: Big Data for Growth and Well-Being*, https://doi.org/10.1787/9789264229358-en.

These various degrees of openness are reflected in the Recommendation through the concepts of "data access and sharing arrangements" that are nested in one another as defined in the Recommendation. That "conditioned access and sharing arrangements" include "non-discriminatory data access and sharing arrangements" as a specific type, which in turn include "open data arrangements" as the more specific type of non-discriminatory arrangement. The main difference in the first relationship is the discrimination (or non-discrimination) based on the users' identity, while in the case of open data, access and sharing is free of charge. Examples of "conditioned access and sharing arrangements" that are discriminatory include data portability arrangements (OECD, 2021[17]; see Box 2).

Box 2. Selected examples of different types of arrangements

Conditioned, discriminatory data access and sharing arrangements (levels 1) are pertinent in cases where data is considered too confidential to be shared openly with the public (as open data) or where there are legitimate (commercial and non-commercial) interests opposing such open sharing. Examples include the Genomic Data Commons (GDC) of the National Cancer Institute (NCI) (OECD, 2019[2]). It provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine. While some data are provided via open data arrangements, many (including individually identifiable data such as low-level genomic sequencing data) are provided with controlled access, requiring authorisation and authentication.

Box 2. Selected examples of different types of arrangements (cont.)

Data portability arrangements (level 1) are a specific type of conditioned data access and sharing arrangements, whereby a natural or legal person can request that a data holder transfer to the person, or to a specific third party, data concerning that person in a structured, commonly used, and machine-readable format on an ad hoc or continuous basis (OECD, 2021[42]; Reimsbach-Kounatze and Molnar, 2024[43]). Data portability initiatives aim to empower users, enhance their informational self-determination, and/or foster competition and innovation. As evidenced by the growing body of legislation, data portability has attracted policy makers' attention in recent years, but there are differences across initiatives. For example, while privacy and data protection frameworks like the GDPR (including in the EU and the United Kingdom) primarily focus on empowering individuals and enhancing informational self-determination, other legal regimes such as Australia's Consumer Data Right (CDR), the EU's Digital Markets Act (DMA), and Digital Services Act (DSA), along with sector-specific regulations like open banking place greater emphasis on fostering consumer choice, competition and innovation.

Open data arrangements (level 3 in Figure 1) "refers to non-discriminatory data access and sharing arrangements, where data is machine readable and can be accessed and shared, free of charge, and used by anyone for any purpose subject, at most, to requirements that preserve integrity, provenance, attribution, and openness" (OECD, 2021[41]). They include most prominently certain open government data on data.gov (United States), data.gov.uk (United Kingdom), data.gov.fr (France), or data.go.jp (Japan).[19] In Finland, for example, the project on opening up and using public data promotes the wider and more effective use of public sector data throughout society in decision-making, business, research and civic engagement. The aim of the project is to draw up a proposal concerning the strategic objectives for opening up and using public sector data that can be implemented by government entities, and to prepare and introduce measures to promote the opening up and use of public sector data.[20] Open data arrangements will typically rely on open licences, such as the Public Domain Dedication and Licence (PDDL), Open Data Commons Attribution License (ODC-BY), Open Data Commons Open Database License (ODbL), Community Data License Agreement (e.g. CDLA-Permissive-2.0) and various Creative Commons Licences. (Leigh Dodds, 2013[44]; Open Knowledge Foundation, n.d.[45]) These licences vary in terms of the levels of restrictions they impose, ranging from complete public domain dedication (CC0 and PDDL) to attribution requirements (CC-BY) and sharing adaptations under the same terms (CC-BY-SA). Each licence serves to promote data openness, allowing users to select the one that best suits their specific needs. The ODI, for example, recommend using a Creative Commons 4.0 licence for open data, which would also be applicable to some AI models (see Box 16).

The concept of data openness is fundamental for data governance across all policy domains. For example, arrangements governing cross-border data flows span a continuum that can be distinguished by four categories of approaches to cross-border data transfers. These include from the most restrictive to the most open: data flows conditional on ad hoc authorisation; conditional on safeguards; ex-post accountability; and no regulation (Casalini and López González, 2019[46]; OECD, 2022[25]).

Another example is the openness of AI models, which also spans a continuum from fully closed AI models to fully open AI models, contrary to some misconceptions about "open source AI" (OECD, forthcoming[16]). Many have criticised the reliance on single features like access or licensing to declare models open or not, and warned about consequential regulatory uncertainties (Castro, 2024[47]), and the risk of "open washing" (Benhamou, 2024[40]; Liesenfeld and Dingemanse, 2024[48]; White et al., 2024[38]) Experts such as Liesenfeld and Dingemanse (2024[48]) have therefore argued that openness of AI models is "necessarily composite (consisting of multiple elements) and gradient (coming in degrees)", and this is true for data

openness in general (see next sections). Solaiman (2023[49]) provides a comparison of various open AI initiatives bases on the degree of openness of their respective AI models.[21]

2.2.1. Factors determining the degree of openness

The degree of openness of a data access and sharing arrangement is determined by the arrangement's applicable technical, financial, legal, or organisational access and use requirements. These requirements will need to be specified on a case-by-case basis, depending in particular on the risks associated with data access and sharing including for example the sensitivity of the data. Data access control mechanisms thereby play an important role. The Recommendation defines these mechanisms as "technical and organisational measures that enable safe and secure access to data by approved users including data subjects, within and across organisational borders, protect the rights and interests of stakeholders, and comply with applicable legal and regulatory frameworks." (see Box 3)

Box 3. Selected examples of data access control mechanisms

Application programming interfaces (APIs) are one of the most prominent technical measures used as data access control mechanisms. An API enable service providers to make their digital resources (e.g., data and software) available over the Internet. APIs thus enable data linkage and the smooth interoperability of the different actors, their technologies, and services, particularly through the use of cloud computing. A key advantage of an API is that it enables a software application 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 requiring the identity of the API user, the scale and scope of the data used (including over time), and even the extent to which the information derived from the data could reveal sensitive / personal information. (OECD, 2019[2])

Trusted data intermediaries (TDIs) are also commonly used as organisational measures to enable safe and secure access to data, often in combination with APIs. In the case of the Genomic Data Commons (GDC) of the National Cancer Institute (NCI) presented in Box 1, access is granted by programme-specific Data Access Committees (DACs). The DACs review, approve or disapprove all requests from the research community for data access. Decisions to grant access are made based on whether the request conforms to the specifications within the NIH Genomic Data Sharing Policy and program specific requirements or procedures (if any). All uses proposed for controlled-access data must be consistent with the data use limitations for the data set as indicated by the submitting institution and identified on the public website for database of Genotypes and Phenotypes (dbGaP). DACs also review and approve or disapprove all requests for access to dbGaP data for programmatic oversight by NIH employees.

Legally enforceable requirements can also influence or determine the degree of data openness. Examples include privacy and data protection frameworks as well as IPR frameworks, including in particular those related to copyright, and the protection of trade secrets. Increasingly, stakeholders have also come to rely on contracts as key legal vehicles for determining rights and obligations related to data access and use. The flexibility inherent to contracts gives stakeholders the freedom to construct well-suited arrangements that reflect their data sharing requirements.

Finally, financial requirements, including pricing and licencing agreements, also play an important role in determining the degree of openness of data. Pricing can be one of the most challenging factors, because optimal pricing can be hard to determine. Pricing is a key distinguishing factor between open data arrangements (where data access is provided "free of charge") and non-discriminatory data access and sharing arrangements, where access is provided "free of charge or for fees".

2.2.2. The optimal level of data openness

While the desired level of data openness is driven by the potential benefits and value that data access and sharing can create, the optimal level of data openness will vary on a case-by-case basis. It will depend on the risks associated with data access and sharing. Risks include confidentiality and privacy breaches and the violation of other legitimate private and public interests such as commercial interests, intellectual property rights, national security and special interests such as threats to endangered species. Privacy and IPRs and other legitimate commercial and non-commercial interests need to be protected, otherwise incentives to contribute data and to invest in data-driven innovation may be undermined, in addition to the risks of direct and indirect harm to right holders, including data subjects. The optimal level may also depend on risk factors such as poor data quality, missing metadata, or inadequate skills of data users that lead to unexpected biases. Where these risks are high, there will typically be a higher need for control and protection mechanisms and thus a tendency for less data openness. In other words: the protection of privacy, IPR and other rights and interests will in many cases justify more restricted arrangements to data access and sharing, such as in particular conditioned data access and sharing arrangements. (see Box 1)

2.2.3. Risk management

To further facilitate, encourage and enhance data access and sharing it is essential that stakeholders assess the level of risk to better balance the benefits of enhancing data openness with these risks, whilst taking into account legitimate private, national and public interests (OECD, 2019[2]). Where the level of risks is perceived as high, trust will tend to be low, unless trust-enhancing measures including access control mechanisms are put in place to reduce risks. This is also why perceptions of inadequate treatment of risks can lead to mistrust, which in turn may lead to outcomes such as unproportioned restrictions to data access and sharing. The Recommendation therefore emphasises the essential role of trust for enabling data access and sharing at multiple occasions. In particular, Section 1 of the Recommendation focusses on “Reinforcing Trust across the Data Ecosystem” (see Section 3 on “Principles to enhance access to and sharing of data, and implementation examples”).

In this regard, a risk management approach to address risks while enhancing data access and sharing is needed as called for by the Recommendation (see next section on “Measures to enhance access to and sharing of data: Applicability of the Recommendation”). Risk management highlights that risk is not a binary concept and there is always certain level of risk with carrying out an activity. Risk management aims to reduce the risk level that is acceptable in light of the context and potential benefits (OECD, 2015[50]; OECD, 2019[2]).

Nevertheless, a risk management approach remains challenging to implement for organisations, especially where the rights of third parties is involved (for instance in the cases of IPRs of organisations and individuals; privacy rights of individuals) (OECD, 2019[2]). The *OECD Privacy Guidelines* as well as the *OECD Recommendation on Digital Security Risk Management* [OECD/LEGAL/0479] provide a set of operational principles that can help inform the adoption of a risk management approach (see Box 4) for certain data types and contexts.

Additionally, the work of the OECD on AI classification (OECD, 2022[32]) and accountability in AI (OECD, 2023[51]) provides a detailed framework for identifying and managing risks throughout the AI systems’ lifecycle based on the following four important steps: (1) Defining the scope, context, actors and criteria; (2) Assessing the risks at individual, aggregate, and societal levels; (3) Treating risks in ways commensurate to cease, prevent or mitigate adverse impacts; and (4) Governing the risk management process as an iterative process.

Box 4. OECD legal instruments calling for a risk management approach

The OECD Recommendation concerning Access to Research Data from Public Funding [OECD/LEGAL/0347], in particular in its first pillar, "Data governance for Trust", is recommending that Adherents "Take steps to transparently manage risks posed by enhancing access to sensitive categories of research data and other research-relevant digital objects from public funding, including personal data, by applying specific risk mitigation measures, as well as providing for a 'right to know' in cases of digital security incidents affecting the rights and interests of stakeholders", and "Clarify roles and responsibilities of researchers and other staff responsible for data access, so as to promote awareness and a culture of confidence and avoid undue risk averseness."

The OECD Privacy Guidelines, and in particular Part three on "Implementing Accountability", recommends that "A data controller should … have in place a privacy management programme that … provides for appropriate safeguards based on privacy risk assessment". These privacy management programmes need to be implemented in such a way "that is tailored to the structure, scale, volume and sensitivity of [data] operations and that provides appropriate safeguards based on privacy risk assessment including plans for responding to inquiries and incidents" (OECD, 2021[52]). Data controller should therefore be prepared to demonstrate their privacy management programme and provide notice, as appropriate, to authorities and data subjects where there has been a significant security breach affecting personal data under its control irrespective of the location of the data. The supplementary *Explanatory Memorandum to the OECD Privacy Guidelines* provides further guidance on privacy management programmes, for instance by listing examples of "appropriate safeguards" a data controller can put in place (e.g., contractual provisions, employee training and education, and audits) and underpinning the importance of risk assessments, which may be achieved through a privacy impact assessment (OECD, 2013[53]).

The OECD Recommendation on Digital Security Risk Management provides a set of general and operational principles that apply to digital security risk management in data governance. The operational principles include in particular: (i) the principles on "risk assessment and treatment cycle" that emphasises that digital security risk assessment should be carried out "as an ongoing systematic and cyclical process" as well as the (ii) the principle on "security measures", according to which the risk should be treated "on the basis of the risk assessment, in order to reduce it to an acceptable level relative to the economic and social benefits expected from those activities while taking into account the potential impact on the legitimate interests of others, and legal requirements."

2.3. The data ecosystem and the overlapping contributions and interests of stakeholders

The Recommendation recognises that data and the value derived from their use are often (co-)created as a result of the interactions and contributions of various parties, in some cases even without their awareness. This encompasses cases where data from multiple sources are linked across organisational borders, as well as cases, where users (businesses and consumers) interact with a digital product (service or good) for instance a digital government service, a portable smart health device, or a social networking service. These products generally collect data inserted by users or data generated by the observed activities of users. Consequently, the Recommendation recognises that users can be (co-)producers of data.

The sum of all these interactions and contributions are referred to in the Recommendation as "data ecosystem". More specifically the 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".

2.3.1. Key actors

The Recommendation recognises that many stakeholders may be involved in the data ecosystem. In addition to the "data subject", which is indirectly defined as an "individual" who is "identified or identifiable" through personal data, the Recommendation explicitly defines "data holders", "data producers" and "data intermediaries" as "relevant stakeholders".[22]

> *'Data holders' refers to organisations or individuals who, according to applicable laws or regulations, are competent to decide on granting access to or sharing data under their control, regardless of whether or not such data are managed by that organisation or individual or by an agent on their behalf.*

Data holders are thus not only in control of data, but must also have the rights to store, process and share the data. In the case of personal data, the data holder is referred to as "data controller" as defined by the OECD Privacy Guidelines (OECD, 2013[53]).[23]

> *'Data producers' refers to organisations or individuals that create, co-create, generate, or co-generate data, including as a by-product of their social and economic activities, and can therefore be considered a primary data source.*

The Recommendation recognises that "data producer" is an inclusive concept. For example, it includes data subjects where they contribute to the (co-)creation or (co-)generation of personal data relating to themselves, in which case they enjoy relevant data protection and privacy rights. Data producers may also hold proprietary rights or intellectual property rights derived from the (co-)creation or (co-)generation of data. Others in the ecosystem may need to protect the rights of data producers, including data subjects, which may override contractual provisions. Contrary to other rights that data producers may possess, the privacy and data protection rights to which data subjects are entitled may not be waived, regardless of any contractual agreements they enter.

> *'Data intermediaries' refers to service providers that facilitate data access and sharing under commercial or non-commercial agreements between data holders, data producers, and/or users. Data holders and trusted third parties can act as data intermediaries.*

Data intermediaries may thus provide additional added value services such as data processing services (including data aggregation), payment and clearing services as well as legal services including the provision of standard licence and certification schemes, irrespective of whether the data is under their control or not. They may also act as *trusted third parties* to foster trust among the different stakeholders, as *data stewards* where they manage data to ensure its long-term quality, security, and accessibility, and/or *data aggregators*, which aggregates data from various sources, curates and structures the data in specific databases. National statistical agencies can for example act as trusted data intermediaries.

2.3.2. Conflicting rights and interests, and 'data ownership'

All stakeholders in the data ecosystem, including individuals (e.g., workers, citizens, consumers) and organisations in the private and public sector, may have overlapping and frequently conflicting rights and interests over the same data and the complementary resources in their different capacities as e.g., data holders, data producers and data subjects. In particular, various legal frameworks might apply differently to stakeholders, within and across countries, including privacy and intellectual property protection frameworks in addition to other access and control rights (provided by e.g., contract, cyber-criminal, and

competition law). These frameworks include essential provisions safeguarding the rights and values that are needed for trust in data access and sharing (such as for the protection of privacy and property rights).

However, the different legal frameworks do not preclude each other, in fact they overlap. However, (Determann, 2018[54]) notes that the "intricate net of existing legal frameworks" combined with the involvement of multiple parties in the creation of data (and its value) may explain current legal uncertainties, in particular those related to 'data ownership'. The challenge is exacerbated where data are created, and expected to be accessed and shared, across national borders.

The Recommendation does not address the question of 'data ownership' *per se* as the concept itself remains controversial and is a source of confusion. This is largely due to the fact that it is used in different contexts with a different meaning (OECD, 2019[2]).[24] The Recommendation, however, does address the issue of control over data, which is a key aspect that is often thought of when referring to the concept of 'data ownership'. For example, in addition to its call to empower individuals and social groups, the Recommendation calls on Adherents to ensure data are as open as possible to maximise their benefits and as closed as necessary to protect legitimate public and private interests, including interests related to national security, law enforcement, privacy and personal data protection, and intellectual property rights…" under its Section 1 on "Reinforcing trust across the data ecosystem" (discussed in the Section 3 on "Principles to enhance access to and sharing of data, and implementation examples").

