



Governing data and artificial intelligence for all

Models for
sustainable and just
data governance

STUDY

Panel for the Future of Science and Technology

EPRS | European Parliamentary Research Service

Scientific Foresight Unit (STOA)

PE 729.533 – July 2022



EN

Governing data and artificial intelligence for all

Models for sustainable and just data governance

With a particular focus on artificial intelligence (AI), this study identifies and examines policy options for the EU's data governance framework that align with a data justice perspective. A data justice approach is one that centres on equity, recognition and representation of plural interests, and the creation and preservation of public goods as its principal goals. The analysis offers both an assessment of the EU data governance strategy overall and specific policy options for the AI act, the data governance act and the data act.

Four benchmarks for good data governance are proposed, in line with the principles of justice: preserving and strengthening public infrastructure and public goods, inclusiveness, contestability and accountability, and global responsibility. Exploring examples of different governance models, we examine how these models and options intersect, and what lessons they offer for the EU case.

AUTHORS

This study was written by Joan Lopez Solano, Aaron Martin, Siddharth de Souza and Linnet Taylor of the Global Data Justice project, Tilburg University, at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit, within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament.

The Global Data Justice project would like to acknowledge valuable contributions to the analysis in this report from: Maria Anagnostu, Shweta Degalahal, Paula Ferreira Vidal, Yash Kaushal, Andrew Key, Janne Joosten, Alexis Manus, Franklyn Ohai, Gargi Sharma and Zsuzsanna Véghné Ujj.

ADMINISTRATOR RESPONSIBLE

Philip Boucher, Scientific Foresight Unit (STOA)

To contact the publisher, please e-mail stoa@ep.europa.eu

LINGUISTIC VERSION

Original: EN

Manuscript completed in June 2022.

DISCLAIMER AND COPYRIGHT

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

Brussels © European Union, 2022.

PE 729.533
ISBN: 978-92-846-9623-9
doi: 10.2861/915401
QA-05-22-170-EN-N

<http://www.europarl.europa.eu/stoa> (STOA website)
<http://www.eprs.ep.parl.union.eu> (intranet)
<http://www.europarl.europa.eu/thinktank> (internet)
<http://epthinktank.eu> (blog)

Executive summary

With particular regard to artificial intelligence (AI), this study aims to identify and examine policy options for Europe's data governance framework that align with a data justice perspective. A data justice approach is one that centres on equity, the recognition and representation of plural interests, and the creation and preservation of public goods as its principal goals. The analysis offers both an assessment of the European data governance strategy overall and specific policy options for the AI act, the data governance act and the data act.

A data justice perspective is a particularly appropriate tool for this analysis, because AI is not a bottom-up class of technology, in terms of either development or use. Developing and deploying AI systems is, and will remain for the foreseeable future, something that is only generally possible for powerful and well-resourced actors in society, whether commercial or public-sector. Unless substantial resources are invested in public alternatives, these activities will also rely on large-scale commercial computing infrastructure that channels the power to analyse and intervene towards those with the most resources and capacity.

As such, the central question this report addresses is how to foster a positive vision of AI as contributing to public goods and creating public value. We find that this can only be achieved through governance approaches that purposefully distribute power over AI systems, and the data ecosystems they rely on, outward towards civil society and democratic institutions, and that strongly incentivise good behaviour on the part of those developing and deploying those systems.

Starting from research on data justice, we propose four benchmarks for good governance: preserving and strengthening public infrastructure and public goods, inclusiveness, contestability and accountability, and global responsibility. We look at the principal ways in which data is currently understood – as a tradeable asset, as a commons, as a strategic national asset, and as a component of individual identities – and demonstrate how these different conceptualisations interact in governance models from various regions around the world.

We provide examples of alternative governance models including indigenous and local forms of digital sovereignty, public data trusts, collaboratives, commons, and systems of personal data sovereignty. We analyse how these either support or undermine good data governance overall, as defined according to our four benchmarks, and explore how these dominant models and options intersect through forms of governance such as private regulation, a precautionary approach, domain-specific and sectoral models for governing, comprehensive approaches, and private and cooperative forms of regulation.

Our key findings:

1. Defining data's potential as a public good

The EU still has work to do in conceptualising what kind of public good data should be. While the legal framework under construction articulates an aim of creating value from data for both public and private purposes, the mechanisms for arbitrating between these often conflicting aims are unclear and the balancing of public and private interests varies across legislative instruments. In particular, the idea that actors will behave altruistically or beneficially without strong constraints – such as human rights impact assessments, enforceable guidelines on how to prioritise rights and interests with regard to data and AI systems, and meaningful ways for affected groups to make claims – power and potentially impunity will accrue to the most powerful and least accessible actors controlling these systems.

2. Constitutionalising the EU approach to data governance

We find that the existing regulatory framework in the EU for data governance runs the risk of becoming fragmented. While the focus on building digital markets is coherent, the different instruments involved create disjunctures in the way technological harms are conceptualised (i.e. through the lenses of data protection, competition and consumer protection) and in turn, this limits the equitable distribution of power both in terms of accessing and using data, and in making claims and seeking redress where necessary. This becomes especially visible in the more complex cases posed by platforms and by the deployment of AI systems, which may simultaneously violate rights, build monopolies and exert unfair trading practices, but which must be separately addressed according to these different logics. Approaching data governance through a constitutional lens – an overarching set of aims regarding rights and the equitable distribution of power – has the potential to provide a coherent path despite this complexity.

3. Centring collective will and decision-making

We argue that AI and data governance should centre collective will and decision-making on the part of societal groups, along with a systemic orientation towards public value. The EU's investments in public infrastructure (named in its data strategy, and implied in the data governance act and the data act) could be reoriented to reflect plural understandings of how data generates value, especially in terms of both large and smaller-scale computing and data infrastructure. Plural thinking and input on digital infrastructure, we argue, is urgently required to support and build public goods within the EU, and to render those public goods resilient to capture by big tech. We also find that the current thinking on AI governance leaves civil society too exposed to exploitation and rights violations, and that there exist multiple paths to civil society power over AI development and deployment that are not yet being explored. This may be due to the framing of AI systems or models as single products rather than as components of a field of research, a set of dynamically evolving systems that will be used in different ways to generate different kinds of value over time.

4. Contextualising tools of data governance

Current trends in data governance involve the development of different tools such as data trusts, various forms of cooperatives and commons, and stewardship processes. We find that none of these are relevant as stand-alone approaches to data governance, but become relevant in relation to particular goals. As such, all are open to misuse if overarching normative goals are not clearly articulated and enforced. For instance, the notion of data trusts and stewardship arrangements has been leveraged by big tech to gain control over data's potential public value in other regions. However, if applied with the goal of creating particular public goods and providing accountability, these tools hold the potential to establish controls for data flows that would in turn give sectoral and interest groups an understanding of how data is being used for AI applications in a given area.

5. Devolving and distributing oversight

Technology regulation enforcement and oversight are increasingly challenged to demonstrate that they can represent the democratic concerns of society. Democratising the process of oversight and enforcement with regard to data and AI could help address this challenge. As powerful technologies are increasingly used on the public in ways that are opaque to individuals, it has become urgently necessary to have oversight and enforcement structures that have a public-facing component, that can demonstrate democratic accountability and therefore that are also more representative of society. It is a challenge to expect legal professionals to recognise emergent harms in diverse societies without a direct connection to civil society, and it is also a challenge to civil society to identify and seek redress in a situation where they do not see themselves represented in oversight bodies.

Policy options:

For the EU to define **how data can become a public good**, AI and data governance could centre collective will and decision-making on the part of societal groups, along with a systemic orientation towards public value. EU investments in public infrastructure (named in its data strategy, and implied in the data governance act and the data act), could reflect plural understandings of how data generate value, not least in terms of large and smaller-scale computing and data infrastructures. Plural thinking and input on digital infrastructure could be mobilised to support and build public goods within the EU, and to render those public goods resilient to capture by big tech.

In relation to **constitutionalising the EU's legislative approach**, the current diversity of legislative instruments under development could be seen as an opportunity to constitutionalise digital spaces. A constitutional law perspective offers a way to limit the power of both public and private actors performing public functions, and to make them accountable to the people. It also provides a language of rights as well as the opportunity to challenge excesses of public and private power. Adopting a framing of digital constitutionalism makes sense of the current diversity of new legislation by testing each for whether it can balance power, provide tools to challenge it, and offer frameworks to hold those with power accountable.

The aim of **centring on civil society** as the most important beneficiary of the data economy could be achieved by providing for equitable infrastructure development and data access, creating and resourcing environmentally sustainable infrastructures and practices, centring on the FAIR principles (findable, accessible, interoperable, reusable) in data policy as a requirement for global responsibility, along with the CARE principles that provide guidance on collective benefit, authority to control, responsibility and ethics, in situations where power over data is asymmetric, and recognising that government is not the only actor who can use data in the public interest. Distributing and devolving the power to access, use and benefit from data is possible, as long as protection from harm is also addressed as a common good to be distributed equitably.

To **contextualise data governance tools**, authorities could pay more attention to defining goals in terms of particular public goods – such as equitable access to education or housing, political representation of marginalised groups, or an equitable social safety net – and apply data governance tools in relation to those goals. Where the goals are public, the tools will also often mean developing public infrastructure for AI and the data it requires. Enabling smaller ecosystems to survive within these infrastructures might be achieved by centring governance and power in the hands of the local developer or community rather than passing it to the owner of the computing infrastructure within which it lives.

In order to **democratise oversight**, we set out options for structuring governance to foster civil society agency in governing data, namely by distributing oversight responsibilities to sectoral and local organisations (such as associations, interest groups, municipalities and provinces). Such a distributed, domain-related oversight infrastructure would be complementary to the current centralised but overburdened approach, which will only become more overburdened once AI oversight is added. This kind of structure would increase the system's capacity to identify the incremental harms that are most common with AI and data analytic systems, where a critical mass of complaints builds over time from individuals affected enough to make a claim.

Finally, we develop a **sustainable approach** that can take account of the data economy's environmental impacts through its operations, and of the sustainability of particular practices within the data economy. The first would require broad conceptualisation of how a data economy can operate sustainably (rather than the false fix of stimulating commercial infrastructure's consumption of green energy at the expense of other industries), while the second would require adoption of public infrastructure and open practices with data and AI. Sustainable practices also involve defining the relationship of these technologies to fundamental rights; identifying where pseudoscience is

being normalised in ways that distort protection from harmful AI practices, and involving civil society on a plural, continuing and structural basis in setting the goals and determining the path of evolution of data and AI technologies.