2.3.3. Interoperability and the interplay between data governance frameworks

2.3.3.1. Technical and legal interoperability

Enhancing access to and sharing of data is sometimes motivated by the objective to improve the interoperability of systems. There is almost always a need for better interoperability, where data are to be shared and re-used. However, interoperability can have different meanings depending on the context. Generally, it is understood in terms of "technical interoperability"' as compatibility, such that systems can work together or "interoperate" in a way that allows for seamless or real-time exchanges, updates or transfers of information or data. In relation to cloud computing, for instance, the International Standards Organization (ISO) defines "interoperability" as the ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged (ISO, 2017[55]).[25]

Commonly used machine-readable formats are not enough to guarantee technical interoperability (OECD, 2019[2]). These 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." However, they do not guarantee *semantic interoperability*, defined as "transferring data to a target such that the meaning of the data model is understood within the context of a subject area by the target." The latter relies on the availability of metadata and data documentation. Both syntactic and semantic interoperability are needed for data interoperability, which is the ability of diverse datasets to be linked, merged or aggregated in meaningful ways (Kush et al., 2020[56]; Network of the National Library of Medicine [United States], n.d.[57]).

The concept of interoperability can also have a legal dimension ("legal interoperability"). The concept of legal interoperability has gained prominence in the context of privacy and data protection, as one of the potential means to address the fragmentation of cumulatively applicable legal regimes, which is considered as one of the main challenges to transborder data flows (OECD, 2021[52]). This is reflected in the *OECD Privacy Guidelines*, under Part Six on "International Co-Operation and Interoperability", which recommends that "Member countries should encourage and support the development of international arrangements that promote interoperability among privacy frameworks that give practical effect to these Guidelines."

However, there exists no internationally agreed-upon definition of what legal interoperability is. In the context of privacy and data protection, consequently, how to achieve legal interoperability among privacy frameworks in practice is unclear. The *OECD Privacy Guidelines* themselves or the Supplementary

Explanatory Memorandum do not define privacy interoperability, although the latter provides examples of a range of initiatives undertaken to bring together different approaches to interoperability among privacy frameworks (OECD, 2013[53]). Robinson, Kizawa and Ronchi (2021[58]) understand "privacy interoperability" "operationally as the ability of different privacy and data protection regimes, or legal frameworks, to work together at multiple levels through policy and practical arrangements and thereby bridge any differences in approaches and systems of privacy and personal data protection to facilitate transborder flows of personal data". In recent years, G7 and OECD governments have reiterated their commitments to "build upon commonalities, complementarities and elements of convergence between existing regulatory approaches and instruments enabling data to flow with trust in order to foster future interoperability".

The Recommendation addresses issues related to interoperability, covering both, the technical aspect of interoperability (e.g., the need for machine readability, common standards and other interoperable technical specifications, including semantic interoperability), as well as the legal aspects of interoperability (e.g., the ability of legal frameworks to work together, based on commonalities, complementarities and convergence).

First, it recognises the importance of technical interoperability that "often depends on machine-readability and interoperable specifications including common licensing arrangements, standards, and metadata that enable findability, accessibility, interoperability, reusability, and the correct interpretation and analysis of data".

But most importantly, the Recommendation includes provisions on technical and legal interoperability in Section 3 on "Fostering effective and responsible data access, sharing, and use across society". In particular, the Recommendation refers to the FAIR (Findability, Accessibility, Interoperability and Re-use) principles (Wilkinson et al., 2016[59]) and the need to promote the development and adoption of interoperable specifications for effective data access, sharing and use, including common standards for data formats and models as well as open source implementations. (See Section 3 on "Principles to enhance access to and sharing of data, and implementation examples").

In so doing, the Recommendation opens new pathways for innovative policy considerations in the governance of data including AI models. For example, it raises the question about what interoperability could mean when applied not only to AI systems but to AI models, and the policy considerations that result from this (Box 5).

Box 5. AI and data interoperability

Interoperability refers to the ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged (ISO, 2017[55]). This definition extends to AI systems, however, with added complexity as information can be exchanged and reused across the following three dimensions of the AI system classification framework (OECD, 2022[32]): (i) input data, (ii) AI models, and (iii) output (data).

AI system interoperability can thus be understood as the ability of two or more AI systems to exchange information, and to mutually use the information that has been exchanged, at the input, output and model level. In his "thoughts on AI interoperability", Cerf (2024[60]) addresses AI system interoperability at both the input and output levels and, most importantly, the interactions of AI systems between these.

- *AI system interoperability at the input level* denotes the ability to use, share and interchange input data between AI systems. This requires *data interoperability* at the input level, which is typically one of the primary objectives of data pre-processing. Cerf (2024[60]) refers in this context to "federated learning, in which multiple ML systems independently ingest training content".
- *AI system interoperability at the output level* is akin to that at the input level, as both rely on data interoperability. However, output level interoperability carries broader implications because it allows to share and interchange AI (sub-)systems. This interoperability can for example help overcome vendor lock-in effects or can foster the integration and interaction of diverse AI systems into a broader and open AI ecosystem. But this requires data interoperability between the input and output levels across AI systems. Cerf (2024[60]), for example, refers in this context to "a more ambitious notion [that] might involve cooperative interaction among ML systems (not only LLMs)".
- *AI system interoperability at the model level* denotes the ability to use, share and interchange AI models between AI systems. It implies the ability of AI models to be trained in one environment or framework (like TensorFlow or PyTorch) and then deployed in another without needing extensive reconfiguration. Therefore, openness of AI model is a condition for interoperability at this level. Furthermore, such interoperability benefit from *AI model interoperability,* which is to be distinguished from AI system interoperability at the model level.

In analogy to the definition of data interoperability*, AI model interoperability* can be understood as the ability of diverse AI models to be linked, merged or aggregated in meaningful ways (as opposed to their ability to be shared and interchanged). Merging AI models seems to remain challenging beyond the technique of federated learning, where multiple model parameters are trained on local devices to be merged into a global AI model. AI models have been also commonly combined using "ensemble learning" techniques to boost overall performance. (Naderalvojoud and Hernandez-Boussard, 2023[61])

Mixture-of-Experts (MoE), a type of ensemble learning technique that involves a system of multiple specialised AI models ("experts"), has gained a lot of attention with the rise of MoE system that perform on par with (or better than) up to 10-time larger LLMs. For example, the Mixtral 8x7B, released by French company Mistral AI ("with open weights" under the Apache 2.0 license), uses a MoE system with a total of 46.7B parameters, yet outperforming for example "Llama 2 70B on mathematics, code generation, and multilingual benchmarks". (Jiang et al., 2024[62])

Currently, the integration of AI models is primarily limited to those that are self-developed. Enhancing open source AI models with comprehensive metadata along with detailed technical documentation including as well data and model cards, could facilitate experimental combinations of different models.

2.3.3.2. Interplay between general and sector-specific data governance frameworks

The successful implementation of the Recommendation also relies on the ability to articulate baseline regulation and sectoral regulation applicable across sectors and data types. General cross-sectoral approaches include privacy and data protection frameworks such as the European Union's GDPR, the GDPR of the United Kingdom, the California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA) as well as other cross-sectoral data governance frameworks such as the DMA and the EU Data Act. Conversely, sectoral approaches are most frequently used for critical infrastructure and cover e.g. financial services (open banking and the EU Second Payment Service Directive of November 2015 [PSD2]), transportation and mobility (the EU Regulation on Motor Vehicles of May 2018), energy (e.g. EU Electricity Directive of 2019) and health care (e.g. HIPAA).

Australia's Consumer Data Right (CDR) is also mainly used for critical infrastructures, although it can be best classified as a hybrid approach in this respect (see Box 4). The CDR is implemented at a sectoral level based on requirements defined with market participants (primarily in infrastructural sectors such as energy, and banking). A 2022 Strategic Assessment of the CDR identified "open finance" as the likely next priority area to expand the CDR, which could include datasets from general insurance, superannuation, merchant acquiring and non-bank lending service providers (Australian Government, 2022[63]). In 2023, the Government announced its focus over the next two years is to mature the CDR in banking and energy and to expand into non-bank lending.

Horizontal data governance frameworks in the past have focussed on a specific type of data, mainly personal data. In an analysis of the legal framework on data portability in the EU, CERRE (Streel, Kramer and Senellart, 2020[64]) shows that horizontal data portability initiatives focus either on personal data or non-personal data with competition law being the exception in many respects. On the other hand, sector-specific data portability initiatives usually cover a range of data types. This has changed with the introduction of the EU DMA as well as the EU Data Act which incorporate data portability provisions that complement Art. 20 GDPR on the right to data portability (see

Table 1).

Table 1. EU legal frameworks for data portability and sharing

	Personal data	Non-personal data	All data
Horizontal	Art. 20 GDPR – Right to data portability	Art. 16 Digital Content Directive – Obligations in the event of termination	Art. 6(9-11) DMA – Obligations for gatekeepers to provide data portability
		Art. 6 Free Flow of Non-personal Data Regulation – Porting of data	Chapter II (Data Act – B2C and B2B data sharing (only covering the IoT sector)
Sector-specific	Art. 66(4) and 67(3) – Second Payment Service Directive (PSD2)		
	Art. 61 Regulation on Motor Vehicles (2018) – Access to vehicle diagnostic, repair and maintenance data		
	Art. 23(2) New Electricity Directive		

Both cross-sectoral and sector-specific approaches to data governance have respective strengths and weaknesses: while sector-specific approaches tend to better address the specific legal, organisational and technical requirements of individual sectors, given that requirements for data transfers may vary by both data type and sector, cross-sectoral approaches may facilitate data sharing both across sectors and within sectors more effectively. This becomes possible as certain industries may not have sufficient incentives to develop a user-driven, data sharing framework on their own. Furthermore sector-specific approaches may create asymmetries. How to ensure the proper interplay between horizontal and sector-specific approaches to data governance is thus critical to take advantage of the strengths of both approaches. Australia's CDR is an example how to reconcile both approaches (Box 6).

The Recommendation acknowledges that national data strategies (NDS) or other scalable, whole-of-government approaches may be needed to address the interplay between cross-sectoral and sector-specific data governance approaches. As called for by the Recommendation, "a strategic whole-of-government approach to data access and sharing" should "integrate[s] cross-cutting economic, social, cultural, technical, and legal governance issues". (See Section 3 on "Principles to enhance access to and sharing of data, and implementation examples").

Box 6. Facilitating the interplay between cross-sectoral and sector-specific data governance frameworks – the case of Australia's Consumer Data Right (CDR)

In August 2019, the Australian Parliament passed legislation introducing a Consumer Data Right (CDR), enabling consumers in designated sectors of the Australian economy (a "CDR consumer") to have certain information disclosed to them or to accredited persons. [Treasury Laws Amendment (Consumer Data Right) Act 2019 (Cth)]. The CDR is a cross-sectoral regime but with sector-specific implementations. That is, the CDR was intentionally drafted to cover all consumer data across the economy, but its implementation is staged by sector. The legislation confers on the responsible Minister an ongoing power to designate sectors of the Australian economy that are subject to the CDR [s 56AC]. The CDR has commenced in the banking and energy sectors and is planned to extend to other sectors, including non-bank lending.

The Department of the Treasury is responsible for developing the rules for the CDR, with the Minister the decision maker. The Australian Competition and Consumer Commission (ACCC) is the delivery agency for accreditation providing the technology that supports data sharing and the enforcement of CDR rules. The Office of the Australian Information Commissioner (OAIC) also has a role in enforcing CDR rules where "CDR data" is also personal data, and the two bodies have agreed a joint compliance and enforcement policy. Both agencies work closely with the Data Standards Body that develops the data standards for data sharing. The latter plays a critical role for establishing standards to facilitate interoperability across sectors.

Source: (OECD, 2021[42]; OECD, 2020[65])

3. Principles to enhance access to and sharing of data, and implementation examples

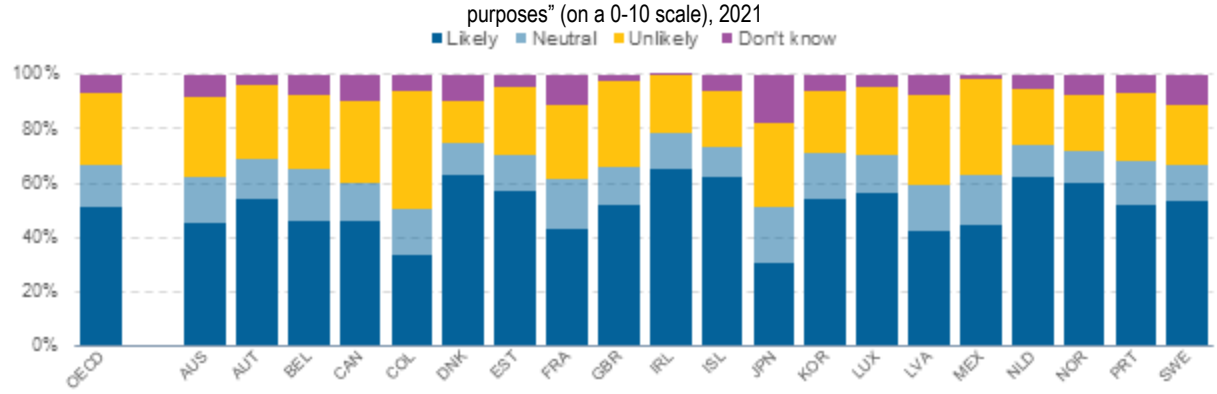
Each of the following sections highlight the specific issues that the Recommendation addresses and how its various provisions address these issues, including by highlighting concrete implementation examples in dedicated boxes.

3.1. Reinforcing trust across the data ecosystem

Trust plays an essential role in data access, sharing and re-use across organisations, sectors, and jurisdictions. Trust can be abused or eroded over time and restoring it can be challenging. For instance, results from the OECD Survey on the Drivers of Trust in Public Institutions show that "on average across countries, 51.1% of respondents say that, if they were to share their personal data with a public agency/office, it is likely that the data would be exclusively used for legitimate purposes" (OECD, 2022[66]) (see Figure).”

Figure 4. Drivers of Trust in Public Institutions: Half of respondents, on average, trust their government to use their personal data for legitimate purposes

Share of respondents reporting different levels of perceived likelihood that their government would use personal data exclusively for "legitimate purposes" (on a 0-10 scale), 2021



Note: Figure presents the within-country distributions of responses to the question "If you share your personal data with a [public agency/office], how likely or unlikely do you think it is that it would be exclusively used for legitimate purposes?". The "likely" proportion is the aggregation of responses from 6-10 on the scale; "neutral" is equal to a response of 5; "unlikely" is the aggregation of responses from 1-4; and "Don't know" was a separate answer choice. Finland and New Zealand are excluded from this figure as data were not available. "OECD" presents the unweighted average across countries. For more detailed information please find the survey method document at http://oe.cd/trust.
Source: OECD (2022), Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions, Building Trust in Public Institutions, OECD Publishing, Paris, https://doi.org/10.1787/b407f99c-en with data from OECD Trust Survey (http://oe.cd/trust)

Along these lines, to help sustain and reinforce trust across the data ecosystem, the Recommendation addresses the following three themes, which are discussed in further detail in dedicated subsections. These include: (i) measures to empower and pro-actively engage all relevant stakeholders alongside broader efforts to increase the trustworthiness of the data ecosystem; (ii) the adoption of a strategic whole-of-government approach to ensure that data access and sharing arrangements effectively and efficiently meet specific societal, policy, and legal objectives that are in the public interest; and (iii) measures to maximise the benefits of data access and sharing, while protecting individuals' and organisations' rights.

3.1.1. Empowering and pro-actively engaging all relevant stakeholders alongside broader efforts to increase the trustworthiness of the data ecosystem

3.1.1.1. Multi-stakeholder engagement and participation

Evidence presented in the joint work of the three committees (DPC, CSTP and PGC) shows that engaging with communities of stakeholders to understand their respective interests and concerns is a major success factor for strengthening buy-in and compliance from all stakeholders and for building trust, in line with ethical values and norms such as fairness, human dignity, autonomy, self-determination.

Pro-actively engaging all relevant stakeholders can help better identify and make explicit the various interests of stakeholders. This is critical because, as highlighted above (in Section 2.3), multiple stakeholders are involved in the collection, control, and use of data at different stages of the data value cycle, providing the rationale for each of them to be claiming rights and interests with respect to those data.

Empowerment and pro-active engagement can also help implement a risk-management approach as discussed in Section 2.1 as their can help policy makers and stakeholders better identify and define responsibilities and the acceptable risk levels across the data ecosystem (Frischmann, Madison and Strandburg, 2014[67]; OECD, 2016[68]; OECD, 2019[2]; OECD, 2021[69]) (OECD, 2020[22]). The Recommendation recommends that Adherents should:

> *Promote inclusive representation of and engage relevant stakeholders in the data ecosystem – including vulnerable, underrepresented, or marginalised groups – in open and inclusive consultation processes during the design, implementation, and monitoring of data governance frameworks (III.a). [They should also] work together with key stakeholders to clearly define the purpose of [data access and sharing] arrangements and identify data relevant to these purposes, taking into account their benefits, costs, and possible risks (IV.a).*

To be effective, multi-stakeholder engagement and dialogues require establishing whole-of-society, open and inclusive processes where:

- All relevant stakeholders are represented; their roles, responsibilities and modalities of participation are identified; and their various interests are recognised and made transparent; and
- Appropriate mechanisms are put in place to ensure that all stakeholders' interests are taken into account in the design, implementation, and monitoring of data governance frameworks; stakeholder feedback is given appropriate attention and consideration; and that a rationale is given where feedback is not taken on board (OECD, 2019[21]).

Governments have implemented different approaches to engage with their citizens on data governance issues. Examples are presented in Box 7.

Box 7. Selected examples for pro-actively engaging stakeholders when developing and implementing data governance policies and arrangements

Australia has used its experience in creating consensus around mining projects which share some commonalities with the data issue: great potential benefits combined with potentially very serious risks. When consulted, the public does accept that those risks do exist, and if they are made familiar with those risks, as well as the risk management practices put in place, this can be made acceptable to a critical number of stakeholders. In order to implement this, the Australian government implemented three classes of activities: 1) Inform the public early on about data initiatives, their objectives, the associated risks and the risk management and mitigation procedures put in place; 2) Ensure smooth and easy compliance with those mitigation procedures by assembling the regulations for data owners into easy to use guidance documents; 3) Engage with the community of stakeholders who have interests in those data, to understand their concerns and respectfully address those concerns.

The Australian Government released the Public Data Policy Statement in 2015, laying out expectations about data sharing and the value it can create, in alignment with OECD standards in the area of privacy protection. Australia switched to an "open by default setting" for data, followed by a dramatic increase in the number of datasets now publicly available.[26] In May 2023 the Government released an initial Data and Digital Government Strategy which includes the Government's commitment to making non-sensitive data open by default and having formal controls for sharing of more sensitive data under agreements to ensure privacy, security and ethical use.

In **Canada**, the Canadian Institutes of Health Research, the National Science and Engineering Research Council and the Social Science and Humanities Research Council organised in-person as well as online consultations on their draft Tri-Agency Research Data Management (RDM) policy in 2017 and 2018. More than 130 written submissions were delivered to the online consultation alone from a broad range of stakeholders including universities, colleges and polytechnics; libraries; government agencies and departments; organisations that raise awareness and build supports for RDM; and academic associations and individual researchers from various health, natural sciences, engineering, social sciences and humanities research disciplines.

Greece set up an informal working group, which is co-ordinated by the Ministry of Digital Governance and operates as a think tank. The working group discusses ideas regarding data, open data policies and data governance, as well as ideas for new legislative acts under the scope of transparency, proportionality and integrity. Besides the Ministry of Digital Governance, this working group involves stakeholders from civil society, other public sector and private sector entities.

Italy promotes inclusive representation of and engage relevant stakeholders in the data ecosystem. Citizens and stakeholders may participate in the innovation process through a specific site that has been set up to discuss on digital public services. Stakeholders are also involved in the process through public consultations published on dedicated sites.

In **Israel**, the DatA-IL Innovation Community is a collaborative effort involving the Ministry of Economy and Industry, the Israel Innovation Authority, the Israel National Digital Agency, and Social Finance Israel (SFI). The community aims to connect people, public data, and AI to foster an ecosystem where data and AI can be used as significant growth drivers for both the public and private sectors. DatA-IL's mission is to leverage data and AI to create social and public benefits through connections, collaborations, knowledge-sharing, and partnerships among startups, companies, public sector entities, NGOs, and academia.[27]

Box 7. Selected examples for pro-actively engaging stakeholders when developing and implementing data governance policies and arrangements (cont.)

Türkiye regularly conducts meetings with the industry associations of certain sectors to increase the effectiveness and functionality of benchmarking studies. In this way, benchmarking studies are carried out more transparently and possible mistakes can be prevented in line with the opinions and suggestions of the sector stakeholders. Specific initiatives across the public sectors also exist. For example, the Turkish Central Bank has initiated a project called "Data Inventory System" which aims to increase the awareness about its databases among the researchers, developers and decision-makers. Furthermore, in 2021, Türkiye partnered with World Bank and conducted two workshops focussing on the challenges of exchanging data across the private sector and scaling up Industry 4.0 applications and the role government should play in developing data spaces.[28]

The United Kingdom has undertaken stakeholder engagement throughout the National Data Strategy's implementation, bringing in perspectives from across industry, academia, civil society and wider public attitudes. The NDS team has worked across Whitehall to refine the scope and direction of the strategy. To date the team has engaged with hundreds of policy officials across government and over 250 external organisations, hosting over 20 roundtables. Additionally, the National Data Strategy Forum was established as a structured programme of engagement designed to ensure that a diverse range of perspectives beyond government and the public sector continue to inform the strategy's implementation. It contributes to building public trust and ensuring that society at large benefits from data. Co-chaired by Digital, Culture, Media and Sport (DCMS) and techUK, it will be an opportunity for engaging and challenging debate, and a forum to deepen the discussions and define a programme of collaborative working.

In **the European Union,** the General Framework of European Standardisation Policy, Regulation 1025/2012 (European Union, 2012[70]), sets an obligation for European standardisation organisations (CEN, CENELEC, ETSI) and national standardisation bodies to be transparent and publicly available and to be inclusive and allow the participation of all relevant stakeholders. This Regulation is highly relevant for standards related to data access and sharing as reflected for instance in the EU Data Act, which refers to "harmonised standard … as defined in Article 2, point (1)(c), of Regulation (EU) No 1025/2012". (European Union, 2023[71]) As another example, Recital 121 of the EU AI Act explicitly encourages the balanced representation of interests from all relevant stakeholders, "in particular SMEs, consumer organisations and environmental and social stakeholders", in the process of developing AI standards "in accordance with Articles 5 and 6 of Regulation (EU) No 1025/2012".

3.1.1.2. Data partnerships

The establishment of data partnerships have been a promising way through which governments have engaged with private sector actors as well as encouraged collaboration within the private sector (Box 8). Within these data partnerships organisations can agree to share and mutually enrich their data sets, including through cross-licensing agreements. This enables all partners to create additional value and insights that a single organisation would not be able to create. The Recommendation therefore calls on Adherents to:

> *Encourage competition-neutral data-sharing partnerships, including Public-Private Partnerships (PPPs), where data sharing across and between public and private sectors can create additional value for society. (III, b)*

When establishing these partnerships, it is important that governments encourage voluntary and sustainable data access and sharing arrangements within such partnerships. Examples have also shown

that data partnerships have the biggest potential for success where co-operation across and between different sectors can create clear value for society, such as in the case of public service delivery, health care and the management of emergencies, including the COVID-19 pandemic. (OECD, 2020[8]; 2022[1])

Data partnerships should be competition neutral. This means that when establishing these partnerships, countries should ensure that "all Enterprises are provided a level playing field with respect to a state's (including central, regional, federal, provincial, county, or municipal levels of the state) ownership, regulation or activity in the market." (OECD, 2021[72]). In line with the *OECD Recommendation on Competitive Neutrality* [OECD/LEGAL/0462] this means that data partnerships should be based on "open, fair, non-discriminatory, and transparent conditions of competition in government procurement processes in order to ensure that no Enterprise, regardless of its ownership, nationality, or legal form is granted any undue advantage". Adhering to open standards and appropriate certification schemes within data partnership can thus be necessary.

Box 8. Selected examples for data partnerships

Korea's initiative to partner with the private sector to foster the release open data on the availability of face masks in response to the Covid-19 pandemic. This initiative provides evidence on the relevance of engaging with key players in the data ecosystem to respond to emergencies. The sharing of data between public and private actors, and the development of APIs to spur real-time integration, contributed to the development of services citizens could use to make better and more informed decisions on how to deal with first-hand needs during the emergency.[29]

Germany's National Research Data Infrastructure (NFDI) has started a working group on data-sharing between industrial actors and research. The project FAIR Data Spaces connects the Gaia-X data infrastructure and the NFDI to a one-stop cloud-based data space for industry and research which provides a legal and ethical foundation to data-sharing. It furthermore establishes a common technical framework to use Gaia-X technology for enterprises while adhering to the FAIR principles. In that way participating businesses get facilitated access to research data.

Norway's initiative focusing on "Digital collaboration between the public and private sector" is an example of public and private collaboration on data sharing where the public sector, the financial industry and bank customers benefit. In the programme, data sharing is competition-neutral and not exclusive. The sharing process is transparent for bank customers who must consent to the sharing of data.

In the **United Kingdom**, Transport for London (TfL), a local government body responsible for the transport system in Greater London (United Kingdom), has engaged in strategic partnerships with major data, software, and Internet services providers such as Google, Waze, Twitter, and Apple via the sharing of data (including through open data). In some cases, this enabled TfL to gain access to new data sources and crowdsource new traffic data ("bringing new data back"), to undertake new analysis and thus to improve TfL's business operation. In doing so, TfL could gain access to updated navigation information (on road works and traffic incidents) and could enhance the efficiency of its planning and operation.

In line with the *OECD Recommendation on Competitive Neutrality*, governments also need to ensure that data partnerships, "regardless of their ownership, location or legal form, are not ultimately responsible for regulating the market(s) in which they currently or potentially compete (especially regarding entry or expansion of existing players)". The Recommendation therefore calls on Adherents to "take all necessary steps to avoid conflict of interest", which could be an issue when governments act as both a partner in a partnership and regulator, or when the data partnership can directly or indirectly impact market regulation.

In such cases governments may need to separate the regulatory functions from their operational public service functions involved in the data partnership.

3.1.1.3. Transparency

Transparency with respect to what, how and by whom data are collected, accessed, shared, and used has become necessary for stakeholders to trust the data ecosystem and make informed decisions. Transparency is also crucial with respect to how data are governed, including information on processing, including with whom the data is shared, for what purpose, under what conditions access may be granted to third parties, the rights, responsibilities, and respective liabilities in case of violations. Therefore, the Recommendation calls on Adherents to:

> *Enhance transparency of data access and sharing arrangements to encourage the adoption of responsible data governance practices throughout the data value cycle that meet applicable, recognised, and widely accepted technical, organisational, and legal standards and obligations, including codes of conduct, ethical principles and privacy and data protection regulation. (III.c)*

The concept of transparency is well established in the context of privacy and data protection, where a variety of mechanisms have been designed and adopted to promote transparency. Obligations of transparency towards individuals and privacy enforcement authorities, but also the individual right to be informed, to data access and to data correction, are all examples of such mechanisms that promote transparency. These mechanisms can help implement risk management practices as discussed in Section 2.1.3. For example, they can help to identify risks related to discrimination as they empower individuals to ascertain the basis on which decisions about them are taken[30].

The Recommendation therefore further specifies:

> *Where personal data is involved, Adherents should ensure transparency in line with privacy and data protection frameworks with respect to what personal data is accessed and shared, including with whom it is shared, for what purpose, and under what conditions access may be granted to third parties. (III.c)*

The provisions presented above therefore should be read in conjunction with the OECD (2013[53]) *Privacy Guidelines*, including its Openness Principle, which recommends that "There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller."

This means that where personal data is involved, necessary steps may require providing clarification about (i) the criteria for consent or other applicable legal basis for collecting, processing and sharing personal data; (ii) the measures taken to enforce the data subject's rights to access and correct, as well as (iii) the designated institution or institutions whose role is to promote these rights and obligations.

In AI, "transparency" has a slightly different meaning than in the context of privacy and data protection and thus also the Recommendation [DSTI/CDEP/AIGO/DGP(2023)1/FINAL], although all three areas (AI, privacy and the Recommendation) share common objectives that are targeted through transparency requirements, namely increasing trust, trustworthiness and fairness. Within AI, further distinctions are drawn between transparency[31], traceability[32], explainability[33], and interpretability[34], each carrying its own specific implications, which explains the noticeable conceptual diversion between the three areas. Transparency in AI refers to the provision and disclosure of information about an AI system, whereas traceability involves the capability to track components of the AI system before, during, and after its implementation (OECD, 2023[51]). The latter (traceability) is thus conceptually more close to above provision on transparency of the Recommendation, which also addresses issues related to the provenance of data; an area where current Large Language Models (LLMs) continue to perform poorly according to

the May 2024 Foundation Model Transparency Index (Bommasani et al., 2024[73]) As the authors conclude:

"Data remains a key area of opacity. […] Almost all developers remain opaque on these matters. […] These low scores reflect the ongoing crisis in data provenance […], wherein companies share no information about the license status of their datasets, preventing downstream developers from ensuring they are complying with such licenses."[35]

3.1.1.4. Empowerment of individuals, social groups, and organisations

The Recommendation foresees a wide range of approaches to "empower individuals, social groups, and organisations" (III.d) which are referred to as:

> *appropriate mechanisms and institutions such as trusted third parties that increase their agency and control over data they have contributed or that relate to them, and enable them to recognise and generate value from data responsibly and effectively (III.d).*

Box 9 presents examples of some "mechanisms and institutions" that can be considered by policy makers, ideally in combination, to enhance agency and control over data. These approaches are in line with data portability rights emerging across countries and the European Union[36].

Box 9. Possible approaches to enhance agency and control over data

The following approaches can be considered by policy makers, ideally in combination as discussed in (OECD, 2022[25]), to enhance agency and control over data:

- *Technological measures*, can refer to a collection of digital tools that permit processing, analysis and sharing of information while protecting the confidentiality of data. They include:
  - *Privacy-enhancing technologies (PETs)*, including tools that allow data subjects to monitor and increase their awareness on how their data us being used and by whom. The OECD (2022[74]) differentiates between the following classes of PETs: (i) data obfuscation, (ii) encrypted data processing, (iii) federated and distributed analytics, and (iv) data accountability tools.
  - *Technological protection measures (TPMs),* which provide legal protection for passwords, paywalls or access control and copy control measures. They are often used in the context copyright protection in combination with rights management information (RMI).[37]
  - *Public sector tools such as citizens' folders* which are typically paired with digital identity management systems to support citizens' monitor and control over their personal data and its processing by public bodies. (OECD, 2023[75]; OECD, 2021[69])
- *Organisational measures*, which refer to institutional arrangements that may involve contractual arrangements often in combination with PETs to help manage control over data. Examples include:
  - *Data sandboxes* which are isolated environments through which data are accessed and analysed. These sandboxes can be realised through PETs but also through physical presence within the facilities where the data are located. (OECD, 2019[2])
  - *Data intermediaries* which the Recommendation defines as "service providers that facilitate data access and sharing under commercial or non-commercial agreements between data holders, data producers, and/or users."

Box 9. Possible approaches to enhance agency and control over data (cont.)

- *Legal measures* can complement technological and organisational measures via legally binding and enforceable obligations to establish and protect the rights of stakeholders. Examples include:
  - *Consent* remains a key pillar in privacy and data protection frameworks to allow individuals to control the collection and (re-use) of personal data about them, in particular if they are also given reasonable means to extend or withdraw their consent over time.
  - *The right to rectification* is a relevant mechanism for ensuring the accuracy of personal data.

Data portability has become an essential tool for enhancing users' control and agency over their data, while enhancing access to and sharing of data across digital services and platforms. (see Box 2). When implementing above approaches, policy makers should recognise that: (i) individuals can be affected by biases, including through information overload, which impact their ability to take rational data-related decisions; and (ii) these vulnerabilities can be exploited through so-called *dark patterns* online. (OECD, 2022[1])

Source: OECD (2022[25]), A Guide for Data Governance Policy-making and OECD (2022[1]), Data Stewardship, Access, Sharing and Control: A Going Digital III module synthesis report

3.1.2. Adopting a strategic whole-of-government approach to ensure that data access and sharing arrangements effectively and efficiently meet specific societal, policy, and legal objectives that are in the public interest

Given resource constraints and the costs of data access and sharing, governments have started to prioritise access to what they consider high-value datasets according to the needs of society. This approach carries however multiple risks. For example, what constitute a high-value dataset today may vary over time and may also depend on the context of its potential use. In order to ensure that data access and sharing arrangements, including those based on high-value datasets, effectively and efficiently meet specific societal, policy, and legal objectives that are in the public interest, governments need to adopt a strategic whole-of-government approach.

"Strategic" within the meaning of the Recommendation is any approach that helps to "effectively and efficiently meet specific societal, policy, and legal objectives that are in the public interest". The Recommendation therefore calls on Adherents to:

> *Prioritise data access and sharing arrangements that help achieve such objectives, taking into account applicable laws and regulations. In so doing, Adherents should work together with key stakeholders to clearly define the purpose of these arrangements and identify data relevant to these purposes, taking into account their benefits, costs, and possible risks. (IV.a)*

3.1.2.1. National data strategies for more coherent, flexible, and scalable data governance frameworks

The Recommendation highlights national data strategies specifically and whole-of-government approaches more generally, as means to ensure that data access and sharing arrangements can help achieve strategic objectives. The prioritisation of data access and sharing arrangements should thereby take into account cross-cutting economic, social, cultural, technical, and legal governance issues. This is why the Recommendation calls on Adherents to:

> *Adopt and regularly review coherent, flexible, and scalable data governance frameworks – including national data strategies, which integrate cross-cutting economic, social, cultural, technical, and legal governance issues – in order to foster data access and sharing within and across society, public and private sectors, and jurisdictions.*

Particular emphasis is put on national data strategies to address cross-cutting data governance issues. These strategies are cross-sectoral by nature and, in many instances, are designed explicitly to contribute to reaching higher level objectives, such as GDP growth, productivity, well-being, and/or combating climate change and fostering sustainable development. To achieve these strategic objectives, many national data strategies build on specific strengths of the country and often integrate pre-existing national strategies related to the digital economy, for example, national digital economy strategies (Gierten and Lesher, 2022[76]), national broadband strategies, digital government strategies, and national AI strategies, which national data strategies often complement. In turn, national data strategies are often complemented by sector specific data strategies, with the most prominent sectoral data strategies being related to health, but also scientific research as well as smart cities, smart transportation, and smart energy systems (see Box 10 for examples of national data strategies).