Table of contents

1. Introduction	1
2. Our approach to analysing data governance	6
2.1. What is governance?	6
2.1.1. What is the place of regulation in governance?	6
2.1.2. How is data currently defined for purposes of data governance?	10
2.1.3. The fluid nature of data and AI	12
3. Review of dominant data governance models and challenges raised	18
3.1. Dominant conceptual bases for governing data	18
3.1.1. Data as a strategic national asset	18
3.1.2. Data as a proprietary asset	21
3.1.3. Personal versus non-personal data	25
3.2. Alternatives to dominant data governance models	25
3.2.1. Public data trusts	26
3.2.2. Data collaboratives	26
3.2.1. Data (semi) commons	27
3.2.2. Data cooperatives	28
3.2.3. Indigenous data sovereignty	29
3.2.4. Personal data sovereignty	31
3.2.5. How these models interact with the dominant narratives over data	31
3.3. Points of interaction between fundamental visions	34
3.4. How broader models are incorporated in current data governance	35
4. What are the key principles and features of 'good' data governance, and what is our basis for judging this in the EU?	38
4.1.1. Preserving and strengthening public infrastructures and public goods	41
4.1.2. Inclusiveness	42
4.1.3. Contestability and accountability	44

4.1.4. Global responsibility	45
5. Review of EU policy context	48
5.1. Existing data strategy and regulation	48
5.2. Relevant legislative files and topical debates	51
5.2.1. AI Act	51
5.2.2. Data Governance Act	54
5.2.3. Data Act	55
6. Policy options and alternatives: a data justice analysis	58
6.1. European Strategy for Data	58
6.2. AI Act	59
6.2.1. AI as a public technology	59
6.2.2. Accountability	59
6.2.3. Defining vulnerability	60
6.2.4. Contestability	61
6.2.5. Beyond the fundamental rights framing	61
6.3. Data Governance Act	61
6.3.1. Data altruism and digital public goods	61
6.3.2. Meaningful data collectivisation	62
6.3.3. Democratising data and civil society agency	62
6.3.4. Governing for a sustainable data economy	63
6.4. Data Act	63
6.4.1. Resisting the commodification of data	63
6.4.2. Enabling public interest use of data	64
6.4.3. Recognition of collective needs and rights around data	64
6.4.4. Interoperability for challenging the dominant data governance model	64
7. Discussion and conclusions	65

List of figures

Figure 1 - Representing the data value chain _____	11
Figure 2 - Features of data as a proprietary model _____	24
Figure 3 - Interactions between different data governance models _____	33

List of tables

Table 1 - An overview of EU legislative frameworks related to data/digital topics _____	2
Table 2 - An overview of data as a strategic national asset in China, Russia and India _____	21
Table 3 - Features and characteristics of alternative data governance models _____	31

1. Introduction

The governance of data is the governance of different interests. As a term, 'governance' has become useful for denoting the whole landscape of actors involved in rulemaking and regulating, including but not limited to governments. As such, governing data involves multiple actors whose task is to arbitrate between competing needs, which in turn relate to much more than data or technology itself. Data represents different social and economic interests, so that for any governance arrangement or model to be credible and to gain traction at scale, it has to have a claim to represent a plurality of needs and perspectives. An important component of that representation can be framed as 'data justice' - the view that data governance should not only seek to do no harm, but should positively contribute to people's autonomy and to their ability to participate in society and make claims about their needs, on a more general level.¹ In contrast, our current worldwide model for data governance represents a very specific set of interests: those of the largest players in the technology sphere, and the states in whose economies they are embedded. In this vision, data is primarily conceptualised as an asset, and citizens as data suppliers. Under this model, data is appropriated in various forms - often without meaningful consent - and used to generate immense value for the most powerful actors internationally, namely those with the capacity to gather and manage it. This means there are imbalances in access to all forms of data, in control over its use, and in the distribution of the costs, benefits and risks that come with participation in the global digital economy.

The European Union (EU) is making strides to align its approach to data governance with European values, but this is happening in a piecemeal process and risks getting lost to an overall market framing. Each legislative act that will be discussed in this report makes a different value statement but as yet the EU is missing an overarching articulation of values with respect to data governance. As a result the existing vision contained in the EU's legislative instruments is normatively limited, consisting predominantly of managerial aspects of governance.² What we propose is a model that frames governance in terms of certain constitutional parameters, where digital constitutionalism articulates how to limit the exercise of power in a digital economy.³ In section 5 of this report, we return to the notion of digital constitutionalism as an overarching framework that provides guidance and values to address how to centre the protection of rights and freedoms of people in their interactions with institutions and actors in the digital economy.

Artificial intelligence (AI) makes a value articulation necessary because it represents a change in the scale and nature of the data governance challenge – it takes what was often an issue of individual rights and control and adds to it problems of collective rights and claims-making, and preserving public values in relation to what are predominantly privately developed and deployed technologies. The EU legislative files we will principally focus on in this context are:

- The **Data Governance Act**, which aims to set up robust mechanisms to facilitate the reuse of certain categories of protected public-sector data, increase trust in data intermediation services and foster 'data altruism' across the EU.

¹ Linnet Taylor, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally,' *Big Data & Society* 4, no. 2 (December 1, 2017): 2053951717736335, <https://doi.org/10.1177/2053951717736335>.

² For example, the Data Governance Act will 'increase trust in data sharing, strengthen mechanisms to increase data availability and overcome technical obstacles to the reuse of data.' European Commission, 'Data Governance Act | Shaping Europe's Digital Future,' November 25, 2020, <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>.

³ Nicolas Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms,' *Social Media + Society* 4, no. 3 (July 1, 2018): 2056305118787812, <https://doi.org/10.1177/2056305118787812>.

- The **Data Act**, which aims at facilitating access to and use of non-personal data, including business-to-business, business-to-government and business-to-customer data, and to review the rules on the legal protection of databases. Meant to complement the Data Governance Act, it seeks to balance rights to data access and incentives for data investment, without altering data protection rules.
- The **AI Act**, which proposes horizontal rules for the development, commodification and use of AI-driven products, services and systems, would introduce a 'product safety framework' for AI based on four risk categories.

A group of other relevant legislative files are part of the broader context of EU lawmaking on data and AI but do not form the main focus of this report. These notably include, but are not limited to:⁴

- The **Digital Services Act**, which will build on the e-Commerce Directive by modernising and harmonising the EU's legal framework for handling illegal or potentially harmful content online, the liability of online intermediaries for third-party content, the protection of users' fundamental rights online and bridging information asymmetries between online intermediaries and users.
- The **Digital Markets Act**, which seeks to address imbalances in digital markets in the EU arising from the dominance of large online platforms (so-called gatekeeper platforms) by setting out harmonised rules that define and prohibit certain unfair practices by gatekeepers and providing an enforcement mechanism based on market investigations.

Table 1 - An overview of EU legislative frameworks related to data/digital topics

Name of file	Status of proposal	Regulatory scope
Data Governance Act	Final approval of text by European Parliament and the Council underway following conclusion of trilogue negotiations	Protected public-sector data, data sharing services and data altruism
Data Act	First text published; awaiting committee opinion	Non-personal data in business-to-business, business-to-government and business-to-customer environments
AI Act	Under discussion in the European Parliament and the Council	Artificial intelligence systems
Digital Services Act	Currently in trilogue discussions	[Digital] intermediary services
Digital Markets Act	Currently in trilogue discussions	Gatekeepers of [digital] platforms

Source: authors' own work

Our main questions in this report are as follows: what are the characteristics of good data and AI governance, and how can the EU embed them in its current and future legislative processes? We will analyse dominant models of data governance and their alternatives, and offer an assessment of their technological, economic, socio-ethical and legal implications and potential for the EU, given its current policy landscape.

To answer these questions, we must first understand what we mean by data governance, as the different definitions in play across different fields have implications for how one understands both the ethical and practical questions of what makes governance good. Data governance is often seen

⁴ For a complete overview of relevant legislative files, see Cristiano Codagnone, Giovanni Liva & Teresa Rodriguez de las Herras Ballel (2022) 'Identification and assessment of existing and draft EU legislation in the digital field'. European Parliament special committee on Artificial Intelligence in a Digital Age (AIDA). Accessible at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703345/IPOL_STU\(2022\)703345_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703345/IPOL_STU(2022)703345_EN.pdf)

as an internal process that companies and bodies handling privately held data go through in order to be compliant with law and standards.⁵ There is another definition of data governance, however, which aims to describe and analyse the whole realm of actors involved in governing data, from private to public and from law to practical arrangements to deal with particular uses and problems. In this study we aim for the second conceptualisation.

In order to provide a basis for reasoning about what the EU's approach should be, we will also outline the different approaches to data governance seen in different regional environments, highlighting the problems and responses that recur around the world, and which constitute the larger landscape in which the EU is deciding on its data governance approach. We will also provide criteria, based on current scholarship, for 'good data governance'. We will then review the EU policy context, relating these broader observations to the particular questions our region is attempting to answer. Finally, we will centre in on specific legislative files relating to data and AI, where discussion is currently underway and intensifying, analysing these directions in data governance in light of our observations.

Most analyses of data governance start from certain assumptions. These include the notion that data is an asset⁶ which can - or should - be traded as a commodity both within and between countries. Another is that the current economic model of the internet and the app economy, where digital platforms and services operate by monetising personal data through adtech,⁷ is the only viable source of sustainability for our digital economy. Yet another is that we understand when data is personal or non-personal, and, connected to this, that effective de-identification is possible so that data can be legally commodified and traded. Relatedly, the supervision and enforcement model embodied by the General Data Protection Regulation (GDPR) – the EU's legal framework for personal data protection – is too often cited as addressing broader data governance challenges, namely ensuring the efficient functioning of digital markets such as those for advertising and for digital services offered to consumers.⁸ In this report, we will show how these assumptions reflect particular interests and views of data, and demonstrate that different starting assumptions are possible. If we change our assumptions, we acquire different possible ways of shaping our digital economy and the technological systems, such as AI, that depend on it. In doing so, we aim to provide ways of understanding and choosing between different possible governance models rather than assuming the status quo of data as an asset as the only possible starting point.

When we consider why we should govern technology, the answer often relates to preventing harm as well as maximising its beneficial potential to the economy and society at large. Governing data technologies such as AI therefore also requires conceptualising how technology affects people (and other interests including our shared environment and environmental resources, our economies and our cultural lives). In turn, this requires conceptualising which people are affected and in what ways, and how the impacts on people are different in different contexts. This leads to complex considerations of harm and vulnerability. In technology governance it is common to make a division between vulnerable and less vulnerable people and groups, based for instance on age, health or ability status, membership of a social group or income level. This is not a given, however: it is

⁵ Tim Stobierski, 'Data Governance: A Primer for Managers | HBS Online,' Business Insights Blog, February 16, 2021, <https://online.hbs.edu/blog/post/data-governance>.

⁶ World Economic Forum and Bain & Company, Inc, 'Personal Data: The Emergence of a New Asset Class' (World Economic Forum, January 2011), https://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

⁷ Notably pushback on this occurred in 2021 from the EDPB (group of DPAs) in Marc Hijink, 'Toezichthouders: verbied online volgen en persoonlijke advertenties,' NRC, November 19, 2021, <https://www.nrc.nl/nieuws/2021/11/19/toezichthouders-verbied-online-volgen-en-persoonlijke-advertenties-a4066161>.