Box 10. Examples of national and sectoral data strategies

The United Kingdom's National Data Strategy (NDS) aims to harness the power of responsible data use, to boost productivity, create new businesses and jobs, improve public services, support a fairer society, and drive scientific discovery. The United Kingdom's NDS pursues five missions, described as follows:

1. To unlock the value of data across the economy, by creating an environment where private sector data is appropriately available, accessible and usable across the economy, while protecting people's data rights and private enterprises' intellectual property.
2. To secure a pro-growth and trusted data regime, and, by taking advantage of the Brexit to improve the United Kingdom's data protection framework, as well as by modernising the Information Commissioner's Office (ICO) so it's better equipped for new challenges.
3. To transform government's use of data to drive efficiency and improve public services, by transforming how government manages, uses and shares data so that it is joined up, reliable, high quality and underpinned by standards of transparency and openness across government.
4. To ensure the security and resilience of the infrastructure on which data relies by delivering on a number of government priorities, including COVID-19 response, contribution to the National Security and Investment Act, and working across government to identify critical data centre and cloud platform infrastructure.
5. To champion the international flow of data, by working with partners to influence global data governance to create an accessible, interoperable data ecosystem that promotes fair digital competition, thereby creating the right global framework of infrastructure, standards and rules to enable greater access and availability of data, protect the country's national and economic security, pursue greater regulatory co-operation, support innovation for science and technology.

Other relevant examples at the national level include the **U.S**. Federal Data Strategy, **Ireland**'s Public Service Data Strategy, **Japan**'s Priority Policy Program for Realizing Digital Society, and **Germany**'s Data Strategy of the Federal German Government (OECD, 2019[21]).

Box 10. Example of national and sectoral data strategies (cont.)

The European Union's Data Strategy provides a comprehensive approach containing both horizontal legislation such as the Open Data Directive, the Data Act and the Data Governance Act as well as more sectorial legislation and non-legislative initiatives such as the Common European Data Spaces. Under the Strategy, rules regarding the making available and the sharing of data in B2B, B2G and B2C relations are established. The Strategy puts in place a cross-sectoral legal framework for European data governance; provides legal certainty regarding data access, use and sharing through regulation; boosts research, innovation and competitiveness through programs such as Digital Europe Program and Horizon Europe Program, including the establishment of the Data Spaces Support Centre; ensures the harmonious application of legislation put in place through creating expert groups (European Data Innovation Board); and helps overcome barriers to sharing, through technical infrastructure, legal rules and ethical guidelines.

Sectoral data strategies

Finland's Act on Transport Services provides that operators who offer passenger transport services must be able to provide the essential information on routes, stops, timetables, prices and accessibility related to their services. This information must be provided through an open interface and in default format that is computer-readable. Interfaces must be provided to all actors under fair, reasonable and non-discriminatory terms. More information: www.lvm.fi/-/act-on-transport-services-955864

Italy's National Public Sector Data Strategy aims to improve the way public data is generated, protected, used, managed, and shared to create citizen-centred public services, support business, and scientific research, and optimize decision making. The goals also include achieving full interoperability of public sector data, facilitating the reuse of private data for the public interest, and stimulating responsible and transparent cross-sectoral collaborations on data.

The European Union's Common European Data Spaces bring together relevant data infrastructures and governance frameworks to facilitate data pooling and sharing. Common European data spaces are decentralised infrastructures where diverse actors can share, access and use data in a secure, reliable and trustworthy manner, following common governance, organisational, regulatory and technical mechanisms. To achieve data exchange across societal actors, they interconnect various data ecosystems in a demand-driven process. Common European data spaces ensure that more data becomes available for use in the economy and society, while companies, organisations and individuals who generate the data retain control. Through its DIGITLA Europe Programme, the European Commission is funding data spaces in 14 strategic sectors and domains of public interest[38]. As these are bottom-up initiatives, the EU is having a supportive role and provides funding and advice.

Brazil's Digital Government Strategy (EGD) provides guidelines that guide Brazil's digital transformation. The EGD structure encompasses six principles, 18 objectives and 59 initiatives. Among the objectives are the Data Governance guidelines of the direct, autarchic and foundational federal public administration on data management actions from the perspective of sharing, architecture, security, quality, operation and other technological aspects. The EGD relies on the exercise of authority by the Central Data Governance Committee, which has control over rules and procedures for management activities. Therefore, the regulation of Data Governance intends to promote the interoperability of information and the integration of public services.

Box 10. Example of national and sectoral data strategies (cont.)

National digital strategies with a focus on data

Denmark's Joint Public Digitalisation Strategy, which aims to strengthen welfare, acceleration of green transition, increase growth, export through digitalisation, and protect the Danish citizens and companies against cyber-attack, includes an initiative (Initiative 24) which addresses data sharing. Initiative 24 supports local, regional, and national authorities in making their data discoverable through a national data portal for public data (Datavejviser.dk). The purpose of this is to facilitate access to public sector data for companies, authorities and scientists to help them generate more innovation, growth, and thereby make public sector data more valuable. The portal offers a searchable overview across the many platforms that house public data. Other relevant initiatives include: (i) Initiative 5 – Better digital access for children's guardian and others care provider who are not a guardian e.g. grandparents, step parents, foster parents etc., (ii) Initiative 10 – Better access to health data for citizens and health professional, (iii) Initiative 14 – Parents' digital access to their children's health information.

3.1.2.2. Whole-of-government approach that enables effective policy coordination and implementation

The prioritisation of data access and sharing arrangements exercise must involve all relevant stakeholders as discussed in Section 3.1.1, including all relevant branches within government. This is particularly relevant for training AI models in a way that represents diverse groups within the country while mitigating risks of discrimination and undue biases. While national data strategies are a promising way to implement such a whole-of-government approach, there are alternatives.[39] Therefore, the Recommendation calls for Adherents more generally to:

> *Demonstrate strong leadership, ideally at the highest level of government, combined with a whole-of-government approach that enables effective policy coordination and implementation of these frameworks with multi-stakeholder participation; (IV.c)*

Box 11 presents examples on how Adherents have implemented a whole-of-governance besides through national data strategies.

Box 11. Implementing a whole-of-government approach for data access and sharing in Finland

Finland's Government established a "Ministerial Working Group on Developing the Digital Transformation, the Data Economy and Public Administration". The new Ministerial Working Group will coordinate activities and situational awareness related to digitalisation and information policy, technology policy and the data economy at the government level, and will coordinate development projects in accordance with its mandate. More information: https://valtioneuvosto.fi/-/10616/hallitukseen-uusi-digitalisaation-datatalouden-ja-julkisen-hallinnon-kehittamisen-ministerityoryhma?languageId=en_US and www.lvm.fi/en/-/ministerial-working-group-to-develop-digital-transformation-data-economy-and-public-administration-1494070

*3.1.2.3. **Technology-neutral and agile legal and regulatory environments***

The legal and regulatory environments and the data access and sharing arrangements they affect need to take into account the various types of data, the various risks associated with their re-use, and the

overlapping and interests of all relevant stakeholders. They also need to internalise and proactively act upon the rapid and continuous, and often also unpredictable changes in digital technologies and analytical advances, and the risks associated with these changes.

A lot will depend on the particular context in which data is accessed and shared, which can vary over time. For instance, trusted third parties that help manage potential risks from data sharing as discussed in Section 3.1.1. need to be able to adjust their terms and conditions in an agile fashion depending on the various types of data, the available digital technologies, risks faced by the actors involved, and their interests. Data access and sharing arrangements may also need to fulfil requirements that may be unknown ex ante, for instance, where they enable access to data required by the foresight communities to identify and anticipate emerging changes and megatrends.

Where appropriate, and where justified by the societal objectives, technology-neutral and agile legal and regulatory environments may be needed to promote responsible data access and sharing and to enable regulatory innovation, while providing the necessary legal certainty and protection with the engagement of all relevant independent enforcement authorities, oversight bodies and stakeholder groups. The Recommendation therefore calls on Adherents to:

> *Adopt technology-neutral and agile legal and regulatory environments that promote responsible data access and sharing and enable regulatory innovation, while providing the necessary legal certainty and protection with the engagement of all relevant independent enforcement authorities, oversight bodies, and stakeholder groups. (IV.d)*

Innovative approaches such as regulatory sandboxes are being adopted by governments and regulators to explore the opportunities and risks presented by data access and sharing under a more controlled environment. (Attrey, Lesher and Lomax, 2020[77])(See Box 12 on Singapore's data regulatory sandbox).

Box 12. The implementation of agile legal and regulatory environments in Singapore

Singapore's Infocomm Media Development Authority (IMDA) offers a Data Regulatory Sandbox to businesses and their data partners to explore and pilot innovative use of data in a safe legal and regulatory "environment". The sandbox helps to reduce uncertainty in compliance with current and planned policies, and limits the exposure of companies and consumers. There are three stages in the Data Regulatory Sandbox, where they are not necessarily sequential. Companies will participate in the sandbox based on their use case and readiness:

- Stage one (engagement) - companies identify areas of interest and provide plans to innovate with data, IMDA/Personal Data Protection Commission (PDPC) review and provide regulatory advice;
- Stage two (providing guidance) – companies provide details of specific use case or requirements for trial, IMDA/PDPC provide either general or practical guidance to help companies have clarity and understanding regarding the uncertainty on the innovative use of data;
- Stage three (policy prototyping) – companies with use case that supports detailing of new policy intent may seek exemption and they are required to conduct risk impact assessment and implement measures to mitigate risks.

3.1.3. Maximising the benefits of data access and sharing, while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives

Enhanced data access and sharing can provide social, and economic benefits, including good public governance. At the same time, it may bring about increased risks to national security, individuals and organisations, including the risks of confidentiality and privacy breaches, and the violation of other legitimate private and public interests such as commercial interests (see Section 2.1.3).

Given these significant risks, some restrictions may be necessary to reinforce trust across the data ecosystem. However, others are maybe unjustified and/or unintended barriers that disproportionately limit the social and economic potential of data, thereby creating significant social and economic opportunity costs and in some cases unethical outcomes. In some instance, these barriers can have negative effects on society, where, for instance they prevent access to, sharing and re-use of health-relevant data needed to address health crises such as the COVID-19 pandemic.

Therefore, data access and sharing may require effective approaches that strike the right balance between the benefits and the risks of data access and sharing, while protecting the rights and interests of stakeholders and in line with transparency, accountability, proportionality principles. This is where policy makers can leverage the data openness continuum as discussed in Section 2.1. The Recommendation refers to the data openness continuum implicitly in provision V. In particular, it recommends that Adherents should:

> *Encourage data access and sharing arrangements that ensure that data are as open as possible to maximise their benefits and as closed as necessary to protect legitimate public and private interests … (V.a)*

The Recommendation furthermore sets out that the steps that Adherents take "to protect these legitimate public and private interests" need to be "necessary and proportionate" (V.b). The Recommendation does not further specify how to assess the necessity and proportionality of the measures taken to protect legitimate public and private interests. But where personal data are involved the proportionality assessment articulated in paragraph 18 of Part Four of the *OECD Privacy Guidelines* is pertinent even if the latter is addresses "transborder flows of personal data": "Any restrictions to [these flows] should be proportionate to the risks presented, taking into account the sensitivity of the data, and the purpose and context of the processing."

Determining the appropriate steps, in line with applicable laws and regulations, requires a risk management approach (see Section 2.1.3). This is why the Recommendation specifically calls on Adherents to:

> *Ensure that stakeholders are held accountable in taking responsibility, according to their roles, for the quality of the data they share and for the systematic implementation of risk management measures throughout the data value cycle, including measures necessary to protect the confidentiality, integrity, and availability of data (data security). To this effect, Adherents should promote the adoption of impact assessments and audits as well as responsible stewardship for data sharing within organisations, … (V.c)*

This also requires enhancing transparency, as discussed in Section 3.1.1.3. Therefore, the Recommendation further sets out that:

> *…. Adherents should strive to ensure that stakeholders are fully informed as to their rights (including their right to information and to obtain redress), responsibilities and respective liabilities in case of violations of privacy, intellectual property rights, competition laws, or other rights and obligations. (V.b)*

The Recommendation also offers practical approaches on how to "ensure that data are as open as possible to maximise their benefits and as closed as necessary to protect legitimate public and private interests". Provision V.d. in particular recommends that Adherents:

> *Foster the adoption of conditioned data access and sharing arrangements with the use of technological and organisational environments and methods, including data access control mechanisms and privacy enhancing technologies, through which data can be accessed and shared in a safe and secure way between approved users, combined with legally binding and enforceable obligations to protect the rights and interests of data subjects and other stakeholders.*

This provision should be read in conjunction with provision III.d on the empowerment of individuals, social groups and organisations (Section 3.1.1.4), given that all the approaches presented in Box 7 to enhance users' agency and control over data are pertinent here as well and more (see Box 13).

BOX

Box 13. Examples on how to facilitate data access and sharing while protecting individuals' and organisations' rights

In **Belgium**, by offering personal data pods to citizens and enterprises, the Flemish Data Utility Company aims to smoothen the path for value creation using data by businesses and stimulating the data-economy. The Flanders Digital Agency, responsible for the digitalisation of the Flemish administration, together with the Flemish Data Utility Company, will take up the role as facilitators and data-ecosystem managers towards governments and businesses.

Canada's Bill C-27, which includes both the Canadian Consumer Privacy Protection Act (CPPA) and the Artificial Intelligence and Data Act (AIDA) is intended to address risks that arise from the use of personal information/ de-identified information (CPPA) and the development of AI technologies (AIDA). They provide for incentives for the responsible use of personal information. CPPA promotes the development and implementation of privacy management programs to ensure compliance with the law, and both proposed laws set out serious consequences for non-compliance, including for egregious behaviour. AIDA regulates the design, development, and deployment of AI systems in the course of interprovincial and international trade, in order to ensure that appropriate measures are in place to address risks of harm and bias. Stakeholders will be integral to this process, and actively consulted to strike the right balance between benefits and risks, while protecting the rights of Canadians. If passed, new measures from AIDA would be required of persons responsible for high impact AI systems, including transparency, and risk assessment and mitigation requirements. Furthermore, a new AI and Data Office headed by a Data Commissioner will be established.[40]

Germany's "Datentreuhandmodelle" (data trustee models), funded by the Federal Ministry of Education and Research ("Bundesministerium für Bildung und Forschung", BMBF), acts as neutral intermediaries between data producers and users by also bridging different domains and sectors. They seek to achieve a fair representation and fair balance of interests of each side to generate trusted venues for innovative data-sharing and partnership. They promote technological as well as organisational access to quality data while respecting data privacy requirements, property rights, scientific openness, and IT security, among other interests.

Box 13. Examples on how to facilitate data access and sharing while protecting individuals' and organisations' rights (cont.)

The United Kingdom has implemented safeguards and enforcement regimes to ensure that data is collected and handled responsibly and securely. The United Kingdom is taking action to secure data, whilst maximising the benefits of data access and sharing. This is vital to enable a thriving, competitive and innovative economy and protect national security interests. Furthermore, a robust regulatory and organisational regime is already in place: there are categories of data sharing that are not permitted subject to a consent framework and/or can only be done in certain ways to manage those risks, which the government continues to keep under review. Levers to manage this include the ICO's data sharing code of practice, the Centre for the Protection of National Infrastructure's Security-Minded approach to Open and Shared Data, the Official Secrets Act, the Information Management Framework which is currently under development, and other relevant legislation and guidance. The Central Digital and Data Office's Data Ethics Framework, which is designed to guide public sector use of data, may also inform how organisations in the private and third sector use data.

In the **European Union** the key principle of the Data Strategy is to facilitate the making available of, access to and use of data while keeping the individual or body generating the data in control of it. With this approach the EU pursues the empowerment of individuals in regard to their data. This is reflected not only in the right of the user to access the data generated on their devices under the Data Act but also in the obligation of data intermediaries to act in the best interest of users that use their services under the Data Governance Act. Moreover, the legislation in place balance between protecting business interests (trade secrets, IPR) of data holders and providing transparency to users as to which data will be used for what reasons. Furthermore, the EU will establish very concrete measures such as a harmonised consent form that will help to make more data available for reuse for objectives of general interest (e.g. for purposes of research and innovation) while complying with existing legislation in place, such as the GDPR to protect personal data.

Singapore's IMDA is exploring the new approaches to facilitate trusted data sharing including based on PETs. In research, since 2018 Singapore has established two programmes dedicated to PETs which have seen some early industry translation projects and won research accolades. To drive the next bound in PET and trust technology innovation, Singapore has set aside SGD 50 million with a focus on applied research and translation. Commercially, the use of PETs is wide and uninhibited among big technology companies. However, in other sectors, despite the maturity of some PETs, the deployment of this technology is facing barriers around lack of knowledge over Use Case-to-PET fit, lack of standards and benchmarks of the technology, and uncertainty in policy positions on the use of PETs. The Singapore Government's IMDA agency is interested in understanding the nature of these barriers and finding solutions to address them by piloting PETs in real-world applications. To this end, the agency signed an MOU with CEIMIA (Canada), under the auspices of the GPAI Data Governance workgroup, for developing a technology demonstration of how PETs can enable an AI system.

3.2. Stimulating investments in data and incentivising data access and sharing

Substantial investments are often required for data management, including creating, collecting, storing, curating, enriching, deleting, providing access to, and sharing data, as well as using data, and managing the associated risks. This may not only include the cost required to assure that the data meets the quality requirements (e.g., cost for data cleaning and data curation). Additional investments are also needed to secure the engagement of all relevant stakeholders, including for the complementary assets such as data-

related skills and infrastructures, which are needed as discussed further below. Means to safeguard returns on investments and other monetary and non-monetary reward mechanisms are thus needed to further incentive investments in data and data sharing in particular when the long-term availability of data needs to be ensured. The Recommendation recognises that these investments may be sustained by a range of different business and financial models over the long term.

3.2.1. Encouraging market-based approaches by providing coherent incentive mechanisms and promoting conditions for the development and adoption of sustainable business models

3.2.1.1. Sound competition policies and regulations for competitive markets for data

As recognised by the Recommendation, "market-based approaches, including commercialisation of data and freedom of contract, are essential for incentivising data access and sharing and related investments, but … there may be costs, risks, and limitations to these approaches' ability to fully meet demand for data". One of these limitations is related to competition, which some countries such as the United Kingdom have addressed (Box 14).

Data markets and intermediaries that provide value-added services such as a payment and data exchange infrastructure can facilitate data access and sharing including through the commercialisation of data. But the dynamics of data markets and the adoption of data-driven business models have raised concerns about the extent to which the control of data may become a source of market power that disincentives non-incumbents from making investments in data and data use. (OECD, 2022[78]) As a consequence, some have suggested a need for competition authorities to better address and correct these developments.[41]

The Recommendation addresses the issue of competition in provision VI.a which recommends that Adherents:

> *Foster competitive markets for data through sound competition policy and regulation that addresses possible exploitation of market dominance and other appropriate measures, including enforcement and redress mechanisms that increase stakeholders' agency and control over data and ensure an adequate level of consumer, intellectual property, and privacy and personal data protection. (VI.a)*

The reference to "other appropriate measures, including enforcement and redress mechanisms" reflects the interplay between competition and other regulatory measures including in particular privacy and data protection and consumer protection. (OECD, 2015[4]; 2022[25]) It thus emphasises the need for enhanced co-operation and co-ordination across regulatory agencies as called for by provision IV.c on a whole-of-government approach (see Section 3.1.2.2). The intersection of competition and data privacy has become increasingly prominent, with recent assessments by competition agencies exploring whether privacy considerations constitute an antitrust issue. (Abate, Bianco and Casalini, 2024[79]) This has accelerated the integration of competition considerations into data protection frameworks and vice versa, while underscoring the urgent need for innovative models for co-operation between competition and data protection authorities.

Provision VI.a also emphasises the need for "increasing stakeholders' agency and control over data" as called for by provision III.d on the empowerment of individuals, social groups, and organisations, thus highlighting mechanisms such as data portability as means to foster competition (see Section 3.1.1.4).

Box 14. Fostering competitive markets for data in the United Kingdom

The United Kingdom is looking to encourage the market in the following ways:

- *Using incentives to maximise value for money data sharing in support of public good*: The United Kingdom is reforming research and development (R&D) tax reliefs to support cutting-edge research methods by expanding qualifying expenditure to include data and cloud costs. This includes modernising the reliefs to better incentivise R&D methods which rely on vast quantities of data that are analysed and processed via the cloud. The United Kingdom is also building its evidence base on how incentives could work in practice and considering the types of data use that most requires incentivisation.
- *Support effective and well-functioning markets by addressing data practices that distort competition and consumer outcomes* — including by widening access to data, where appropriate: The United Kingdom is creating the Digital Markets Unit (DMU) within the Competition and Markets authority, which will have statutory powers to support effective and well-functioning digital markets, including by addressing data practices by larger firms that distort competition and consumer outcomes. Interventions could include opening up or widening access to specific datasets, where there is evidence, this will drive up competition, building on the successes of the Open Banking initiative by delivering the full potential of Smart Data solutions.

3.2.1.2. Self- or co-regulation mechanisms for data access and sharing

The intricate net of existing legal frameworks including IPR and privacy protection frameworks, combined with the involvement of multiple parties in the creation of data (and its value) including across national borders, has led to a number of legal uncertainties (OECD, 2022[24]; OECD, 2019[2]). This is supported by work by the Global Partnership on AI (GPAI) that shows that many organisations that want to share data and AI models voluntarily and responsibly, face challenges due to a lack of regulatory harmonisation and clarity, insufficient technical tools, insufficient contractual tools and codes of conduct, and difficulties valuing the data and AI models. (GPAI, 2022[80]; GPAI, 2023[81])

As a result, stakeholders have come to rely on contract law as the primary legal means for determining rights related to data access and use. While freedom of contract may give stakeholders the ability to construct well-suited contractual arrangements that best fit their needs, uncertainties remain that may increase transaction costs, and expose particularly those stakeholders that are in a weaker position to negotiate fair terms and conditions.

Voluntary guidance, codes of conduct and templates that define a set of contractual clauses with a focus on potentially contentious issues are promising, where they can help reduce legal uncertainties and transaction costs (see Box 15). The Recommendation therefore calls on Adherents to:

> *Promote, where appropriate, self- or co-regulation mechanisms – including voluntary guidance, codes of conduct and templates for data access and sharing agreements – that provide legal flexibility while ensuring that all relevant stakeholders have certainty as to applicable laws and regulations. (VI.b)*

Box 15. Government examples of guidance, codes of conduct and templates for data access and sharing

The **Australian** Research Council (ARC) is dedicated to maximising the benefits of funded research by promoting increased access to research data. The ARC encourages researchers to consider how they can effectively manage, store, disseminate, and reuse data. Under ARC's data management framework and Open Access policy, researchers, in consultation with institutions, hold a responsibility to assess the management and future potential of their research data, accounting for the diverse approaches, standards, and uses that may exist across different institutions, disciplines, and research projects. The ARC's Data Management Statement outlines requirements consistent with the Australian Code for the Responsible Conduct of Research (2018), which establishes guidelines for the proper management of research data and primary materials by researchers. This includes adherence to institutional policies covering data ownership, storage, retention, and ensuring "appropriate access... by the research community."[42]

Japan formulated its 2018 *Contract Guideline on Utilisation of AI and Data*, which summarises the issues and factors to be considered when drafting a contract on the utilisation of AI or data. It is intended to be used as a reference when private businesses conclude contracts related to data re-use or development and use of AI-based software. In 2021, it developed the "Basic Guidelines for the Management of Health and other personal data by Private-sector PHR Business Operators", to promote the use of safe and secure private Personal Health Record (PHR) services. The document requests that private PHR providers comply with the guidelines. In addition, Japan supported the establishment of the "PHR Service Business Association (PSBA)" by private PHR providers, and the PSBA established additional guidelines aiming for higher service standards to the Basic Guidelines by the end of June 2024.

The Netherlands' Code of Conduct for Research Integrity was introduced in 2018 and encompasses five principles and 61 standards for good research practices. The principles are honesty, scrupulousness, transparency, independence and responsibility. When it comes to making research data openly available, the principle of transparency lays out that it should be clear, at least to peers, what data the research was based on, how it was obtained, what and how results were achieved. In case parts of the research data are not intended to be made public, researchers are committed to providing a good account of why this is not possible.

Singapore's IMDA has published the Trusted Data Sharing Framework in June 2019 to help organisations put in place a set of baseline practices through a common "data-sharing language" and suggests a systematic approach to the broad considerations for establishing trusted data sharing partnerships. It includes sample legal templates to kickstart early discussions in contractual bilateral data sharing. In response to industry need for multilateral data sharing collaborations, IMDA further developed two additional legal templates to guide businesses who are establishing data partnerships to either collaboratively solve a common use case or co-create new data-driven insights.

The development of these guidance, codes of conduct and templates should be undertaken by, and in active collaboration with, the relevant stakeholder groups in line with provision IV.a on multi-stakeholder engagement and participation (see Section 3.1.1.1).

In the context of its "Protecting Innovation, Intellectual Property (IP) Project", the GPAI is working on standardised contractual terms to facilitate AI data and model sharing that would contribute to existing private sector initiatives for standardised contractual terms for data access and sharing (Box 16).[43] At a recent GPAI workshop, attendees suggested the possibility of needing various standard contract forms to

accommodate the diverse AI and data use cases and applications. However, some participants also raised concerns that an overly broad range of standardised contractual terms could counteract the purpose of standardisation. (GPAI, 2023[81]) The most prominent private sector initiatives in this area emerge out of the open source / open knowledge community.

While self- or co-regulation mechanisms offer valuable flexibility in managing data access and sharing, there are also limitations and challenges. These can include the lack of uniform legal obligations across jurisdictions; the absence of regulatory oversight, which may render self-regulation ineffective; and gaps in dispute resolution or the exercise of rights.

In fact, when the European Commission's (2023[71]) proposed the EU Data Act, it acknowledged these shortcomings, citing “the limited efficacy of the self-regulatory frameworks” and “the general unavailability of open standards and interfaces”. It thereby referred to the EU's Free Flow of Non-Personal Data Regulation, which requires businesses to adhere to “self-regulatory codes of conduct” monitored by the European Commission. Regulatory frameworks like the EU Data Act and the EU Digital Markets Act aim to close these gaps by imposing mandatory obligations on data access and sharing.

Thus, while self- and co-regulatory measures have their merits, a nuanced approach that addresses these limitations is crucial for ensuring that they serve all stakeholders fairly and effectively. It is therefore not only advisable for any self- or co-regulatory initiatives to be developed in conjunction with regulatory bodies and multiple stakeholder groups to ensure comprehensive coverage, transparency, and enforceability. Governments may want to reserve the option to enforce compulsory regulations if voluntary measures proved insufficient as examples from the United Kingdom's midata initiative, the Australia’s CDR and European Commission have demonstrated.

Box 16. Non-governmental initiatives for standardised contractual terms for data access and sharing

ALI-ELI Principles for a Data Economy (Cohen and Wendehorst, 2021[82]) is a collaborative effort between the American Law Institute (ALI) and the European Law Institute (ELI). It aims to provide a contractual framework for data-related transactions and rights and is intended for policymakers, legal professionals, and stakeholders in the data industry. Divided into four parts, covering general provisions, data contracts, data rights, third-party aspects, and multi-state issues, the Principles explores the relationship between these topics and existing contract laws. It also covers contractual issues in both B2C and B2B contexts, highlighting the differences between continental European legal systems and common law jurisdictions. It also addresses contractual issues in data-related transactions in the context of AI.

Microsoft’s data sharing agreement templates as a private sector initiative (Microsoft, 2019[83]): In July 2019, Microsoft published three data sharing agreements to be used as template: (i) the Open Use of Data Agreement (O-UDA), (ii) the Computational Use of Data Agreement (C-UDA), and (iii) the Data Use Agreement for Open AI Model Development (DUA-OAI). These were complemented by a fourth one later in November 2019: (iv) the Data Use Agreement for Data Commons (DUA-DC). The O-UDA addresses free, unrestricted data use for both commercial and non-commercial purposes, while the C-UDA addresses indirect data utilisation rights through computational methods. The DUA-OAI focuses on rights for data in the context of AI development. Lastly, the DUA-DC defines guidelines for a shared data repository accessible to multiple entities.

Box 16. Non-governmental initiatives for standardised contractual terms for data access and sharing (cont.)

The Linux Foundation (2017[84]) in collaboration with a broad set of participating organisations, introduced the Community Data License Agreement (CDLA) as a family of open data agreements. The CDLA licenses define a licensing framework to support collaborative communities built around curating and sharing "open" data (see Box 2). There are two CDLA licenses: a Sharing license that encourages contributions of data back to the data community (CDLA-Sharing-1.0) and a permissive license that puts no additional sharing requirements on recipients or contributors of open data (CDLA-Permissive-2.0). Both encourage and facilitate the productive use of data.

The Open Knowledge Foundation (n.d.[85]) has also introduced open data commons (ODC) licences for databases, where users are free to copy, distribute and use the database; produce works from the database; and modify, transform and build upon the database. Besides the Open Data Commons Public Domain Dedication and License (PDDL 1.0), which has no requirements, Open Data Commons Attribution License (ODC-By 1.0) require attribution. The Open Data Commons Open Database License (ODbL 1.0) requires further that any public use or distribution of an adapted version of the database, or works derived from it, must also make the adapted database available under the same ODbL terms (share-alike). Redistribution of the database or an adapted version that includes restrictions, such as DRM, must also include a version free of such measures (keep-open).

The most recent version of **Creative Commons** (Creative Commons, 2024[86]) (CC 4.0) can be used to license databases subject to copyright and, where applicable, sui generis database rights. The variations include CC0 (public domain), CC BY (attribution), CC BY-SA (attribution-sharealike), CC BY-ND (attribution-noderivs) preventing derivative work, CC BY-NC (attribution-noncommercial), CC BY-NC-SA (attribution-noncommercial-sharealike), and CC BY-NC-ND (attribution-noncommercial-noderivs). CC does not recommend the use of its NonCommercial (NC) or NoDerivatives (ND) licenses on databases intended for scholarly or scientific use. In contrast, the CC0 Public Domain Dedication is recommended where database owners want "to waive all copyright and related rights in the database and its contents, placing it as close as possible into the worldwide public domain".

3.2.2. Promoting conditions for the development and adoption of sustainable business models and markets for data access and sharing

As highlighted above, market-based approaches have their limitations. Where data is used to produce a public or a social good (e.g., scientific knowledge and democratic participation), market-based approaches can be suboptimal, and further government action may be warranted. In these cases, the positive externalities generated by the creation and use of data as public or social goods may justify public financing for data access and sharing arrangements, in particular, for open or publicly accessible data.[44]

However, increasing costs combined with budgetary constraints have challenged the sustainability of open data repositories, as open data is increasingly expected to be provided free of charge, in addition to the requirement that it be provided in machine-readable and non-proprietary formats without any barriers blocking its access, sharing, use and re-use (OECD, 2021[69]; 2019[2]; 2022[25]; 2019[21]). This raises the question of how the provision of data can be funded sustainably, given that in the case of open data costs are most often borne by data holders, while benefits accrue to data users.

The Recommendation therefore calls on Adherents to:

> *Support long-term investments in data access and sharing arrangements to ensure their sustainability, including in open data arrangements. Adherents should consider a combination of various structured financing and revenue models to support these arrangements where appropriate. (VI.c)*

Financing models that can be leveraged in combination include among others: structural funding, host or institutional funding, annual deposit-side contract, data deposit fees, project funding. In addition, different revenue models such as the freemium revenue model can be used for cross-subsidising various activities to assure the collection and commercialisation of data.

Box 17. Financing models for open science data repositories

For science data repositories more specifically, policy makers, research funders, and other stakeholders need to consider the ways in which data repositories are and can be funded, and the advantages and disadvantages of those business models in different circumstances:

- *Structural funding* typically involves a trade-off between funding for data repositories and funding for other research infrastructure or for research itself. That allocation will best be made by informed actors making choices, such as through a funding allocation process involving widespread research stakeholder participation, expert consultation, and “road-mapping”.
- *Host or institutional funding* may divorce informed actors from the funding decisions or require additional processes to ensure greater stakeholder understanding of the value of the repository services.
- *Data deposit fees*: bring the trade-off closer to the researchers, but their success in optimising allocation will depend on the extent to which the actors are informed and on their freedom of choice. The latter may be constrained by open data mandates (regulation).

Source (OECD, 2017[92])

3.2.3. Providing coherent and appropriate incentive mechanisms

Data access and sharing may create externalities that benefit others more than the data holder and data subject. Therefore, data holders and controllers may not necessarily be incentivised to share their data. This is in particular true if the costs and risks of data access and sharing are perceived to be higher than the expected returns. In addition, there is growing evidence of reluctance to data sharing by individuals and organisations due to uncertainties in the implementation of existing privacy protections and other legal requirements. These also include liability considerations in case of unforeseen adverse consequences from the release and sharing of data.

Furthermore, misalignments of incentives in reward and evaluation systems may undermine the willingness of individuals to share data. This is for instance the case in science, where researchers are primarily rewarded for their scientific papers and not for the data they share (see Box 18). This is also true in the public sector where current incentives to civil servants may not be well aligned with overall societal and policy objectives due to technical challenges, institutional barriers, or confidentiality and digital security concerns.