⁸ Vagelis Papakonstantinou and Paul De Hert, 'EU Lawmaking in the Artificial Intelligent Age: Act-Ification, GDPR Mimesis, and Regulatory Brutality,' *European Law Blog* (blog), July 8, 2021, <https://europeanlawblog.eu/2021/07/08/eu-lawmaking-in-the-artificial-intelligent-age-act-ification-gdpr-mimesis-and-regulatory-brutality/>.

important to take account of structural inequality, where people and groups have been historically subject to discrimination, harm or unfair treatment, and to the ways in which these vulnerabilities become translated into the present by applications of technology.⁹ However, it is also possible to conceptualise data technologies and AI as creating new vulnerabilities which may be experienced by anyone, and may not relate to historical and structural harms. In this report we will deal with both, and will attempt to distinguish where data governance treats certain groups or people as inherently vulnerable, versus cases (often harder to pin down in terms of pre-emptive governance) where the technology already does, or is likely to, create new forms of vulnerability.

One of the central questions for technology regulation is always whether a technology requires its own set of regulatory rules, or whether it can be covered by existing frameworks - i.e. whether the law needs to change, or our interpretation and enforcement of the law.¹⁰ Do we need to consider the impact of data and AI on particular communities or sectors of interest in order to make sense of the practical challenges of governing? Or is it enough to centre on an overall approach which will be interpreted and concretised sector by sector or application by application? There are at least three benefits to such a general approach (which characterises most of the EU legislative instruments and files we will discuss in this report). First, it stimulates public and legal debate, and second, through such debates also provides explicit warning signs and red lines indicating when a law is not fit for purpose. Third, it is not constrained by the sectoral influence of market actors. In contrast, the effect of regulating technology at the sectoral level, for example by making a data governance framework for the gig economy or for financial data,¹¹ may be to create an incentive for technology firms in particular to redefine themselves in order to escape regulation, for instance by claiming that an entire workforce is not employed but freelancing,¹² or that holiday rentals should not be classified as hotels.¹³

All these concerns also matter for governing AI, which in this report we define in a broad way to indicate a range of computational techniques including machine learning and deep learning, logic- and knowledge-based approaches and statistical methods, used on data stemming from the public and private sectors as well as the activities and behaviour of individuals.¹⁴ The current phase of AI development and application highlights many of the problems with data which remain despite advances in data protection regulation. The use of identification techniques in relation to

⁹ In this report we do not assume that vulnerability is inherent. We start from the position articulated by EDRI: 'Taking a leaf out of the book of anti-racism movements – instead of being 'vulnerable' to discrimination, exploitation and other harms, we know they are imposed on us. Rather than vulnerable, some groups are marginalised, as active processes with people, institutions and structures of power as the cause.' European Digital Rights (EDRI), 'Digital Rights for All,' European Digital Rights (EDRI), August 7, 2020, <https://edri.org/our-work/digital-rights-for-all/>.

¹⁰ See, for instance, the case for and against treating technology as exceptional: Frank H Easterbrook, 'Cyberspace and the Law of the Horse,' *THE UNIVERSITY OF CHICAGO LEGAL FORUM* 207 (1996): 11.

¹¹ Such as, for example: Monetary Authority of Singapore, 'MAS-Led Industry Consortium Publishes Assessment Methodologies for Responsible Use of AI by Financial Institutions,' Monetary Authority of Singapore, April 2, 2022, <https://www.mas.gov.sg/news/media-releases/2022/mas-led-industry-consortium-publishes-assessment-methodologies-for-responsible-use-of-ai-by-financial-institutions>.

¹² As, for instance, Uber has repeatedly claimed and been challenged on: Owen Bowcott, 'Uber to Face Stricter EU Regulation after ECJ Rules It Is Transport Firm,' *The Guardian*, December 20, 2017, sec. Technology, <https://www.theguardian.com/technology/2017/dec/20/uber-european-court-of-justice-ruling-barcelona-taxi-drivers-ecj-eu>.

¹³ As has Airbnb on: Will Feuer, 'Europe's Top Court Just Delivered Airbnb a Major Victory as the Company Prepares to Go Public,' *CNBC*, December 19, 2019, sec. Technology, <https://www.cnbc.com/2019/12/19/airbnb-wins-legal-victory-from-europes-top-court-as-it-looks-to-ipo.html>

¹⁴ See, e.g., definitions in European Parliament and European Council, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts,' Pub. L. No. COM(2021) 206 final, 2021/0106 (2021), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

intervention on the population level; problems of collective privacy and autonomy,¹⁵ and issues of power asymmetry between the users and subjects of AI technologies all require further thinking beyond what is contained in existing data protection legislation. AI affects core concepts such as purpose limitation, data minimisation and the definition of personal data. Much of the regulatory and ethics discussion about AI in the EU has also centred the concept of fairness, which has been demonstrated to be particularly contentious because it means different things to technical and non-technical communities.¹⁶

These are just a few of the areas of data governance where the stakes for those using and managing data are high: regulation impacts the work of companies and markets, and the terminology and framing of principles and criteria matters because it determines how instruments of governance interpret and operationalise issues of rights and agency. With this work, we aim to contribute to greater clarity on both fronts.

¹⁵ Linnet Taylor, Luciano Floridi, and Bart van der Sloot, eds., *Group Privacy: New Challenges of Data Technologies* (Cham: Springer International Publishing, 2017), <https://doi.org/10.1007/978-3-319-46608-8>.

¹⁶ Andrew Selbst, danah boyd, Sorelle Friedler, Suresh Venkatasubramanian, & Janet Vertesi (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68). <https://dl.acm.org/doi/pdf/10.1145/3287560.3287598>

2. Our approach to analysing data governance

2.1. What is governance?

The concept of governance takes on different forms: it provides an institutional architecture from which different systems are organised, it offers a framework and regulation for participation by different actors, it allows for negotiations between competing authorities by providing policies and procedures and provides ways in which conflicts can be mediated.¹⁷

As a concept, it has many meanings and attributes attached to it, including in terms of its association to outcomes such as accountability or transparency, to processes in terms of collaboration or participation, to thematic areas whether corporate, environmental or data, and much of it is focused on building a narrative of progress in terms of policy and reform.¹⁸ Governance also takes place in a multi-jurisdictional manner so that it can apply at the local, regional, national and transnational levels, and its relations can manage power within domains such as markets and networks, as well as across the private and public sector divide.¹⁹

In this report, we are interested in the *power* that different actors have in shaping and making governance mechanisms. This includes, at the regulatory level, the state and public institutions that legislative instruments to govern data; it also includes the private power that corporations exercise in creating internal mechanisms such as oversight boards to govern data, or sectoral initiatives such as codes of conduct for responsible data handling. Furthermore, it also covers the approaches that civil society organisations and communities offer in terms of alternative visions for data governance, which place an emphasis on how communities can own, collect and determine the application of their own data.²⁰ By focusing on power, we are also able to examine the unevenness that emerges as different governance frameworks interact with one another. An account of governance models therefore needs to explore the ways in which they relate to each other, how they may complement or conflict, and what implications this has within a data economy.

We are also interested in governance in terms of *practice*, i.e. the ways in which institutions, relationships and policy instruments come to shape technology. This happens through practices of global or local governance, partnerships between different stakeholders such as international organisations and corporations, but also between different sectors and across borders between states.

Finally, we are also interested in governance in terms of *impacts*, and whether as a system of rules, policies and procedures, it enables values such as legitimacy, participation and inclusion. In the context of data with implications for people, this includes not only how data is governed, but also the governance effects produced by data.²¹

2.1.1. What is the place of regulation in governance?

These conceptual points on governance can manifest themselves practically through different kinds of regulatory arrangements, which are relevant to governing data and the use of AI, and may also

¹⁷ Gunnar F. Schuppert, 'Governance,' in *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, ed. James D. Wright (Oxford: Elsevier, 2015), 292–300, <https://doi.org/10.1016/B978-0-08-097086-8.75020-3>.

¹⁸ Franco Moretti and Dominique Pestre, 'Bankspeak,' *New Left Review*, no. 92 (April 1, 2015): 75–99.

¹⁹ Mark Bevir, 'Governance as Theory, Practice, and Dilemma,' in *The SAGE Handbook of Governance* (London: SAGE Publications Ltd, 2011), 1–16, <https://doi.org/10.4135/9781446200964>.

²⁰ John Taylor and Tahu Kukutai, *Indigenous Data Sovereignty* (ANU Press), <https://doi.org/10.22459/CAEPR38.11.2016>.

²¹ Fleur Johns, 'Governance by Data,' *Annual Review of Law and Social Science* 17, no. 1 (October 13, 2021): 53–71, <https://doi.org/10.1146/annurev-lawsocsci-120920-085138>.

serve as inspiration for future iterations of governance models in the EU. For illustrative purposes and to ground the discussion, here we provide examples of different forms of established and emerging data regulation at state, private and global levels. While each is presented separately, it is important to stress that in practice these levels interact with one another extensively.

State-based regulation, often referred to as command-and-control or top-down regulation, typically comes in the form of legislation, i.e. 'legal rules backed by criminal sanctions'.²² In this regard, the EU's most recognisable regulatory activity around data has been in the form of secondary legislation, namely the General Data Protection Regulation (GDPR), which came into force in 2018 and aims at the 'protection of natural persons with regard to the processing of personal data and on the free movement of such data'. Less discussed, however, is a separate EU instrument for regulating the free flow of non-personal data. In contrast to GDPR, the non-personal data regulation (NDPR) does not follow a fundamental rights rationale and instead is meant to stimulate the free movement of non-personal data within the EU. It also puts restrictions on data localisation in EU Member State legislation (unless justified on public security grounds and true to the proportionality principle). As commentators have observed, 'the NPDR's lasting contribution to European data law may lie in establishing non-personal data as a distinct category in law, ostensibly regulated without regard to data protection concerns'.²³ As will be analysed in depth later in this report, emerging proposals in the EU, i.e. the Data Governance Act and the Data Act, are moving the legislative needle in the area of non-personal data governance.

There are also various forms of private (i.e. non-state) regulation that govern the collection, protection and use of data and deployment of AI in different ways. By definition, these governance arrangements decentre the role of the state as a regulatory authority,²⁴ and are often referred to as 'self-regulation', which in the EU context describes 'the possibility for economic operators, social partners, non-governmental organisations or associations to adopt amongst themselves and for themselves common guidelines at European level (particularly codes of practice or sectoral agreements)'.²⁵ In many cases, however, private regulation can be enabled by legislation, meaning the state is not entirely absent from these regulatory practices. This phenomenon is referred to as co-regulation²⁶ or 'regulated self-regulation',²⁷

Codes of conduct, for example, may be created and enforced by industry associations and other bodies to help promote good data governance practices. In the context of personal data protection, Article 40 of the GDPR even 'encourage[s] the drawing up of codes of conduct intended to contribute to the proper application' of the regulation, with data protection authorities in Member States performing an oversight role for their approval and use. Similarly, voluntary standards have

²² Julia Black, 'Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a 'Post-Regulatory' World,' *Current Legal Problems* 54, no. 1 (February 21, 2001): 103–46.

²³ Thomas Streinz, 'The Evolution of European Data Law,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, January 18, 2021), <https://doi.org/10.2139/ssrn.3762971>.

²⁴ Julia Black, 'Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a 'Post-Regulatory' World,' *Current Legal Problems* 54, no. 1 (February 21, 2001): 103–46.

²⁵ European Parliament, European Commission, and European Council, 'Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-Making,' 123 OJ L (2016), http://data.europa.eu/eli/agree_interinst/2016/512/oj/eng, para 22.

²⁶ Christopher T. Marsden, *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge: Cambridge University Press, 2011), <https://doi.org/10.1017/CBO9780511763410>.

²⁷ Robert P. Kaye, 'Regulated (Self-)Regulation: A New Paradigm for Controlling the Professions?,' *Public Policy and Administration* 21, no. 3 (September 1, 2006): 105–19, <https://doi.org/10.1177/095207670602100308>.

been developed to supplement the practice and enforcement of technical and organisational measures required by GDPR.²⁸

In the domain of AI, and largely in the absence of governing legislation, AI ethics committees are an increasingly common feature of corporate data governance, which are proposed as a way to institutionalise ethical decision making to help eradicate bias from — and to promote fairness in — AI systems.²⁹ New non-corporate initiatives, such as the Distributed Artificial Intelligence Research (DAIR) Institute, which seeks to be a 'space for independent, community-rooted AI research free from Big Tech's pervasive influence', are also developing industry-wide standards for bias mitigation in AI datasets 'by making it common practice for researchers to write accompanying documentation about how they gathered their data, what its limitations are and how it should (or should not) be used'.³⁰ Inspired by innovations in cybersecurity governance, we have also seen the emergence of so-called bug bounty programs to promote the discovery and mitigation of algorithmic harms.³¹

In the domain of content moderation, social media platforms have also created so-called 'oversight boards' — an attempt at strengthening 'independent judgement' regarding decisions to remove harmful or illegal content on platforms. Facebook's quasi-judiciary oversight board, for example, became operational in 2020 and at the time of writing consists of 20 members representing 18 countries.³²

To date, regulatory interest at the global level has largely focused on governing trade-related aspects of data and AI. The World Trade Organisation (WTO), for example, is examining the relationship between digital trade and the governance of various types of data, as well as how established trade principles might hinder or facilitate interoperability among data governance approaches. The WTO is concerned with whether trading in data differs from trading in goods or services governed by traditional trade agreements and how open trade principles, as well as domestic regulation (or the absence thereof), might make it easier or more difficult for nations to develop a trustworthy environment for data-driven economic development, including AI innovation.

Other international actors like UNESCO – the UN specialised agency that promotes world peace and security through international cooperation in education, arts, sciences and culture – have pursued ethical questions related to AI. In November 2021, the Recommendation on the Ethics of Artificial Intelligence was adopted by UNESCO's General Conference (importantly, all 193 members of UNESCO signed on). The first global standard-setting instrument on the ethics of AI, it advances a number of principles including fairness and non-discrimination, sustainability, privacy, safety and security, transparency and explainability, and awareness and literacy. While non-binding, the recommendation can be viewed as a form of soft law and could serve as the basis for legislative

²⁸ Irene Kamara, 'Data Protection Standardisation: The Role and Limits of Technical Standards in the EU Data Protection Law' (Doctoral Thesis, Tilburg University, 2021), <https://research.tilburguniversity.edu/en/publications/data-protection-standardisation-the-role-and-limits-of-technical->

²⁹ Emanuel Moss and Jacob Metcalf, 'Ethics Owners: A New Model of Organizational Responsibility in Data-Driven Technology Companies' (Data and Society, September 2020), https://datasociety.net/wp-content/uploads/2020/09/Ethics-Owners_20200923-DataSociety.pdf.

³⁰ Billy Perrigo, 'Why Timnit Gebru Isn't Waiting for Big Tech to Fix AI's Problems,' *Time*, January 18, 2022, <https://time.com/6132399/timnit-gebru-ai-google/>.

³¹ Josh Kenway et al., 'BUG BOUNTIES FOR ALGORITHMIC HARMS? Lessons from Cybersecurity Vulnerability Disclosure for Algorithmic Harms Discovery, Disclosure, and Redress' (Algorithmic Justice League, January 2022), <https://www.ajl.org/bugs>.

³² Note that in two cases, board members represent two different countries <https://www.oversightboard.com/meet-the-board/>

instruments down the line. It is also worth pointing out that the UNESCO principles are but one of very many ethical and human rights-based frameworks for AI to be promulgated in recent years.³³

Across the state, private and global levels, data governance models can have different purposes from providing an institutional architecture which ascertains different functions and hierarchies for how data is managed across different models; a policy and regulatory infrastructure which determines how data is gathered, managed, stored, used, shared and protected in different situations; a technical aspect which provides guidelines and procedures, for instance in regard to data security or debiasing AI training data; and a civic aspect which relates to questions of data literacy, partnerships and the participation of people in governance.³⁴

We believe it is difficult to separate these different functions and address them in isolation. Doing so would mean not acknowledging the ways in which governance can materialise in different contexts. For instance, while there may be a robust regulatory infrastructure for data, without adequate institutional independence and stability, it will be difficult to enforce and implement regulatory policies, and without a flourishing civil society, questions of accountability and monitoring become more fragmented and ad hoc.³⁵ Therefore data governance needs to be able to examine the ways in which data creates relations between people (horizontal data relations) and between people and institutions or platforms (vertical data relations).³⁶

In thinking about data relations, and how these emerge in different contexts, it is also useful to examine the connections between the purposes for which different models have been set up and what values they espouse. For instance, the FAIR principles, which emerged in open science, provide a basis upon which data can be shared by making data more findable, accessible, interoperable and reusable.³⁷

However, the idea of greater sharing has also been critiqued for not accounting for historic and contextual ways in which data has marginalised different groups of people and also reduced scope for agency and control over data, for instance by indigenous people. In response to this, CARE principles have emerged which include ideas of collective benefit, authority to control, responsibility and ethics.³⁸ These contrasts highlight a fundamental tension that data governance

³³ See a mapping of these approaches by Berkman Klein Center here: Jessica Fjeld and Adam Nagy, 'Principled Artificial Intelligence Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI,' Berkman Klein Center, February 5, 2020, <https://cyber.harvard.edu/publication/2020/principled-ai>.

³⁴ United Nations Department of Economic and Social Affairs, 'UN/DESA Policy Brief #89: Strengthening Data Governance for Effective Use of Open Data and Big Data Analytics for Combating COVID-19 | Department of Economic and Social Affairs,' United Nations Department of Economic and Social Affairs, December 21, 2020, <https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-89-strengthening-data-governance-for-effective-use-of-open-data-and-big-data-analytics-for-combating-covid-19/>.

³⁵ Chinmayi Arun, 'AI and the Global South: Designing for Other Worlds,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, June 9, 2019), <https://papers.ssrn.com/abstract=3403010>.

³⁶ According to Salomé Viljoen, vertical data relations, which describe the relationship between people and data collectors (often digital platforms), are encoded technically via data flows and legally by contracts, privacy rules and consumer protection law. Horizontal data relations involve information extracted from people that puts them in a relationship with others who are connected to them in some way (voluntarily or not), and who may be acted upon (via behavioural manipulation, for example) on the basis of this collected information. See: Salome Viljoen, 'Data Governance for a Society of Equals,' LPE Project, March 22, 2021, <https://lpeproject.org/blog/data-governance-for-a-society-of-equals/>.

³⁷ The first step in (re)using data is being able to find them (i.e. making them findable), including by both humans and computers. Making data accessible involves consideration of authentication and authorisation measures for data access. Interoperability concerns integration with other data as well as with applications/workflows for data analysis, storage and processing. The reusability of data, which is the ultimate goal of the FAIR principles, involves describing data and metadata sufficiently well that they can be replicated and/or combined in different settings. FAIR Principles,' GO FAIR, accessed April 26, 2022, <https://www.go-fair.org/fair-principles/>.

³⁸ According to the principles, collective benefit requires that data ecosystems are designed to function in ways that enable indigenous peoples to derive benefit from data. Authority to control entails empowering indigenous peoples' rights and

frameworks need to address, one where there is an emphasis on value maximisation and extracting the most economic benefit from data, and the second, a more grounded approach that focuses on context while emphasising the role of accountability, representation and rights for different members.³⁹