The Recommendation therefore calls on Adherents to:

> *Promote appropriate incentive mechanisms that enable the fair distribution of the benefits of data access and sharing arrangements and ensure that stakeholders are enabled, encouraged, recognised, and rewarded for engaging in data access and sharing arrangements.*

It is worth noting that under some legal frameworks, data access and sharing practices have been tied to less stringent regulatory obligations under certain conditions, thereby indirectly incentivising the adoption of these practices. For example, according to Article 2 (Scope), 5g of the EU AI Act:

> *The obligations laid down in this Regulation shall not apply to AI systems released under free and open source licenses unless they are placed on the market or put into service as high-risk AI systems or an AI system that falls under Title II and IV.*

Similarly, Article 28 (Responsibilities along the AI value chain), 2b (emphasis added):

> *The provider of a high risk AI system and the third party that supplies an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI system shall, by written agreement, specify the necessary information, capabilities, technical access and other assistance based on the generally acknowledged state of the art, in order to enable the provider of the high risk AI system to fully comply with the obligations set out in this Regulation.* ***This obligation shall not apply to third parties making accessible to the public tools, services, processes, or AI components other than general-purpose AI models under a free and open licence.***

Recital 57e further specifies that:

> *Third parties making accessible to the public tools, services, processes, or AI components other than general-purpose AI models, shall not be mandated to comply with requirements targeting the responsibilities along the AI value chain, in particular towards the provider that has used or integrated them, when those tools, services, processes, or AI components are made accessible under a free and open licence. …*

While there is ongoing debate about the implications of these conditional exemptions (Castro, 2024[47]), some of which related to the unclear scope of what constitutes a "general-purpose AI models under a free and open licence" and the risk of "open washing" (Liesenfeld and Dingemanse, 2024[48]; White et al., 2024[38]; Bommasani et al., 2024[73]), it is worth recognising that conditional exemptions may be an indirect way to incentive the adoption of desired data governance approaches. The exemptions of "anonymised data" from the scope of privacy and data protection frameworks is another such example, which also demonstrates the regulatory challenges that can emerge from these exemptions.

Box 18. Incentivising data access and sharing in the area of open science

Innovative approaches to increase the recognition and rewards for engaging in data access and sharing arrangements can be observed in the area of open science. This includes in particular the dissemination of scientific data together with data citations, which are increasingly considered as factors for the performance evaluation of scientists and as a requirement for granting public funding.

The OECD Recommendation concerning Access to Research Data from Public Funding highlights a number of measures to recognise and reward the provision of access to, and maintenance of, research data and other research-relevant digital objects from public funding. These measures include:

- Developing criteria for researcher recruitment, advancement, and grant review that take into account the accessibility, quality, and impact of research data,.
- Supporting the development of robust and open indicators on the impact of access to research data, including through the tracking of data and software citations;
- Facilitating giving credit for all contributions to the research endeavour, including data acquisition, curation, analysis, validation, documentation, packaging, and final write-up;
- Promoting data and software citation in academic practice, including the development of data and software citation standards and acknowledgement of data and code creators and maintainers as key contributors.

Source: OECD Recommendation concerning Access to Research Data from Public Funding

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

The availability of data does not guarantee effective use and reuse across organisations, sectors and jurisdictions or even re-used at all. The Recommendation therefore calls on Adherents to consider the following issues to foster effective re-use of trustworthy data across society: (i) Further improving conditions for cross-border data access and sharing with trust; (ii) Fostering the findability, accessibility, interoperability and reusability of data across organisations, including within and across the public and private sectors; and (iii) Enhancing the capacity of all stakeholders to effectively use data responsibly along the data value cycle.

3.3.1. Further improving conditions for cross-border data access and sharing with trust

As data access and sharing increasingly occur across jurisdictions, governments need to consider means that reinforce trust across the global data ecosystem, to further improve conditions for cross-border data access and sharing with trust. This also includes promoting international co-operation, commonalities, complementarities and elements of convergence across legal frameworks to foster legal interoperability. (See Section 2 on “Understanding key concepts”).

Provision VII of the Recommendation aims to “further improve conditions for cross-border data access and sharing with trust”. Given its emphasis on trust, this provision needs to be read in conjunction with Section 1 of the Recommendation on “Reinforcing trust across the data ecosystem” (Provisions III – IV) (see Section 3.1). This means in other words, that all the approaches presented above can be considered by policy makers to further improve conditions for cross-border data access and sharing with trust.

When the Recommendation calls on Adherents to:

> *Assess, and to the extent possible minimise, restrictions to cross-border data access and sharing, in particular for purposes of global public interest, taking into account the need to ensure respect for fundamental rights and vital interests, including the protection of privacy and intellectual property rights and the right to access public information. (VII.a)*

This should be read in combination with provision V the Recommendation that:

> *Recommends that Adherents seek to maximise the benefits of measures for enhancing data access and sharing, while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives, alongside broader efforts to promote and enable a culture of responsibility for data governance throughout the data value cycle. (V)*

The approaches put forward by the Recommendation under provision V thus also apply in the context of cross-border data access and sharing. This includes in particular provision V.a that recommends to "encourage data access and sharing arrangements that ensure that data are as open as possible to maximise their benefits and as closed as necessary to protect legitimate public and private interests".

Provision VII.b provides further clarity on the aspects that are specific to cross-border data access and sharing, when it recommends to:

> *Ensure that measures that condition cross-border data access and sharing are non-discriminatory, transparent, necessary, and proportionate to the level of risk, taking into account, among others, the sensitivity of the data, the purpose and context of data access, sharing, and use, and the extent to which measures are in place to enforce accountability irrespective of the jurisdiction in which the data is stored. (VII.b)*

The concept of non-discrimination is used in the Recommendation when referring to "Non-discriminatory data access and sharing arrangements", which are defined as "a specific type of data access and sharing arrangement, where data can be accessed and shared, free of charge or for fees, based on terms that are independent of the data users' identities."

As highlighted in Section 3.1.1.3, the concept of transparency is well established in the context of privacy and data protection and articulated in the *OECD Privacy Guidelines*' principles on Openness which recommends that "There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller."

Similarly, as highlighted in Section 3.1.3, the concept of necessity and proportionality is articulated in paragraph 18 of Part Four of the *OECD Privacy Guidelines* recommending that "Any restrictions to transborder flows of personal data should be proportionate to the risks presented, taking into account the sensitivity of the data, and the purpose and context of the processing."

Further clarity on the specific aspects to cross-border data access and sharing are also provided by provision VII.c which recommends to:

> *Promote continued dialogue and international co-operation on ways to foster data access and sharing across jurisdictions – including through the implementation of trust-enhancing measures as set out above – as well as the interoperability and mutual recognition of data access and sharing arrangements, taking into account applicable legal requirements and global standards. (VII.c)*

Provision VII.c, which explicitly refers to Section 1 of the Recommendation on "Reinforcing trust across the data ecosystem" (Provisions III – IV) emphasises the need for cross border dialogue and international co-operation as well as the interoperability and mutual recognition of data access and sharing arrangements as discussed in Section 2.3.3.

Box 19. Improving conditions for cross-border data access and sharing with trust

Nordic Cooperation

The Nordic Council of Ministers is an important arena of Nordic cooperation. Under the Norwegian presidency of the Nordic Council of Ministers in 2017, a common Nordic-Baltic ministerial declaration on digitalisation – Digital North -– was prepared and adopted. A separate Nordic Council of Ministers for Digitalisation (MR-DIGITAL) was established at the same time. MR-DIGITAL will, among other things, support the exchange and use of digital services across national borders in the region, as well as joint efforts to promote 5G, artificial intelligence and data sharing. Similarly, the Cross Border Digital Services (CBDS) Programme, a strategic initiative from the Nordic Council of Ministers (NCM), will increase mobility and integration across the region through the development and deployment of cross-border digital services with the objective to benefit citizens, businesses, and public authorities in the region.

Estonia-Finland data exchange

Shared data governance arrangements can also help to improve cross-border public service delivery tapping on trustworthy data exchange. In 2013 Estonia and Finland agreed on the development of a common agenda to support the implementation cross-border digital services, which in consequence enabled the deployment of Estonia's X-Road data-sharing in Finland. Later in 2018, the integration of these platforms led to greater, automated and secured cross-border data sharing, thus benefiting service users and supporting the future development of additional cross-border services in the region. Yet, success did not only rely on technical issues for in 2017 Estonia and Finland agreed on the creation of the Nordic Institute for Interoperability and Solutions, which "ensure(s) the development and strategic management of the X-Road and other cross-border components for eGovernment infrastructure" (NIIS, 2019).

Source: (OECD, 2019[21])

3.3.2. Fostering the findability, accessibility, interoperability and reusability of data across organisations, including within and across the public and private sectors

Government action may be needed where the lack of common standards (or the proliferation of incompatible standards) for data format (syntax) and data models (semantics) present a barrier to data access and sharing. Furthermore, data quality standards need to be defined and relevant information on data quality need to be provided with regards to the original context of data collection and use. This is because the information that can be extracted from data depends on the quality of the data. Poor-quality data will almost always lead to poor data re-use, analysis and results. At the same time, data quality will typically depend on the intended use of the data: data that are of adequate quality for certain applications can be of insufficient quality for other applications.

The Recommendation addresses these issues with a focus on the FAIR (findability, accessibility, interoperability and reusability) principles. It recognises (in the preamble):

> *that effective and efficient data access and sharing often depends on machine-readability and interoperable specifications including common licensing arrangements, standards, and metadata that enable findability, accessibility, interoperability, reusability, and the correct interpretation and analysis of data, and that standard-*

> *setting organisations and industry consortia as well as open source play a critical role in the development and adoption of these interoperable specifications;*

It then addresses in particular two aspects related to FAIR, namely the need to:

> *Strive to ensure that data are provided together with any required meta-data, documentation, data models and algorithms in a transparent and timely manner, supported by appropriate data access control mechanisms, including application programming interfaces (APIs). (VIII.a)*

And the need to:

> *Assess and, whenever possible, promote the development and adoption of interoperable specifications for effective data access, sharing, and use, including common standards for data formats and models as well as open source implementations. (VIII.b)*

The Recommendation emphasises the role of stakeholders, including standard-setting organisations, to promote the development and adoption of interoperable specifications through "open, accessible, voluntary, and consensus-based efforts" and to raise awareness about the benefits of these specifications. There is a significant number of government initiatives aimed at fostering the findability, accessibility, interoperability and reusability of data, and public sector data in particular (see Box 20).

Box 20. Examples of government initiatives fostering the findability, accessibility, interoperability and reusability of data

The **Australian** Code for the Responsible Conduct of Research 2018 and the associated Management of Data and Information in Research guide requires that researchers should adhere to established national and international standards for data description and structuring to facilitate tracking of references. These standards include using Digital Object Identifiers (DOIs) for datasets, ORCID IDs for researchers, and standard terminology for scientific concepts. Published research data generally require some kind of online description (i.e. metadata) and should be findable, accessible, interoperable, and re-usable, both manually and with automated tools. This requires researchers to include appropriate context (descriptive, technical, methodological, access, and provenance information). The ARC is furthermore considering the use of Crossref global grant DOIs for all ARC-funded projects which would increase findability and linking of project data.

Canada's Standard on Service and Digital contains a recently updated "standard on systems that manage information and data", which outlines the requirements for systems to support the sharing, creation, collection, accessibility, and interoperability of data. To facilitate improved interoperability and reusability of open data, the Government of Canada has added various new features to Open.Canada.ca. These include in particular:

- Canadian Government has implemented a new tool to release open datasets, which automates data quality validation processes to ensure that tabular data is structured correctly. This ensures that tabular data can be used by others with reduced costs for data cleaning and reshaping.
- Canadian Government is able to load those validated datasets into a publicly accessible data warehouse with modern features such as Web APIs, automated file format conversions, integrated data dictionaries and the ability to preview the data on screen before downloading. This ensures that data is rapidly reusable in various formats and consumption methods.

Box 20. Examples of government initiatives fostering the findability, accessibility, interoperability and reusability of data (cont.)

Italy has implemented a number of initiatives to fostering the findability, accessibility, interoperability and reusability of public sector data. These include in particular:

- The National Digital Data Platform, which has been implemented with the goal to enable the fulfilment of the "once-only" principle of the "Single Digital Gateway" directive of the European Union and to ensure the full interoperability of key datasets and services across central and local public entities as well as to enable the harmonisation with other EU countries of the service procedures prioritised by the "Single Digital Gateway" directive.
- The development of a National Data Catalog for semantic interoperability, which is being funded with EUR 10.7 million in total. The main purpose of the Catalog is to provide a one-stop shop for ontologies, controlled vocabularies and data schemas in the public sector to facilitate the findability, accessibility, interoperability and reusability of national semantic assets. Public bodies will be able to publish their semantic assets on the platform, while developers will be able to use them to develop semantically and syntactically interoperable digital public services. The National Institute of Statistics is the central public agency in charge of developing and maintaining the Catalog.
- The National Digital Data Platform (PDND) is a central catalogue of APIs shared across central and local administrations for the interoperability between base registries. Public and private entities will be able to use in an authorised and certified way the published APIs (e-services) in compliance with EU privacy laws. Semantic and syntactic interoperability of the e-services will be supported by the National Data Catalogue for Semantic Interoperability (NDC).

Chile has established the Chilean National Agency for Research and Development (ANID) in 2018, which is currently improving upon and promoting usage of its institutional repository 'repistorio.anid.cl'. In order to ensure and promote adherence with FAIR principles, ANID is undertaking the following actions:

- Improving findability by encouraging the deposit of research results of projects financed by the agency in its repository with DOIs and it is planned to automatically assign DOIs to deposited results without such an identifier. ANID has launched a communications campaign to promote adherence to its Open Access Policy among universities by laying out the various components and obligations of open access practices and standards.
- ANID is designing a "national node for access to scientific information" which aims to improve accessibility to metadata of digital objects deposited in university or its own repositories. Such data may be retrieved using persistent identifiers.
- The "national node for access to scientific information" seeks to improve interoperability by following LA Referencia standards which universities must follow (as a minimum) too and which the ANID repository is already in compliance with. In addition, the ANID Open Data Policy makes use of the controlled vocabulary used by the Confederation of Open Access Repositories (COAR) for document types subject to deposit requirements and thereby contributes to interoperability between repositories.

- In order to enhance data re-usability, ANID makes uses different licenses for data use and seeks to promote the Creative Commons license. The "national node for access to scientific information" will furthermore develop standards for the description of metadata.

Box 20. Examples of government initiatives fostering the findability, accessibility, interoperability and reusability of data (cont.)

The Netherlands' National Programme Open Science (NPOS) organises and coordinates a FAIR data table for the purpose of building a national FAIR-compliant federated data ecosystem which provides for access to data without unnecessary restrictions across scientific domains and society. This ecosystem is envisioned to a constantly and collaboratively evolving framework based on a well-supported digital infrastructure for interoperable FAIR data from local research-performing and -supporting organisations to improve data (re-)usability. It also involves a highly professional community of data stewards in order to facilitate the local implementation of quality machine-actionable FAIR data and the associated metadata.

In the United Kingdom, the Data Standards Authority (DSA) play an important role in a cross-government proposal developed by the Cabinet Office, the Office for National Statistics (ONS) and the DCMS to address the government's data foundations. The DSA's vision is to use data standards to improve public services for the public through stronger policies, analysis and insights by ensuring that data can be easily found, accessed, shared responsibly and combined. The focus of the DSA's work is to mitigate data linkage risks and issues by adopting data standards when users need to share data within or between the public and private sector. It is envisioned that DSA takes the following actions:

- To improve data interoperability, the DSA builds cross-government consensus on defining and agreeing standards. This improves collaboration, encouraging experts across sectors to consider problems and risks in detail.
- The DSA also considers how to implement new controls to make sure data standards are adopted and supported by an assurance process. The DSA looks for opportunities and incentives to increase adoption of standards across the wider public sector.
- To assess and articulate the benefits of adopting data standards, the DSA forms processes to help identify and showcase implementation of standards, or pilots of new standards to demonstrate the value of adopting them.

Brazil's Digital Government Strategy seeks to integrate data and services across all administrative levels of the country, while reducing operational costs, expanding the offer of digital services and reducing administrative burdens for citizens. Along these lines, the government seeks to implement a "data interoperability bus" that facilitates the integration and reuse of this information for the provision of e-government services to citizens. In 2019, the first information reference register was created, the Citizen Base Register, providing agencies with a reliable source of consultation for fundamental registration data and identification for the service, exempting the redundant supply of information at every interaction. Since then, the issuance of passports and other official documents in electronic format and the electronic exchange of information on electoral and military discharges has been accelerated significantly.

FAIR data and AI

FAIR data and data quality are becoming increasingly a critical source of concerns with the rise of generative AI models, including LLMs, which rely significantly on data scraping methods. As highlighted in Section 3.1.1.3 on Transparency, there is compelling evidence that data provenance remains a key source of opacity not only in the context of AI. After reviewing more than 1,800 of the widely used datasets, the Data Provenance Initiative found that 70 percent of the datasets lacked information about licensed uses or stated more permissive uses (Data Provenance, 2024[87]).[45] This is a major concern not only from a transparency perspective, but also for trust, trustworthiness and fairness in AI.

AI data and model opacity also risk undermining the benefits, and exacerbating the risks, of open (source) AI. These benefits rely on AI (system and model) interoperability which in turn requires a sufficient understanding about the relevant components of the AI system (input, output and AI model), including data provenance and the associated legal obligations, so that they can be accessed, shared and re-used more effectively and with a sufficient level of legal certainty.

An important development in this respect has been the introduction of data statements/sheets (McMillan-Major, Bender and Friedman, 2024[28]) and model cards (Mitchell et al., 2019[29]). "These resources have seen considerable uptake in the scientific community, though their adoption by for-profit entities lags behind" (Liesenfeld, Lopez and Dingemanse, 2023[14]) These formal documents can assist AI model developers in recording the procedures for curating, distributing, and maintaining a dataset or AI model, thus enabling their users to critically assess the underlying assumptions, potential risks and harms, and the possibilities for wider application. It is therefore no coincidence that the provision of data and model cards has become a legal requirement under the EU AI Act under certain conditions.[46] And it remains to be seen to what extend this legal requirement may become a standard policy that encourages similar data governance practices in other policy domains beyond AI.