2.1.2. How is data currently defined for purposes of data governance?

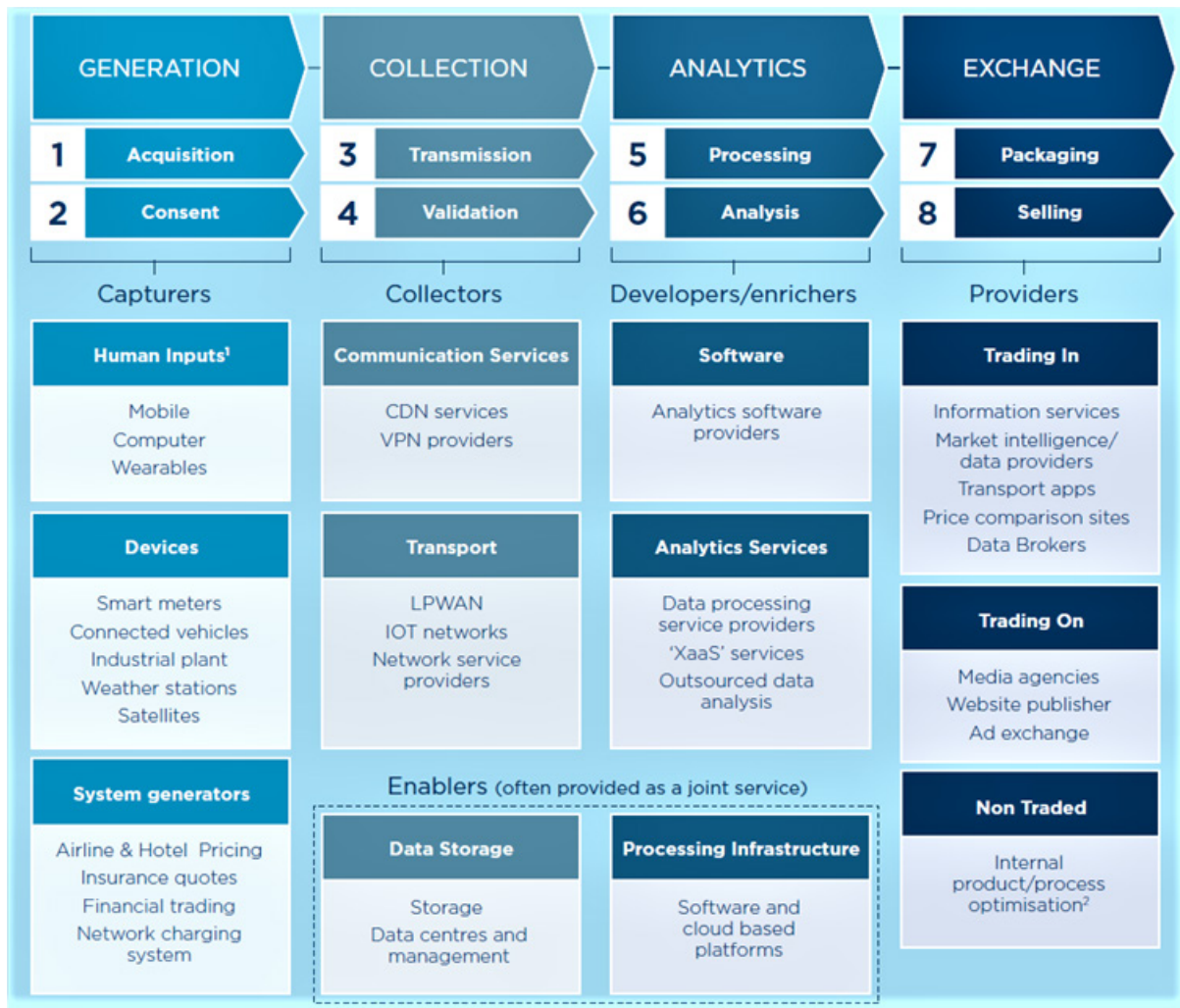
Definitions of data in the sphere of governance and legislation tend to rely on the assumption that the focus should be on personal data when it comes to operationalising rights and preventing harms, and non-personal data when it comes to opening data up for broader uses - for example, when patient data becomes translated into public health data, or shared for purposes of academic research. This translation process requires consideration of how what may begin as personal data becomes part of models which then have collective impacts, and potentially harms. In this report we will outline the pitfalls of that approach and why, if we want good data governance – or even adequate data governance that is fit for purpose in the 2020s and beyond – we should think about how our different categories of 'data' are in fact overlapping rather than mutually exclusive.

Figure 1, showing the GSMA's overview of data value chains - a term denoting the models and approaches that companies take in order to monetise data - serves to explain the challenges of strict categorisation: For each of the sources or practices covered there (certainly not an exhaustive list) we can infer the types of data involved, and can evaluate the value and risks to different parties given different uses. However, almost all of these types of data can, depending on the circumstances, be defined as personal or non-personal and therefore fall into radically different categories in terms of governance.

interests in their data as well as their authority to control such data, including how data about indigenous lands, territories, resources, knowledge and geographical indicators are represented and identified. Responsibility suggests that people working with indigenous data should share how the data is used to support indigenous peoples' self-determination and collective benefit, while accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to indigenous communities. Regarding ethics, indigenous peoples' rights and well-being should be the primary concern at all stages of the data lifecycle and across the data ecosystem. Global Indigenous Data Alliance, 'CARE Principles of Indigenous Data Governance,' Global Indigenous Data Alliance, accessed April 26, 2022, <https://www.gida-global.org/care>.

³⁹ Sean Martin McDonald, 'Data Governance's New Clothes,' Centre for International Governance Innovation, July 5, 2021, <https://www.cigionline.org/articles/data-governances-new-clothes/>.

Figure 1 - Representing the data value chain



Source: GSMA and A.T. Kearney, 'The Data Value Chain' (GSMA, June 2018), https://www.gsma.com/publicpolicy/wp-content/uploads/2018/06/GSMA_Data_Value_Chain_June_2018.pdf.

This blurring of boundaries has increasingly caused problems for data governance because it confounds the categorisation on which our architectures and conceptualisations currently rest: the personal/non-personal divide, and with it, the different streams of thinking and methods of governing it.

One main reason why we wish to govern data is to determine how it enters and circulates via the various digital markets that have come into existence over the last decades. These markets include online advertising markets of the sort managed by Google and Facebook, where a data collection behemoth brokers profiles of users for firms who want to advertise to them, and rents those users' attention to the highest bidder, but also more informal ones where the interests of different actors, both commercial and public-sector, may come together around a particular type of data or datafied process. One example of this is the data produced by connected objects, which map, track and follow behaviour and the environment over time. In such a market, tradeable data assets may or may not reflect individual identities: individual profiles are valuable to advertisers but so are insights about types of people who may share preferences and behaviours. Another example of data gaining value for different actors along its lifecycle is mobile phone location data - originally customer billing records, but which have acquired value for law enforcement, for academic sociologists, for urban

planners and for public health bodies, to name but a few. Such location data is also valuable on a secondary level for advertisers who wish to understand which online profiles belong to humans (who generate location records) rather than bots (which do not) or to illegitimately 'redline' certain groups to exclude them from particular services or goods.

In most of the world, data in its existence as capital is mainly traded by a few huge monopolies acting as exchanges, where behavioural futures deriving from human raw materials can be traded. In this market,⁴⁰ data quickly becomes separated from the humans who originate it, and who have little knowledge of where it resides, or ability to claim rights over it. Ironically data's value, particularly as an input to AI systems, depends to a great extent on human work cleaning, tagging and labelling data points that can then be translated into financial value.⁴¹ This suggests that data can be both capital and labour at once.⁴²

Data's value in the market is currently conceptualised mainly around the ways in which it can be used to target either individuals or groups for intervention, both commercial and non-commercial. Profiles have value in terms of making it possible to target individuals, for example in personalised advertising, but also because they can be aggregated into types so that they can be targeted as groups, as happens for instance when local authorities buy aggregated and de-identified⁴³ location data from brokers to manage crowds in public space.⁴⁴ This 'non-distributive' profiling practice makes it possible to categorise people probabilistically and to target those most likely to have characteristics of interest, such as the propensity to buy a particular product or service, to pose a security risk, to migrate, or any other characteristic of choice. These profiles constitute data derivatives, and may come in the shape of models, inputs for other processes, or combinations of profiles. This complexity has implications for our current regulatory framework for data: if the most valuable commodity is not personal data but data derivatives which may not relate to individuals directly, but which nevertheless have the kind of effects that privacy rights and data protection were originally designed to address, then data governance needs to extend far beyond data protection.

2.1.3. The fluid nature of data and AI

Although legal and regulatory perspectives tend to treat 'personal data' as the most important defining category for digital data, this is not the only story about how data is connected to people

⁴⁰ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2016), <https://www.hup.harvard.edu/catalog.php?isbn=9780674970847>; Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: Ingram Publisher Services, 2019); Linnet Taylor et al., '(Re)Making Data Markets: An Exploration of the Regulatory Challenges,' *Law, Innovation and Technology*, 2022, <https://doi.org/10.31235/osf.io/pv98s>.

⁴¹ Andrea Fumagalli et al., 'Digital Labour in the Platform Economy: The Case of Facebook,' *Sustainability* 10, no. 6 (June 2018): 1757, <https://doi.org/10.3390/su10061757>.

⁴² Imanol Arrieta Ibarra et al., 'Should We Treat Data as Labor? Moving Beyond "Free," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, December 27, 2017), <https://papers.ssrn.com/abstract=3093683>.

⁴³ The term 'deidentified' is purposely used here to denote the scientific consensus that for most forms of data, despite increasingly sophisticated techniques for removing identifiability, full anonymisation is not practically possible due to advances in computing over the 2000s. Location data, in particular, is almost impossible to anonymise (Yves-Alexandre De Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 'Unique in the crowd: The privacy bounds of human mobility.' *Scientific reports* 3, no. 1 (2013): 1-5.

<https://www.nature.com/articles/srep01376>); For a detailed review and explanation, see: Daniel Nunan, and Maria Laura Di Domenico. 'Exploring reidentification risk: is anonymisation a promise we can keep?.' *International Journal of Market Research* 58, no. 1 (2016): 19-34. <https://doi.org/10.2501/IJMR-2016-004>.

⁴⁴ Linnet Taylor, Luciano Floridi, and Bart van der Sloot, eds., *Group Privacy: New Challenges of Data Technologies* (Cham: Springer International Publishing, 2017), <https://doi.org/10.1007/978-3-319-46608-8>.

and publics: in fact data can transition between different definitions depending on who uses it, for what purpose, and whether or not it is de-identified.

i. All data is potentially security data⁴⁵

In the absence of robust governance frameworks that can account for, and appropriately regulate around, data's fluidity, we are witnessing perverse outcomes following from the explosion of digital data in different contexts. As an example, telematics data which includes information from a vehicle's GPS system and onboard diagnostics regarding things like speed, location, airbag and seatbelt status, engine temperature, maintenance requirements and service needs was originally intended to be used to improve the performance and efficiency of connected vehicles. It is not generally considered to be sensitive data as it is meant to serve technical objectives not directly related to human subjects. And while this data is of growing interest to the automobile insurance sector to adjust premiums,⁴⁶ law enforcement and immigration authorities are also increasingly requesting it from private firms for purposes entirely unrelated to vehicle management.

In the US, for example, Customs and Border Protection (CBP) and Immigrations Customs Enforcement (ICE) officials have demanded telematics data from the companies that collectively track the movements of tens of millions of vehicles on a daily basis.⁴⁷ They do so for the purposes of determining both the historic and live locations of individuals of interest. While in this context it appears CBP and ICE access to such data is regulated by a court order or warrant, which would set out the reasoning for data requests and thus provides some degree of protection from the misuse of data, military contractors in the US are also marketing the ability to pinpoint 'the real-time locations of specific cars in nearly any country on Earth' using the same kinds of telematics data.⁴⁸

Geotab, a Canadian firm and leading provider of telematics data, says its customers 'own their data' and thus 'have rights over it'⁴⁹, but it is increasingly unclear what such notions of data ownership and rights mean in securitised contexts such as these. Another provider, GM OnStar, explains in its privacy policy that it shares user data across its subsidiaries and with business partners, research institutes, car dealers, rental companies and other third parties for marketing activities 'with necessary consents'. Where GM OnStar anonymises telematics data, it may share it with third parties for 'any legitimate business purpose.' The company's spokesperson says users have to agree to its privacy statement before using OnStar services.

In another case, the legal research and risk 'solutions' firm, LexisNexis, which has historically offered its data brokerage services to financial institutions and insurance companies, is now selling billions of records to ICE along with analytical tools to help the immigration agency conduct investigations.⁵⁰ Whereas civil society has often struggled to articulate the harms and risks posed by

⁴⁵ cf. Rob Aitken, 'All Data Is Credit Data: Constituting the Unbanked,' *Competition & Change* 21, no. 4 (August 1, 2017): 274–300, <https://doi.org/10.1177/1024529417712830>.

⁴⁶ Verbelen Roel, Katrien Antonio, and Gerda Claeskens, 'Unraveling the Predictive Power of Telematics Data in Car Insurance Pricing,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 7, 2017), <https://doi.org/10.2139/ssrn.2872112>.

⁴⁷ Thomas Brewster, 'These Companies Track Millions Of Cars – Immigration And Border Police Have Been Grabbing Their Data,' *Forbes*, January 4, 2021, sec. Cybersecurity, <https://www.forbes.com/sites/thomasbrewster/2021/04/01/these-companies-track-millions-of-cars-immigration-and-border-police-have-been-grabbing-their-data/>.

⁴⁸ Joseph Cox, 'Cars Have Your Location. This Spy Firm Wants to Sell It to the U.S. Military,' *Vice*, March 17, 2021, <https://www.vice.com/en/article/k7adn9/car-location-data-telematics-us-military-ulysses-group>.

⁴⁹ Brewster, 'These Companies Track Millions Of Cars – Immigration And Border Police Have Been Grabbing Their Data.'

⁵⁰ Sam Biddle, 'LexisNexis to Provide Giant Database of Personal Information to ICE,' *The Intercept*, February 4, 2021, <https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis/>.

the virtually unregulated data brokerage industry,⁵¹ the use of these services in the immigration context, particularly to facilitate deportations, is clearly problematic and a threat to fundamental rights. These risks emerge closer to home as well: in the EU, in early 2022 Europol was instructed by the European Data Protection Supervisor (EDPS) to delete vast stores of illegally collected personal data.⁵² These are just a few illustrative examples of breakdowns in the concept and practice of data protection in the security context.

ii. All data is potentially health data

Health data is one of the categories that data governance is most keen to pin down due to its assumed sensitive nature, and a history of abuse over the last century by unethical actors. At first, health data was principally defined as patient data collected in relation to treatment or to clinical research, and was thus assumed to circulate between patients, their doctors, and related clinical or research institutions such as hospitals, universities or public health bodies. As such it was defined as sensitive and particular conditions were applied to its creation, storage and (re)use. With the advent of e-health and the digitisation of healthcare records in many countries, however, a new dimension was added: that of insights derived from cross-tabulating categories in databases, including insurance and billing data. Since the 2000s, a new dimension has been added: data collected as a by-product of activities which may or may not be related to healthcare provision or public health. For instance, signs of early-stage Parkinson's disease can be detected using key-stroke analysis from subject's typing patterns⁵³ - an analytic tool developed using consenting patients aware of being part of a study, but which could also be used to analyse the health of anyone typing on a connected device, without their knowledge. In another example, mobile phone location records, used as billing data, have been used to analyse people's risk of being subjected to domestic violence⁵⁴ and the same type of data has also been shown to provide early identification of people's infection with flu.⁵⁵

The point at which these data sources produce 'personal data' depends on the decisions of the researcher as to whether to de-identify them or not – and even when de-identified, data can remain sensitive. The recent case of Crisis Text Line⁵⁶, a US nonprofit which offers counselling for people with mental health issues and suicidal thoughts, demonstrates this fluidity. Data was de-identified and aggregated before being shared with a related for-profit company which used it as training data for customer service software. This caused a very public backlash from people who felt that the data,

⁵¹Aaron Rieke et al., 'Data Brokers IN AN OPEN SOCIETY' (Open Society Foundations, November 2016), <https://www.opensocietyfoundations.org/uploads/42d529c7-a351-412e-a065-53770cf1d35e/data-brokers-in-an-open-society-20161121.pdf>; Sharon Bradford Franklin, Greg Nojeim, and Dhanaraj Thakur, 'Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers' (Center for Democracy and Technology, September 12, 2021), <https://cdt.org/insights/report-legal-loopholes-and-data-for-dollars-how-law-enforcement-and-intelligence-agencies-are-buying-your-data-from-brokers/>.

⁵² Apostolis Fotiadis et al., 'A Data 'Black Hole': Europol Ordered to Delete Vast Store of Personal Data,' *The Guardian*, January 10, 2022, sec. World news, <https://www.theguardian.com/world/2022/jan/10/a-data-black-hole-europol-ordered-to-delete-vast-store-of-personal-data>; Teresa Quintel, 'European Union · The EDPS on Europol's Big Data Challenge in Light of the Recast Europol Regulation,' *European Data Protection Law Review* 8, no. 1 (2022): 90–102, <https://doi.org/10.21552/edpl/2022/1/14>.

⁵³ Goldberger, A., L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. 'PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.' (2000).

⁵⁴ Ting Chang et al., 'The Role of Alcohol Outlet Visits Derived from Mobile Phone Location Data in Enhancing Domestic Violence Prediction at the Neighborhood Level,' *Health & Place* 73 (January 1, 2022): 102736, <https://doi.org/10.1016/j.healthplace.2021.102736>.

⁵⁵ Emory Health Sciences. 'Anonymous cell phone data can quantify behavioral changes for flu-like illnesses: Iceland study links cell phone metadata with public health data.' ScienceDaily. www.sciencedaily.com/releases/2021/01/210126140056.htm (accessed April 26, 2022).

⁵⁶ <https://www.protocol.com/bulletins/crisis-line-ends-data-sharing>

albeit de-identified, was not appropriate as training material for a for-profit service - bearing out Nissenbaum's 'contextual integrity' theory⁵⁷ which holds that people will see unrelated third-party uses of data as illegitimate.

These cases demonstrate just a few of the myriad ways in which apparently innocuous metadata from people's everyday devices becomes potentially highly sensitive health data as soon as it is subjected to particular research practices. This is just one area where it makes sense to think of data and data technologies, including AI, as creating vulnerability, rather than exploiting inherent vulnerability in different groups. Regardless of whether people are in a vulnerable situation, the exploitation of data that offers insights about their current and future health conditions creates vulnerability of various kinds. It also illustrates how data moves between categories depending on who has access to it, what their purpose is, and what analytical methods are used. There is no way to cover this kind of fluidity without declaring all social data off-limits as sensitive.⁵⁸

iii. Digitising crime data

Increasingly, judicial systems around the world are finding ways to digitise their practices as well as their repositories for data relating to crime in order to build more efficient judicial processes. One such proposal from India is the creation of an Interoperable Criminal Justice Database. This database has been designed to include data from the police along with data from prisons and courts. The stated objective of this initiative is that for the criminal justice system to work more effectively, data needs to be shared across different institutions in the system, and be made accessible and interoperable.⁵⁹ The Supreme Court of India, in its vision document for a digitised judiciary, has called for the 'seamless exchange of live data' rather than data being shared on a need-to-know basis. The purpose being that this will improve the execution of judicial functions which in the court's view currently operate in silos.⁶⁰

The prospect of having such a centralised database has raised questions of judicial independence where court information will now be housed with the government and thereby create implications for separation of powers as well as fundamental rights implications for people, including questions of privacy and liberty.⁶¹ Additionally, biases at the stage of data creation at one institution will now circulate between different institutions without sufficient interrogation of the methods used to create the data registries.⁶²

This example highlights how, when data flows between institutions that have different functions and different modes of independence and accountability to the public, the capacity to interrogate the data changes. Further, if there is a seamless exchange between data which is not on a need-to-

⁵⁷ Helen Nissenbaum, 'Contextual Integrity Up and Down the Data Food Chain,' *Theoretical Inquiries in Law* 20, no. 1 (January 1, 2019): 221–56, <https://doi.org/10.1515/til-2019-0008>.

⁵⁸ Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law,' *Law, Innovation and Technology* 10, no. 1 (January 2, 2018): 40–81, <https://doi.org/10.1080/17579961.2018.1452176>.

⁵⁹ High Court of Tripura, 'Interoperable Criminal Justice System,' accessed April 26, 2022, <https://thc.nic.in/user%20manual/ICJS-manual.pdf>.

⁶⁰ E-COMMITTEE SUPREME COURT OF INDIA, 'Digital Courts Vision & Roadmap Phase III of the ECourts Project,' accessed April 26, 2022, https://www.livelaw.in/pdf_upload/digital-courts-e-courts-project-supreme-court-e-committee-391425.pdf.

⁶¹ E-COMMITTEE SUPREME COURT OF INDIA, 'Digital Courts Vision & Roadmap Phase III of the ECourts Project,' accessed April 26, 2022, https://www.livelaw.in/pdf_upload/digital-courts-e-courts-project-supreme-court-e-committee-391425.pdf.

⁶² It was found for instance that the creation of registers of repeat offenders had an underlying caste bias, and rather than challenging the existence of such registers, they would now be part of a centralised data base where their use would be further cemented. The Indian Express, 'The Dangers of a Centralised Database for Justice System,' May 28, 2021, <https://indianexpress.com/article/opinion/columns/the-dangers-of-a-centralised-database-for-justice-system-7333252/>.

know basis, it exposes risks of marginalising people for whom data collected in one context is used against them in another, thereby reducing their capacity and autonomy to take action about the ways in which the data is used in relation to them.⁶³

iv. Data's potential for collective harms

Another illustration of how data may travel between categories can be seen in the problem of group data. Nearly ten years ago, researchers identified the problem that data without links to identifiable individuals can nevertheless pose a risk to them, and data that applies to a given individual can also be made to offer insights about whole groups, in ways that are not predictable to the people involved. A good example of the first type of problem is satellite data showing human settlements in South Sudan during periods of genocidal attacks by pro-Sudanese militias.⁶⁴ Although people were not visible in these images, once hacked by the aggressors in this conflict the images identified where people were living, which settlements had not yet been attacked, and provided a guide for hostile militias to find surviving communities and wipe them out - making these images without humans in them the most sensitive personal data possible. In a more innocuous example, algorithmic analysis by Cambridge Analytica⁶⁵ of data about 300,000 Facebook users' online behaviour and preferences made it possible to create a dataset of 'lookalikes' which then allowed for the attempted political manipulation of 50 million US voters. With lookalike audiences, the opportunity for exploitation stems from their beliefs, values and personal experiences, which are categorised into psychographic information and used to target groups based on types, rather than individuals based on unique attributes.

These two opposite examples, of data showing no identifiable individual attributes versus data showing detailed attributes which then are used to create collective types and exploit people psychologically, both demonstrate that governing data by categorising it at its starting point as sensitive, personal or in a special category is no guarantee that we will capture the actual effects of data on people during its lifecycle. However it is initially categorised, data can have surprising consequences for people not initially part of its creation, and poses a problem for governance models that aim to control and channel data based on its initial relation to individuals, its ability to identify people, or its sensitivity. Whole business models in the tech sector are based on subverting these categorisations, as occurred with Cambridge Analytica, which was able to sell its ability to target people for political manipulation, based mainly on unrelated people's data, in 68 countries worldwide.⁶⁶

v. Digitisation overlaps with AI

In many of the issues that have come up with AI, decisions made about how to gather and use data are the source of risk. For instance in the Dutch government's misuse of data to predict fraud among

⁶³ Linnet Taylor, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally,' *Big Data & Society* 4, no. 2 (December 1, 2017): 2053951717736335, <https://doi.org/10.1177/2053951717736335>.