While metadata, data statements/sheets and model cards can contribute to enhancing the reusability of AI models and their interoperability, questions remain on how to enhance the findability and to some extent also the accessibility of AI models and AI input data. In this context, AI data aggregators such as Common Crawl (2024[88]), LAION (2024[89]), EleutherAI (2024[90]) and Hugging Face (2024[91]) can make scraped AI data and/or open source third party datasets available.[47] And some such as Hugging Face (2024[91]) have become a key facilitator for the findability of AI models as well, thus contributing to what could be refered to as FAIR AI models (Box 21).

Box 21. Towards FAIR AI model

In analogy to the FAIR Data Principles (Wilkinson et al., 2016[59]), FAIR AI models could be understood as follows:

Findable: Metadata about AI models should be easily found to aid in the search process of AI models. It should be provided in machine-readable format for automatic discovery by both humans and computers. A unique and persistent identifier for AI models, such as DOI may be required.

Accessible: AI models, once found, should be easy to access, possibly through a well-defined open protocol, including e.g. direct download or an API. Access to such protocol should be free of charge, and to be used by anyone for any purpose subject, at most, to requirements that preserve integrity, provenance, attribution, and openness of the model.

Interoperable: AI models should be interoperable with other AI models and systems. This requires the use of standard formats for AI models such as TensorFlow, Keras, PyTorch or ONNX. This will require

access to metadata and documentations including, but not limited to, data statements/sheets and model cards.

Reusable AI models: The ultimate goal of FAIR is to optimise the reuse, in this case of the AI model. To achieve this, metadata and documentations is also required together with clear licensing agreements.

3.3.3. Enhancing the capacity of all stakeholders to use data more effectively and responsibly

3.3.3.1. Awareness raising and skills development

Better data-related skills are needed along the whole value cycle of data to assure the effective sharing and (re-)use of data. Data holders need data management and data curation capabilities to assure the long-term quality and availability of data. Data users, on the other hand, need adequate digital and data analytics skills to effectively re-use data. Evidence also shows that skills and competences can improve awareness of the actual risks of data access and sharing. The Recommendation therefore first calls on Adherents to:

> *Foster awareness about the benefits and risks of data access, sharing, and use to encourage responsible data governance throughout the data value cycle by engaging in dialogues with all relevant stakeholder groups and partnerships. To this effect, Adherents should disseminate good practices on data access, sharing, and use that help address barriers to accessing and sharing data responsibly and increase the capacity of individuals and organisations to manage, access, share, and use data responsibly.(IX.a)*

Provision IX.b then recommends that Adherents:

> *Promote the development of the data-related skills and competencies needed, including by workers and public servants, to harness the benefits of data access, sharing, and use throughout the data value cycle in a manner consistent with the strategic approach to data access and sharing as set out above. This should include promoting data literacy among the public and increasing citizen's capacity to understand relevant data governance issues and exert their rights.*

There are a significant number of government initiatives aimed at enhancing stakeholders' capacity to use data more effectively and responsibly (see Box 22).

3.3.3.2. Adoption of data-related information and communication technology infrastructures

Governments may also need to support the development of the information and communication technology (ICT) infrastructures needed for data storage, processing and analytic infrastructures as access to data storage, processing and analytic infrastructures is a major condition for the effective re-use of data across society. This is particularly critical for SMEs, but also for individuals including scientists as it involves the cost of operating, maintaining as well as scaling data infrastructure.

The Recommendation therefore calls on Adherents to:

> *Facilitate access to and the adoption of the sustainable, open, scalable, safe, and secure foundational infrastructures needed along the data value cycle, including for connectivity, storage, and computing, by promoting digital security risk management practices throughout the data value cycle, incentivising investments in and the adoption of such infrastructures across the data ecosystem, and leveraging PPPs where practicable and appropriate.*

Box 22. Examples for measures to enhance stakeholders' capacity to use data more effectively and responsibly

Belgium: One of the four pillars of the Flemish Data Strategy focusses on 'data literacy'. This is split up in three subdimensions:

- Awareness development and creation of support for a better data governance via campaigns and the organisation of knowledge sessions for middle management
- Deepening and broadening practical data skills by educational programs
- Capacity enlargement via collaboration and learning tracks with external partners, employer branding initiatives towards data experts and the creation of a talent pool of experts that is broadly applicable in different entities of the Flemish Government (Vlaanderen Connect).

Latvia will establish a national Data Stewards program within its higher education curriculum to enhance the skills of researchers to manage and share their research data. Training of employees and customers for effective data sharing and use is also planned in Agricultural Data Center. Latvia is also working on consolidation of state information and communication technology resources and competencies and use of cloud computing services in public administration. Latvia plans to consolidate the ICT infrastructure of the public administration by setting up public administration Federated Cloud by placing it provisionally in four data centres. It will be done using Recovery and Resilience Facility implementing four projects:

- Development of cloud computing services of the Information Center of the Ministry of the Interior within the framework of the state federated cloud;
- Development of cloud computing services of the National Library of Latvia within the state federated cloud; and
- Development of cloud computing services of the Latvian State Radio and Television Center within the state federated cloud.

Singapore: IMDA launched the Better Data-Driven Business programme (or BDDB) to provide free tools and guidance to help businesses better safeguard their customers' personal data while making more effective use of data. The BDDB programme provides businesses, particularly SMEs, with free tools and guidance to use their data responsibly to drive business growth.

- It offers a free plug-and-play business intelligence (BI) tool for businesses to convert their data into interactive visual dashboards, which will help them glean business insights to achieve common business objectives. A step-by-step guide is also provided on how to use the BI tool and interpret the data for insights to develop actionable business plans.
- To ensure customers' personal data are safeguarded, IMDA worked with PDPC to incorporate good data protection practices in the tool, such that only data that are necessary for the insights are used, and the data is generally pseudonymised (such as using customer ID instead of names). Additional guidance is provided for related actions, such sending of marketing messages. For SMEs who are using point-of-sales or human resource management systems from several IMDA's SMEs Go Digital solution providers, they will find their data already mapped to the requirements of the BI tool and can easily export their data from these systems to the tool

to generate the insights. For SMEs that already collect and use data regularly, the BDDB programme equips them with the knowledge and tools to use data responsibly for wider and more complex use cases.

References

[Former] Article 29 Data Protection Working Party (2017), *Guidelines on the right to data portability under Regulation 2016/679, WP242 rev.01*, https://ec.europa.eu/newsroom/dae/redirection/document/44099 (accessed on 7 June 2024). [31]

Abate, C., G. Bianco and F. Casalini (2024), *The intersection between competition and data privacy*, OECD Publishing, Paris, https://doi.org/10.1787/20758677. [79]

Attrey, A., M. Lesher and C. Lomax (2020), "The role of sandboxes in promoting flexibility and innovation in the digital age", *OECD Going Digital Toolkit Notes*, No. 2, OECD Publishing, Paris, https://doi.org/10.1787/cdf5ed45-en. [77]

Australian Government (2022), *Consumer Data Right - Strategic Assessment Outcomes*, https://treasury.gov.au/publication/p2022-242997. [63]

Australian Research Council (2018), *Australian Code for the Responsible Conduct of Research 2018*, https://www.arc.gov.au/about-arc/program-policies/research-integrity/australian-code-responsible-conduct-research-2018 (accessed on 6 June 2024). [99]

Bar-Sinai, M., L. Sweeney and M. Crosas (2016), "DataTags, Data Handling Policy Spaces and the Tags Language", https://doi.org/10.1109/SPW.2016.11. [93]

Benhamou, Y. (2024), "Open Source AI – definition and selected legal challenges", https://copyrightblog.kluweriplaw.com/2024/04/15/open-source-ai-definition-and-selected-legal-challenges/#content (accessed on 5 June 2024). [40]

Benjamin, M. et al. (2019), "Towards Standardization of Data Licenses: The Montreal Data License", https://arxiv.org/abs/1903.12262v1 (accessed on 10 June 2024). [101]

Bernstein, F. et al. (1977), "The protein data bank: A computer-based archival file for macromolecular structures", *Journal of Molecular Biology*, Vol. 112/3, pp. 535-542, https://doi.org/10.1016/S0022-2836(77)80200-3. [15]

Bommasani, R. et al. (2024), "The Foundation Model Transparency Index v1.1 May 2024", https://crfm.stanford.edu/fmti (accessed on 5 June 2024). [73]

Casalini, F. and J. López González (2019), "Trade and Cross-Border Data Flows", *OECD Trade Policy Papers*, No. 220, OECD Publishing, Paris, https://doi.org/10.1787/b2023a47-en. [46]

Castro, D. (2024), *The EU's AI Act Creates Regulatory Complexity for Open-Source AI – Center for Data Innovation*, https://datainnovation.org/2024/03/the-eus-ai-act-creates-regulatory-complexity-for-open-source-ai/ (accessed on 5 June 2024). [47]

Cerf, V. (2024), "Thoughts on AI Interoperability", *Communications of the ACM*, Vol. 67/4, p. 5, https://doi.org/10.1145/3649475. [60]

Cohen, N. and C. Wendehorst (2021), *ALI-ELI Principles for a Data Economy: Data Transactions and Data Rights*, American Law Institute, European Law Institute, https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ALI-ELI_Principles_for_a_Data_Economy_Final_Council_Draft.pdf. [82]

Common Crawl (2024), *Common Crawl*, https://commoncrawl.org/ (accessed on 27 May 2024). [88]

Creative Commons (2024), *Frequently Asked Questions*, https://creativecommons.org/faq/#Can_I_apply_a_Creative_Commons_license_to_databases.3F. [86]

Data Provenance (2024), *Data Provenance Explorer*, https://www.dataprovenance.org/ (accessed on 27 May 2024). [87]

Deloitte (2013), *Market Assessment of Public Sector Information: A Report to the Department for Business Innovation and Skills*. [6]

Deng, J. et al. (2010), "ImageNet: A large-scale hierarchical image database", pp. 248-255, https://doi.org/10.1109/CVPR.2009.5206848. [13]

Determann, L. (2018), "No One Owns Data", *UC Hastings Research Paper* No. 265, https://doi.org/10.2139/ssrn.3123957. [54]

EleutherAI (2024), *EleutherAI*, https://www.eleuther.ai/ (accessed on 27 May 2024). [90]

European Commission (2022), *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act)*, https://eur-lex.europa.eu/legal-content/en/ALL/?uri=COM:2022:68:FIN (accessed on 7 October 2022). [103]

European Commission (2017), *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions European Interoperability Framework – Implementation Strategy*, https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF (accessed on 17 March 2022). [92]

European Union (2023), *REGULATION (EU) 2023/2854 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202302854 (accessed on 7 May 2024). [71]

European Union (2022), *European Declaration on Digital Rights and Principles*, https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles (accessed on 6 June 2024). [100]

European Union (2016), *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679 (accessed on 29 September 2022). [30]

European Union (2012), *REGULATION (EU) No 1025/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council*, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012R1025#d1e32-29-1 (accessed on 10 June 2024). [70]

Frischmann, B., M. Madison and K. Strandburg (eds.) (2014), *Governing Knowledge Commons*, Oxford University Press. [67]

Gierten, D. and M. Lesher (2022), "Assessing national digital strategies and their governance", *OECD Digital Economy Papers*, No. 324, OECD Publishing, Paris, https://doi.org/10.1787/baffceca-en. [76]

GPAI (2023), *Fostering Contractual Pathways for Responsible AI Data and Model Sharing for Generative AI and Other AI Applications*, Global Partnership on AI, https://gpai.ai/projects/responsible-ai/IC_Intellectual%20Property%20project.pdf. [81]

GPAI (2022), *Protecting AI innovation, Intellectual Property (IP): GPAI IP Expert: (I) Guidelines for Scraping or Collecting Publicly Accessible Data and (II) the Preliminary Report on Data and AI Model Licensing*, Global Partnership on AI, https://gpai.ai/projects/innovation-and-commercialization/intellectual-property-expert-preliminary-report-on-data-and-AI-model-licensing.pdf. [80]

Hauser, M. (2023), *Machine Learning Frameworks in Open-Source Software: An Exploratory Study on Code and Project Smells*, University of Stuttgart. [37]

Hugging Face (2024), *Datasets: imagenet-1k*. [91]

ISO (2017), "ISO/IEC 19941:2017(en) Information technology – Cloud computing – Interoperability and portability", webpage, https://www.iso.org/obp/ui/#iso:std:iso-iec:19941:ed-1:v1:en (accessed on 9 March 2021). [55]

IWGDPT (2024), *Working Paper on Large Language Models*, Federal Commissioner for Data Protecfion and Freedom of Information (BfDI), Bonn, https://www.bfdi.bund.de/SharedDocs/Downloads/EN/Berlin-Group/20241206-WP-LLMs.pdf?__blob=publicationFile&v=1. [35]

Janapa Reddi, V. (ed.) (n.d.), *Machine Learning Systems with TinyML*, Harvard University, https://harvard-edge.github.io/cs249r_book/ (accessed on 5 June 2024). [36]

Jiang, A. et al. (2024), "Mixtral of Experts", https://arxiv.org/abs/2401.04088v1 (accessed on 6 June 2024). [62]

Kush, R. et al. (2020), "FAIR data sharing: The roles of common data elements and harmonization", *Journal of Biomedical Informatics*, Vol. 107, https://doi.org/10.1016/J.JBI.2020.103421. [56]

LAION (2024), *LAION*, https://laion.ai/ (accessed on 27 May 2024). [89]

Lateral Economics (2014), *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, http://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FN. [7]

Leigh Dodds (2013), *Publisher's Guide to Open Data Licensing*, ODI, https://theodi.org/insights/guides/publishers-guide-to-open-data-licensing/ (accessed on 5 June 2024). [44]

Liesenfeld, A. and M. Dingemanse (2024), *Rethinking open source generative AI: open-washing and the EU AI Act*, https://www.mpi.nl/publications/item3588217/rethinking-open-source-generative-ai-open-washing-and-eu-ai-act (accessed on 5 June 2024). [48]

Liesenfeld, A., A. Lopez and M. Dingemanse (2023), "Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators", *Proceedings of the 5th International Conference on Conversational User Interfaces, CUI 2023*, https://doi.org/10.1145/3571884.3604316. [14]

Linux Foundation (2017), *Linux Foundation Debuts Community Data License Agreement*, http://www.linuxfoundation.org/press/press-release/linux-foundation-debuts-community-data-license-agreement. [84]

Lorenz, P., K. Perset and J. Berryhill (2023), "Initial policy considerations for generative artificial intelligence", in *OECD Artificial Intelligence Papers*, OECD Publishing, Paris, https://www.oecd-ilibrary.org/science-and-technology/initial-policy-considerations-for-generative-artificial-intelligence_fae2d1e6-en (accessed on 5 June 2024). [34]

McKinsey Global Institute (2013), *Open Data: Unlocking innovation and performance with liquid information*. [5]

McMillan-Major, A., E. Bender and B. Friedman (2024), "Data Statements: From Technical Concept to Community Practice", *ACM Journal on Responsible Computing*, Read_Status: To Read Read_Status_Date: 2024-05-31T23:21:13.841Z, pp. 1:1–1:17, https://doi.org/10.1145/3594737. [28]

Metz, C. et al. (2024), *How Tech Giants Cut Corners to Harvest Data for A.I.*, https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html (accessed on 27 May 2024). [98]

Microsoft (2019), *Removing Barriers to Data Innovation: Empowering people and organizations to share and use data more effectively*, https://3er1viui9wo30pkxh1v2nh4w-wpengine.netdna-ssl.com/wp-content/uploads/prod/sites/560/2019/12/Backgrounder-FAQ-Sheet_FINAL.pdf (accessed on 10 May 2020). [83]

Mitchell, M. et al. (2019), *Model Cards for Model Reporting*, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3287560.3287596. [29]

Naderalvojoud, B. and T. Hernandez-Boussard (2023), "Improving machine learning with ensemble learning on observational healthcare data", *AMIA Annual Symposium Proceedings*, Vol. 2023, p. 521, http:///pmc/articles/PMC10785929/ (accessed on 6 June 2024). [61]

Network of the National Library of Medicine [United States] (n.d.), *Data Glossary - Data Interoperability*, National Library of Medicine, https://www.nnlm.gov/guides/data-glossary/data-interoperability (accessed on 6 June 2024). [57]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. [27]

OECD (2023), "Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI", *OECD Digital Economy Papers*, No. 349, OECD Publishing, Paris, https://doi.org/10.1787/2448f04b-en. [51]

OECD (2023), "AI language models : Technological, socio-economic and policy considerations", *OECD Digital Economy Papers*, No. No. 352, OECD Publishing, Paris, https://www.oecd-ilibrary.org/science-and-technology/ai-language-models_13d38f92-en (accessed on 5 June 2024). [33]

OECD (2023), *Recommendation of the Council on the Governance of Digital Identity*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0491 (accessed on 24 October 2023). [75]

OECD (2022), *Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions*, Building Trust in Public Institutions, OECD Publishing, Paris, https://doi.org/10.1787/b407f99c-en. [66]

OECD (2022), *Data Shaping Firms and Markets*, OECD Publishing, Paris, https://doi.org/10.1787/7b1a2d70-en. [105]