⁶⁴ Linnet Taylor, 'Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, October 6, 2016), <https://papers.ssrn.com/abstract=2848825>.

⁶⁵ Nadeem Badshah, 'Facebook to Contact 87 Million Users Affected by Data Breach,' *The Guardian*, April 8, 2018, sec. Technology, <https://www.theguardian.com/technology/2018/apr/08/facebook-to-contact-the-87-million-users-affected-by-data-breach>.

⁶⁶ Carole Cadwalladr, 'Fresh Cambridge Analytica Leak Shows Global Manipulation Is out of Control,' *The Observer*, January 4, 2020, sec. UK news, <https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation>.

social welfare recipients (the SyRI case,⁶⁷ and the toeslagenaffaire, or child benefit case⁶⁸), the cases have been reported as problems caused by automation and AI. In fact, these were black-boxed systems where automation does not appear to have been the defining factor - instead the main problem was the use of illegitimate forms of analysis (based on racialised and ethnic attributes in the child benefit case, and data maximisation by Dutch public institutions with SyRI). This took place in combination with algorithmic approaches which grouped people by particular sensitive attributes, using those attributes to proxy for fraud risk.

These approaches were designed and applied by humans, albeit using digital records and algorithmic logics, and there was no question of a lack of human oversight. Both cases were examples of bad public policy that happened to be processed using digital records. Despite this, however, the information uncovered by ensuing legal challenges has demonstrated that the practices used by public ministries in these cases is indistinguishable from social scoring - in the SyRI case, myriad data points were brought together from different public-sector sources into a single system to calculate fraud risk, with direct effects on individuals due to that; and in the child benefit case, families were flagged for fraud based on attributes such as dual nationality. These cases demonstrate the overlap between the activities that require 'data governance' and suggest that statements about what kind of action we wish to prevent, and why, become more essential as data technologies become more ubiquitous and complex.

The examples in this section are not intended to argue that data is too complex to govern effectively, or that it will always escape from measures taken to control it. Instead, they demonstrate that data has a lifecycle during which it may have diverse meanings, purposes and effects, and that it can be used simultaneously in multiple ways at different levels, from local to international. Effective data governance therefore has a temporal dimension: checks and controls are needed at points where data's use or users change, which may occur throughout the open-ended lifecycle of data, across its different sectoral uses, and in line with the different aims of the actors using it. It also requires a spatial dimension, in that as data flows between sectors, institutions, or markets, there is a recalibration of the terms on which it is used in the new space. This is critical to ensure that it does not subvert or destabilise the existing conditions or rules that govern that space.

⁶⁷ Jenny Gesley, 'Netherlands: Court Prohibits Government's Use of AI Software to Detect Welfare Fraud | Library of Congress,' Library of Congress, 2020, <https://www.loc.gov/item/global-legal-monitor/2020-03-13/netherlands-court-prohibits-governments-use-of-ai-software-to-detect-welfare-fraud/>.

⁶⁸ Ingo Dachwitz, 'Childcare benefits scandal: Dutch government to pay million Euro fine over racist data discrimination,' netzpolitik.org, January 6, 2022, <https://netzpolitik.org/2022/childcare-benefits-scandal-dutch-government-to-pay-million-euro-fine-over-racist-data-discrimination/>.

3. Review of dominant data governance models and challenges raised

3.1. Dominant conceptual bases for governing data

We tend to approach data governance with the idea that the current model - which aims for a worldwide free market for data, with some restrictions in a number of countries based on the notion of personal data - is the only possible normative starting point. This section will look at a range of alternative starting points for how data can be conceptualised and consider the different framings and configurations that these make available for data governance.

There are three main overarching framings - what one might term ideal types - at play in data governance which overlap with each other in almost every discussion: making data proprietary to those who produce and manage it; categorising data according to whether it relates to individuals or not; and the alternative visions. Each of these approaches has its own politics: libertarian, neoliberal, communitarian and totalitarian values can be discerned in the structuring of different governance models worldwide. We seldom scrutinise these underlying assumptions, but it is worth doing so because most data governance is an exercise in balancing between these political and configurational pivot points, so that choices about how far to go in one direction usually impact on our ability to make use of the others. As will be explored below, we also tend to make assumptions about each approach that are not necessarily true and which merit scrutiny if we are to think clearly about options and choices.

3.1.1. Data as a strategic national asset

For states with an established digital economy, the question of who has access to data and for what purposes has political as well as economic importance. Data is a key resource for population control and influence, for surveillance and monitoring both domestically and abroad, and for formulating economic and political strategy. In short, data is central to control. This manifests in different ways according to political systems and citizens' control over states, but the last decades have shown that states jealously guard their control over all kinds of data, and are eager to extend their reach into the digital spheres of other nations. The Snowden revelations demonstrated that even between allies, data is an important resource for both domestic and international surveillance and monitoring, and digital power has become a central topic of global diplomacy, with an ongoing search for new cyber norms to regulate power over the internet and forms of digital aggression between states. This search for a balance of interests is also continually surfacing on the national level, where regulation focusing on rights and democratic power over data is in a constant struggle with states' interests in maximising their ability to access and control data about people.⁶⁹

This balance can be seen in all states, but reaches an extreme in those where strong state control over the population and the economy is a central strategy for the government's survival. In China, for example, this manifests in a state-led data economy where population control and state security are both supported through data regulation. Since the 2000s, when China began constructing a data governance framework, the main objective has been to ensure that the nation becomes a 'digital superpower'. Its governance model rests on three pillars: 'informatisation', or the process of constructing policies that leverage data to generate economic growth and effective governance;

⁶⁹ Julia Pohle and Thorsten Thiel, 'Digital Sovereignty,' *Internet Policy Review* 9, no. 4 (December 17, 2020), <https://policyreview.info/concepts/digital-sovereignty>.

the introduction of digital technologies in social, economic, and political life; and cybersecurity, understood as protection from vulnerabilities that hostile actors could exploit to affect the country.

China's framework is based around two main regulations: the Personal Information Protection Law and the Data Security Law. The Personal Information Protection Law (PIPL) could be framed under the objective of informatisation: its main idea is to regulate and balance the relationship between individuals and entities collecting data, but with a focus on preserving state power over data. The PIPL recreates some of the consumer protection aspects of the GDPR but does not include privacy as a fundamental right. Likewise, the management of data by the government remains outside of that regulation, thus excluding the applications that could help to protect Chinese citizens against corruption.

The Data Security Law aims to offer a framework to protect China's national digital assets from a perspective of national security, constructing a framework to evaluate and manage potential risks to national security and the stability of the Chinese government. This framework assesses all kinds of data including personal and non-personal data in the light of national security and public interests, and essentially puts forward a risk-based approach where the state's interests are central.^{70,71}

Two further dimensions of the Chinese governance approach are also worth noting. First, the country's emerging governance framework for AI involves increasing competition between different parts of the state, which are promoting regulations for online algorithms (i.e. recommendation systems), testing and certification for AI systems, and even ethical guidelines.⁷² Second, experts have also observed a phenomenon they refer to as the 'Beijing Effect' whereby China is shaping data governance at the transnational level by supplying digital infrastructure to emerging markets, which means that China's data governance values and technological systems are increasingly relevant to other economies and societies.⁷³

Russia offers another example of a data governance model that prioritises state and security interests, while also providing a framework for regulating the way businesses process data about people. Russia's model frames data as a way to exert power domestically and internationally, with an overall policy articulation that prioritises first state priorities, then corporate ones, and only afterwards citizens' interests.⁷⁴ The model focuses particularly on the control of internet infrastructures as a way of ensuring that speech can be effectively controlled, with regulators cooperating with owners of internet infrastructure in order for the state to coordinate control over

⁷⁰ Data is categorised based on 'the level of importance to the State's economic and social development, as well as the degree of damages to the national security, social interests or the lawful interests of citizens and organizations, if the data is tampered, damaged, leaked or illegally obtained or used.' Covington, 'China Released Updated Draft Data Security Law and Personal Information Protection Law for Public Comments,' March 5, 2021, <https://www.cov.com/-/media/files/corporate/publications/2021/05/covington-alert--china-released-updated-draft-data-security-law-and-personal-information-protection-law-for-public-comments-may-3-2021.pdf>.

⁷¹ Rogier Creemers, 'China's Emerging Data Protection Framework,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 16, 2021), <https://doi.org/10.2139/ssrn.3964684>.

⁷² Matt Sheehan, 'China's New AI Governance Initiatives Shouldn't Be Ignored,' Carnegie Endowment for International Peace, April 1, 2022, <https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127>.

⁷³ Matthew S Erie and Thomas Streinz, 'The Beijing effect: China's digital silk road as transnational data governance' 54, no. 1 (2021): 92

⁷⁴ Stanislav Budnitsky and Lianrui Jia, 'Branding Internet Sovereignty: Digital Media and the Chinese–Russian Cyberalliance,' *European Journal of Cultural Studies* 21, no. 5 (October 1, 2018): 594–613, <https://doi.org/10.1177/1367549417751151>.

online speech.⁷⁵ As with China, data localisation is a key pillar of the strategy: all processing of data that relate to Russian citizens has to take place on servers in Russian territory, and any actors doing such processing are subject to reporting requirements to the Russian government. This both stimulates the growth of domestic technology businesses and locates control over information with the state. Unlike China's data protection law which does name the public interest in relation to people rather than only the state, Russia's data protection law focuses principally on security⁷⁶ and does not include a category of 'public interest' as a basis for processing data. Both countries' frameworks also lack the Data Protection Impact Assessment, which distributes the power to define risk and effect away from the state and toward businesses in particular, but (as in China) does include a security impact assessment.

In India, as of December 2021, the Joint Committee on Personal Data Protection submitted its report on the country's draft data protection legislation to Parliament.⁷⁷ The report in contrast to the previous versions recommended the regulation of both personal and non-personal data and recommended a regulatory authority that would be responsible for both. The Committee focused in its report on the economic value of data, and also advanced that it was an asset that was of 'national importance'. The report is also clear that there must be data localisation, advancing that data is valuable for matters of national security, innovation, geopolitical and economic purposes. It recommended a comprehensive policy that would ensure that there was the necessary infrastructural apparatus to achieve data localisation and that this did not compromise the ease of doing business.