OECD (2022), *Declaration on Government Access to Personal Data Held by Private Sector Entities*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0487 (accessed on 3 October 2023). [96]

OECD (2022), "Emerging privacy enhancing technologies: Opportunities and challenges based on the OECD Privacy Guidelines' Principles". [74]

OECD (2022), *Financing Growth and Turning Data into Business: Helping SMEs Scale Up*, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://doi.org/10.1787/81c738f0-en. [97]

OECD (2022), *Going Digital Guide to Data Governance Policy Making*, OECD Publishing, Paris, https://doi.org/10.1787/40d53904-en. [25]

OECD (2022), *Going Digital to Advance Data Governance for Growth and Well-being*, OECD Publishing, Paris, https://doi.org/10.1787/e3d783b0-en. [24]

OECD (2022), "Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint", in *OECD Digital Economy Papers*, OECD Publishing, https://www.oecd.org/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications-7babf571-en.htm (accessed on 7 June 2024). [17]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [32]

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. [1]

OECD (2022), *The evolving concept of market power in the digital economy: Background note by the Secretariat*, https://one.oecd.org/document/DAF/COMP(2022)5/en/pdf. [78]

OECD (2021), *Good Practice Principles for Data Ethics in the Public Sector*, OECD Publishing, Paris. [69]

OECD (2021), *Mapping Data Portability Initiatives, Opportunities and Challenges*, https://www.oecd-ilibrary.org/docserver/a6edfab2-en.pdf?expires=1646401600&id=id&accname=ocid84004878&checksum=225184A31C46ED7EDAC34CD3DFB1BACA. [42]

OECD (2021), "OECD Good Practice Principles for Data Ethics in the Public Sector", *OECD Public Governance Policy Papers*, No. 57, OECD Publishing, Paris, https://doi.org/10.1787/caa35b76-en. [23]

OECD (2021), *OECD-Global Privacy Assembly Online Workshop: "One Year Later: Addressing the Data Governance and Privacy Implications of the COVID-19 Pandemic and the Road to Recovery*, https://one.oecd.org/document/DSTI/CDEP/DGP(2021)12/FINAL/en/pdf. [10]

OECD (2021), *Recommendation of the Council concerning Access to Research Data from Public Funding*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347 (accessed on 13 March 2022). [102]

OECD (2021), *Recommendation of the Council on Competitive Neutrality*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0462 (accessed on 4 October 2022). [72]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463 (accessed on 6 March 2023). [41]

OECD (2021), *Report on the Implementation of the Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, https://one.oecd.org/document/C(2021)42/en/pdf. [52]

OECD (2020), *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, OECD Publishing, Paris, https://doi.org/10.1787/947717bc-en. [22]

OECD (2020), "Ensuring data privacy as we battle COVID-19", https://www.oecd.org/coronavirus/policy-responses/ensuring-data-privacy-as-we-battle-covid-19-36c2f31e/ (accessed on 10 March 2021). [8]

OECD (2020), *OECD Digital Economy Outlook 2020*, https://www.oecd-ilibrary.org/docserver/bb167041-en.pdf?expires=1644491639&id=id&accname=ocid84004878&checksum=9835C6E4364EFB6DE8F0F8B6B9C5C65D. [3]

OECD (2020), "Summary of discussion of the roundtable on Consumer data rights and competition: Annex to the Summary Record of the 133rd Meeting of the Competition Committee, held virtually on 10-16 June 2020". [65]

OECD (2020), "Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics", http://www.oecd.org/coronavirus/policy-responses/tracking-and-tracing-covid-protecting-privacy-and-data-while-using-apps-and-biometrics-8f394636/ (accessed on 10 March 2021). [12]

OECD (2020), *Why open science is critical to combatting COVID-19*, https://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combatting-covid-19-cd6ab2f9/. [9]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use Across Societies*, https://www.oecd-ilibrary.org/docserver/276aaca8-en.pdf?expires=1644917622&id=id&accname=ocid84004878&checksum=244E5EB48309BE972654F87847D735C0. [2]

OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/059814a7-en. [21]

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. [20]

OECD (2016), *Research Ethics and New Forms of Data for Social and Economic Research*. [68]

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264229358-en. [4]

OECD (2015), *Digital Security Risk Management for Economic and Social Prosperity*, https://www.oecd.org/digital/ieconomy/digital-security-risk-management.pdf. [50]

OECD (2015), *Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity*. [94]

OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188 (accessed on 21 May 2020). [53]

OECD (2013), *The evolving privacy landscape: 30 years after the OECD Privacy Guidelines*, OECD Publishing, https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf. [104]

OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, https://doi.org/10.1787/9789264034020-en-fr. [19]

OECD (forthcoming), "Benefits and risks of open-sourcing advanced foundation models", *OECD Artificial Intelligence Papers*, OECD Publishing, Paris. [16]

OECD (n.d.), *OECD Legal Framework*, http://www.oecd.org/en/about/legal.html (accessed on 1 October 2024). [26]

OECD-Harvard Global Health Institute (2017), *Expert Consultation on "Mobile Technologies Based Services for Global Health and Wellness: Opportunities and Challenges*, https://one.oecd.org/document/DSTI/CDEP/SPDE(2017)10/en/pdf. [11]

Open Knowledge Foundation (n.d.), *Open Data Commons*, https://opendatacommons.org/. [85]

Open Knowledge Foundation (n.d.), *The Open Data Handbook*, http://opendatahandbook.org/ (accessed on 5 June 2024). [45]

Open Source Initiative (n.d.), *The Open Source AI Definition – draft v. 0.0.8*, https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-8 (accessed on 5 June 2024). [39]

Paic, A. (2021), "Open science - Enabling discovery in the digital age", *OECD Going Digital Toolkit Notes*, No. 13, OECD Publishing, Paris, https://doi.org/10.1787/81a9dcf0-en. [18]

Perez Rivera, A., C. Emilsson and B. Ubaldi (2020), *OECD Open, Useful and Re-usable data (OURdata) Index: 2019*. [95]

Reimsbach-Kounatze, C. and A. Molnar (2024), "The impact of data portability on user empowerment, innovation, and competition", *OECD Going Digital Toolkit Notes*, No. 25, OECD Publishing, Paris, https://doi.org/10.1787/319f420f-en. [43]

Robinson, L., K. Kizawa and E. Ronchi (2021), "Interoperability of privacy and data protection frameworks", *Going Digital Toolkit Note* 21, https://goingdigital.oecd.org/data/notes/No21_ToolkitNote_Privacy&DataInteroperability.pdf. [58]

Solaiman, I. (2023), "The Gradient of Generative AI Release: Methods and Considerations", *ACM International Conference Proceeding Series*, pp. 111-122, https://doi.org/10.1145/3593013.3593981. [49]

Streel, A., J. Kramer and P. Senellart (2020), *Making data portability more effective for the digital economy*, https://cerre.eu/publications/report-making-data-portability-more-effective-digital-economy/. [64]

White, M. et al. (2024), "The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence", https://arxiv.org/abs/2403.13784v3 (accessed on 5 June 2024). [38]

Wilkinson, M. et al. (2016), "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data*, Vol. 3, https://doi.org/10.1038/sdata.2016.18. [59]

Endnote

[1] To collect implementation examples, a questionnaire was circulated to delegates of the OECD Working Party on Data Governance and Privacy (DGP) on 19 May 2022, which 15 OECD Members and partner economies responded to. Additional implementation examples were collected through work undertaken in the context of Phase III of the OECD Going Digital project focusing on data governance for growth and well-being. This includes in particular work on the OECD (2022[25]) *Going Digital Guide to Data Governance Policy Making* and the OECD (2022[97]) project on SME data governance. Overall, Members and partner economies covered include: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Türkiye, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States, as well as Brazil and Singapore, plus the European Union.

[2] In the public sector, in particular, enhancing data access and sharing can contribute to operationalise digital government principles such as *once only,* thus reducing the burden on citizens and businesses when interacting with public agencies.

[3] Data-intensive science is seen as the fourth paradigm of science, and has strongly boosted many scientific disciplines.

[4] For example, when OpenAI had reportedly exhausted its text input data, it developed a speech recognition tool that was said to transcribe more than 1 million hours of YouTube videos to create AI input data, even though YouTube's terms of use prohibit using its videos for independent applications (Metz et al., 2024[98]).

[5] The development involved a Joint Steering Group (JSG) of experts nominated by the three Committees (DPC, CSTP and PGC) to support the work. The JSG included more than 90 experts, representing over 30 OECD Member and partner economies as well as Business at OECD (Business and Industry Advisory Committee, BIAC), the Trade Union Advisory Committee (TUAC), the Civil Society Information Society Advisory Council (CSISAC), and the Internet Technical Advisory Committee (ITAC). In addition, a targeted multistakeholder consultation on the draft Recommendation was undertaken in February 2021 to seek additional input from major stakeholders in the data ecosystem as well as from academics, whose participation in the JSG had been relatively limited.

[6] Among the legal instruments developed by the OECD to guide policy making in relation to the governance of data, four Recommendations are about enhancing access to and sharing of data (EASD) specifically, setting out guidance and best practices on common issues such as data openness, transparency, stakeholder engagement, intellectual property rights (IPR), and pricing. These Recommendations (referred to hereafter as "the current Recommendations") include: (i) the *OECD Recommendation concerning Access to Research Data from Public Funding* [OECD/LEGAL/0347] (hereafter, the Recommendation on Research Data); (ii) the OECD *Recommendation for Enhanced Access and More Effective Use of Public Sector Information* [OECD/LEGAL/0362] (hereafter, the PSI Recommendation); (iii) the *OECD Recommendation on Digital Government Strategies* [OECD/LEGAL/0406]; and (iv) the *OECD Recommendation on Health Data Governance* [OECD/LEGAL/0433].

[7] "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can

influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." (OECD, 2024[27])

[8] The order of Section 2.1 and 2.2 was swapped since the section on the "data openness" (new Section 2.2) required sufficient understanding about AI models as data.

[9] See also (OECD, 2024[27]).

[10] "Datafication" refers to "data generation through the digitisation of content, and monitoring of activities, including real-world (offline) activities and phenomena, through sensors". (OECD, 2015[4])

[11] Assuming data meets minimal requirements in terms of quality.

[12] Including AI algorithms and software.

[13] See also Bar-Sinai, Sweeney and Crosas (2016[93]), who propose an extensible, formal, theoretical model for dataset handling policies that reflect the risk associated with the data.

[14] Observed data are created where activities are captured and recorded. In contrast to volunteered data where the data subject is actively and purposefully sharing its data, the role of the data subject in case of observed data is rather passive and it is the data controller that plays the active role. Examples of observed data include location data of cellular mobile phones and data on web usage behaviour.

[15] Volunteered (or surrendered or contributed or provided) data are data provided by individuals when they explicitly share information about themselves or others. Examples include creating a social network profile and entering credit card information for online purchases.

[16] Derived (or inferred or imputed) data are created based on data analytics, including data "created in a fairly 'mechanical' fashion using simple reasoning and basic mathematics to detect patterns". In this case, it is (only) the data controller or processor that plays the active role in the creation of data. The data subject typically has little awareness over what is inferred about her or him, especially since that personal information can be derived from several pieces of seemingly anonymous or non-personal data. Examples of derived data include credit scores calculated based on an individual's financial history.

[17] Acquired (purchased or licensed) data are obtained from third parties based on commercial (licensing) contracts (e.g. when data are acquired from data brokers) or other non-commercial means (e.g. when data are acquired via open government initiatives). As a result, contractual and other legal obligations may affect the re-use and sharing of these data.

[18] As clarified by CC (2024[86]), "licensors agree to waive or not assert any moral rights, publicity rights, personality rights, or privacy rights they themselves hold, to the limited extent necessary to allow exercise of the licensed rights. Any rights outside of the scope of the license may require clearance (i.e., permission) in order to use the work as you would like. … For example, if you have licensed a photograph of yourself, you may not assert your right of privacy to have the photo removed from further distribution. … If there are any third parties who may have publicity, privacy, or personality rights that apply, those rights are not affected by your application of a CC license, and a reuser must seek permission for relevant uses. If you are aware of any such third party rights in the material you are licensing, we recommend marking the material to give notice to reusers."

[19] See (Perez Rivera, Emilsson and Ubaldi, 2020[95]) presenting the framework for comparing government open data initiatives via the OECD Open, Useful and Re-usable data (OURdata) Index.

[20] The project was set up for the period 2020–2022. For more information: https://vm.fi/en/opening-up-and-using-public-data.

[21] For a more detailed comparison, see (Liesenfeld and Dingemanse, 2024[48]; Bommasani et al., 2024[73]).

[22] 'Data users' was not defined [in part] as self-explanatory.

[23] The term "data controller" could have been used instead of "data holder". However, given that the term "data controller" was a term of art specific to privacy frameworks and referred to the controllers of personal data, a different term needed to be introduced.

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

[25] See also (European Commission, 2017[92])

[26] The number of datasets increased from 500 available before 2015 to over 108 000 available in November 2023.

[27] Further information about the DatA-IL Innovation Community can be found via https://data-il.org/.

[28] The first workshop focussed on manufacturing firms from pharmaceuticals, chemicals, and textiles manufacturing firms and the second focussed on firms from machines, motorized vehicles, and electronics manufacturing.

[29] For more information see: https://oecd-opsi.org/covid-response/release-of-open-api-dataset-for-the-availability-of-public-masks-at-designated-stores/

[30] The Council of Europe recommends that in some circumstances the transparency extend to include the logic underpinning the processing in the context of profiling. In EU GDPR, the principles of fair and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes, which include the right to be informed of the existence of profiling and the consequences of such profiling (Recital 60).

[31] "Transparency describes responsible disclosure to ensure people are aware that AI is being used in a prediction, recommendation or decision, or in an interaction (e.g. a chatbot)" (OECD, 2023[51]).

[32] Traceability describes "the need to maintain a complete account of the provenance of data, processes, code, and other elements in the development of an AI system." (OECD, 2023[51])

[33] Explainability "refers to the ability to accurately describe the mechanism, or implementation, that led to an algorithm's output". (OECD, 2023[51])

[34] Interpretability "refers to whether a human can derive meaning from a system's output for a specific use case". (OECD, 2023[51]).

[35] BigCode/Hugging Face/ServiceNow is the only developer that scores points on indicators relating to data creators, data copyright status, data license status, and personal information in data. Only 3 developers (Aleph Alpha, BigCode/Hugging Face/ServiceNow, IBM) score points on 6 or more of the 10 data indicators, while 6 developers score 2 or fewer points. (Bommasani et al., 2024[73])

[36] Another example is the *European Declaration on Digital Rights and Principles* committing to "ensuring that everyone has effective control of their personal and non-personal data in line with EU data protection rules and relevant EU law; [and] ensuring effectively the possibility for individuals to easily move their personal and nonpersonal data between different digital services in line with portability rights" (European Union, 2022[100]).

[37] The 1996 WIPO "Internet" Treaties (WCT and WPPT) contain provisions to prevent the unauthorized removal or alteration of electronic 'rights management information' (RMI) such as information about the author, rights holder, or terms or conditions of copyright protected works.

[38] These are health, mobility, industrial and manufacturing, agriculture, finance, green deal, energy, public administration, skills, open science cloud, tourism, cultural heritage, media, language.

[39] One approach for supporting a whole-of-government approach, for example, may involve creating and making accessible data repositories containing high-value data sets of public interest, either via open data anrangements or conditioned data access and sharing arrangements, as appropriate. These data repositories can for instance serve as a central infrastructure for efficient information sharing between public entities, thereby improving services for residents and supporting the formulation of data-driven public policies.

40 The Commissioner will have opportunities to promote and engage in policy work on responsible data sharing (including data use for the broader public good); increase public understanding of the benefits and risks of the data-driven economy; and seek stakeholder input on required regulation for range of risks due to AI that does not stifle innovation.

[41] As noted in (OECD, 2022[105]), for competition authorities: "this can include considering multi-sidedness and business models that involve the provision of products at a price of zero, often in exchange for consumer data. Similarly, the role of data in reinforcing demand-side characteristics of a market, including in relation to search and switching costs, and choice and information overload, could further contribute to entrenching market power for a firm in a dominant position."

[42] Supporting this, a guide aligned with the Australian Code for the Responsible Conduct of Research (Australian Research Council, 2018[99]) delineates the responsibilities of both institutions and researchers in facilitating the sharing of research data.

[43] See also Benjamin et al. (2019[101]) for a possible taxonomy for the licensing of data in the fields of artificial intelligence and machine learning.

[44] For instance, results from the 2023 edition of the Open, Useful and Re-usable data show how open government data have maintained their relevance across OECD Members as a result of their role as a trustworthy input users can use to develop, train and apply data-intensive systems including AI systems. Open government data also helps in better managing risks related to the provenance of data and their sources.

[45] See also the May 2024 Foundation Model Transparency Index by Bommasani et al., 2024[65]).

[46] According to Recital 57e to the EU AI Act, "Developers of free and open source tools, services, processes, or AI components other than general-purpose AI models should be encouraged to implement widely adopted documentation practices, such as model cards and data sheets, as a way to accelerate information sharing along the AI value chain, allowing the promotion of trustworthy AI systems in the Union."

[47] But there are also commercial providers of input data such as Bright Data. But in many cases, as the authors highlight "it remains unclear whether or under which license the datasets were acquired".