In a further development aiming to limit the power of tech companies, many of which are global corporations, the committee recommended that there should be a streamlined process for data breaches, where companies were mandated to report said breaches within 72 hours. The report also acknowledged that existing regulation was not sufficient to regulate social media companies and recommended that they be treated as publishers in certain contexts. It recommended that all foreign social media companies have Indian offices, and, failing to do so, would limit their capacity to work in the Indian market. Social media companies were also recommended to be classified as significant data fiduciaries which would have to meet a higher compliance standard.⁷⁸

⁷⁵ Liudmila Sivetc, 'State Regulation of Online Speech in Russia: The Role of Internet Infrastructure Owners,' *International Journal of Law and Information Technology* 27, no. 1 (March 1, 2019): 28–49, <https://doi.org/10.1093/ijlit/eay016>.

⁷⁶ OneTrust DataGuidance, 'Russia - Data Protection Overview,' DataGuidance, April 12, 2022, <https://www.dataguidance.com/notes/russia-data-protection-overview>.

⁷⁷ Lok Sabha Secretariat, 'Report of the Joint Committee on the Personal Data Protection Bill,' December 2021, http://164.100.47.193/Isscommittee/Joint%20Committee%20on%20the%20Personal%20Data%20Protection%20Bill,%202019/17_Joint_Committee_on_the_Personal_Data_Protection_Bill_2019_1.pdf.

⁷⁸ Ikigai Law, 'Summary of the JPC report on data protection,' *Ikigai Law* (blog), December 17, 2021, <https://www.ikigailaw.com/summary-of-the-jpc-report-on-data-protection/>.

Table 2 - An overview of data as a strategic national asset in China, Russia and India

Country	Logics
China	<ul style="list-style-type: none"> ➤ State-led data economy with population control and state security as key foci ➤ Policy to create a digital superpower which has three pillars <ul style="list-style-type: none"> ✓ Economic growth and effective governance ✓ Digitalisation of social, economic and political life ✓ Cybersecurity, especially for threats from outside ➤ Data localisation is central to the model
Russia	<ul style="list-style-type: none"> ➤ Prioritises state and security interests. Does not include a category of public interest ➤ Ensures control of digital infrastructure, especially for instance to control online speech ➤ Aims to stimulate domestic technology businesses
India	<ul style="list-style-type: none"> ➤ Focuses on the economic value of data ➤ Emphasises data localisation, in particular for matters of national security ➤ Aims to limit the power of global tech companies, in particular concerning data breaches

3.1.2. Data as a proprietary asset

Despite the state interests outlined above, both national and international data markets predominantly treat data as an asset that behaves like private property,⁷⁹ and an asset that should be governed in relation to individual rights and claims. This is underpinned by the fundamental assumption that the individual is the relevant unit of analysis when we think about rights and claims relating to data, rather than, for example, groups or society as a whole. This model is based on the notion that, if individuals are properly compensated with rights over data and (sometimes) financial or other utility, data can constitute both a kind of capital and a commodity that can be freely traded. This in turn is the basis for digital markets around the world, and is a key feature of data governance propositions on the global level (for example by the G20 and WTO) due to its importance to economic growth. Data as an essential input to AI models, as a source of inferences and insights, including consumer and risk profiles, and of business process inputs, represents a set of informational infrastructures on which both business and policy increasingly rely.⁸⁰ Economists predominantly address data as non-rival and non-excludable in its nature, meaning that different actors can use a dataset at the same time without decreasing its utility.⁸¹

Wherever data is not governed otherwise by regulation, a free market tends to come into being where data is extracted and traded, principally to the benefit of those with the ability to gather and process it. The US, for example, has an almost completely free market with regard to most forms of data, which interacts with a federal structure where states can make their own decisions about how to regulate data (as is demonstrated by the flurry of state-level privacy laws in recent years⁸²).

⁷⁹ World Economic Forum and Bain & Company, Inc, 'Personal Data: The Emergence of a New Asset Class.'

⁸⁰ For research on infrastructures and their implications for public and private sector benefits from data and AI, see the work of Seda Gürses (TUDelft) and Martha Poon (Data & Society). e.g. Tobias Fiebig, Seda Gürses et al. (2021). 'Heads in the Clouds: Measuring the Implications of Universities Migrating to Public Clouds.' *arXiv preprint arXiv:2104.09462*.

⁸¹ Nestor Duch-Brown, Bertin Martens, and Frank Mueller-Langer, 'The Economics of Ownership, Access and Trade in Digital Data,' *JRC Digital Economy Working Paper 2017-01*, 2017, <https://doi.org/10.2139/ssrn.2914144>.

⁸² Taylor Kay Lively, 'US State Privacy Legislation Tracker,' IAPP, April 21, 2021, <https://iapp.org/resources/article/us-state-privacy-legislation-tracker/>.

Canada has a similar federal approach, where rules may differ radically on the subnational level. The logic of this kind of system - when formulated by policy - tends to be to prioritise the growth of the technology sector and to establish market dominance in terms of trading and brokering data internationally. The US's lack of regulation, for example, has been the chief force driving the development of what are estimated to be trillion-dollar international markets for advertising, profiling and consumer data which now affect other nations' ability to control and shape their data markets.

The 'data as asset' model has also given rise to a vision of data governance based on self-regulation by technology firms and other actors handling data at scale. Often citing 'data ethics' or 'responsible AI', corporations have made the case that they can be left to self-govern their use of data, sometimes adding in the promise of oversight bodies to assess content moderation decisions⁸³ or an AI development strategy.⁸⁴ These claims are tied to a model of what might be termed strong data capitalism, which centres corporate actors as key engines and decision-makers in economic development and growth, and the state as facilitator of this growth through legal and policy tools.

Within this model, the international data market also offers opportunities for states to serve as hubs for powerful market actors' data transactions. One example is Singapore, whose data governance framework focuses on maximising data's value to the private sector, including international firms. It therefore prioritises legal certainty and the streamlining of processes for firms trading data within and through the nation-state's servers and companies, but does not prioritise the right of individuals to make claims (something reflected in the act's history of enforcement to date),⁸⁵ or provide any controls on the state's own use of data. In order to provide greater legal certainty to data controllers, Singapore's data protection regulation defines a series of precise categories which proxy for degree of sensitivity, including personal data, 'derived personal data', 'publicly available data', 'user activity data', and 'user-provided data'.

This asset-focused approach to data governance is supported by policy on the global level, with the G20, WTO⁸⁶ and other international organisations such as UNCTAD aligning with the assumption that free international trade in data is a core priority for global data governance,⁸⁷ and other

⁸³Kate Klonick, 'Inside the Making of Facebook's Supreme Court,' *The New Yorker*, February 12, 2021, <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court>.

⁸⁴ Kelsey Piper, 'Exclusive: Google Cancels AI Ethics Board in Response to Outcry,' *Vox*, April 4, 2019, <https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board>.

⁸⁵ The Personal Data Protection Act was established in 2012. On 19 February 2019, the State Court of Singapore dismissed a claim brought against the Singapore Swimming Club for defamation and breach of the PDPA. Although written grounds of judgement are not available, this case is significant as it appears to be the first time where the Singapore courts were asked to consider whether there was a breach of the PDPA, even though the PDPC had not made any decision in respect of any purported contravention of the PDPA.

⁸⁶ Chris Foster, 'A First Look at the WTO/JSI Discussions on Digital Trade,' *Digital Trade Tracker* (blog), March 16, 2021, <https://digitaltradetracker.org/2021/03/16/a-first-look-at-the-wto-jsi-discussions-on-digital-trade/>.

⁸⁷ According to UNCTAD, 'any international framework for governing cross-border data flows needs to complement and be coherent with national policies for making the data-driven digital economy work for development. It will need to be flexible, so that countries with different levels of readiness and capacities to benefit from data have the necessary policy space when designing and implementing their development strategies in the data-driven digital economy.' UNCTAD's position on data ownership versus access and control is as follows: 'The particular characteristics of data suggest that they need to be treated differently from conventional goods and services, including in their international transfers. In the new context of the data-driven digital economy, concepts such as ownership and sovereignty are being challenged. Rather than trying to determine who 'owns' the data, what matters is who has the right to access, control and use the data.' On data sharing versus data trade, UNCTAD states: 'There are significant difficulties in reconciling the notion of national sovereignty traditionally associated with country territories and the borderless nature, globality and openness of the digital space in which data flow. Digital sovereignty is often associated with the need to store data within national borders, but the link between the geographic storage of data and development is not evident. Assigning territoriality to cross-

governance measures will fall into place around it. This approach is also taken by the OECD which also seeks to develop data governance frameworks to stimulate investment.⁸⁸ The G20's proposal centres around interoperability as the first priority, with the harmonisation of norms and rules as a second.⁸⁹ The G20's core statement about data states that 'data is anchored in human identity and should be safeguarded and traded with care', but frames this as a necessary corollary to a system designed to let data flow - i.e. the fundamental assumption is that data flows freely *unless* states mandate otherwise, rather than data may flow freely *if* certain conditions are met. The G20 also accepts a scenario where states disagree on norms and rules for safeguarding the people affected by those data flows, and that data will continue to flow regardless: 'This definition [of data as anchored in human identity], as it is, does not bring up any contestable aspects of 'identity' such as human rights, data privacy, and freedom of information, which may be met with hesitation by some States. Therefore, we believe that such a definition of identity in data can be agreed upon by most, if not all, countries.'

Similarly, the WTO's drafted policy on digital trade⁹⁰ names digital privacy as an area where states will retain controls over trade (p.45), along with state security which is also named as an exception (p.78). The draft agreement does not cite any other rights as meriting controls on free trade, instead working from the assumption that data should flow freely and any controls on the part of states should not be 'burdensome' to business. This market view of data reflects neoliberal values and gives rise to a governance framework in which data is treated as 'terra nullius' (not in principle attached to or implicating the individuals who originate it) in the context of trade.

Despite this overall trend on the international scale, it is worth thinking about the conditions under which data is made excludable by law (which these G20 and WTO governance models sketch out as possible extras rather than fundamental anchor points). What we see in practice worldwide is that when data has effects on those whom it represents, either on the collective or individual level, other reasoning tends to conflict with the pure market vision of data-as-asset. For instance, data has also been defined under various conditions as human labour,⁹¹ human social relations⁹² and a component of human identity-building processes.⁹³ The data circulating in markets is multiple, and this is both the strength and weakness of the data-as-asset framing: it can exist and be processed in several places at the same time, meaning that any approach that gives people 'control' over 'their' data tends to be by its nature partial and illusory. In fact, data currently has an almost infinite lifecycle once it is de-identified, aggregated and reused. This challenges us to think about what we mean

border data flows is also a challenge. Data can be better understood as shared, rather than as traded or exchanged.' See: <https://unctad.org/webflyer/digital-economy-report-2021>

⁸⁸ The 2021 OECD Recommendation on Enhancing Access to and Sharing of Data aims to help governments develop coherent data governance policies and frameworks to unlock the potential benefits of data across and within sectors, countries, organisations and communities by reinforcing trust across the data ecosystem, stimulating investment in data and incentivising data access and sharing, and fostering effective and responsible data access, sharing and use across sectors and jurisdictions. See: <https://www.oecd.org/sti/ieconomy/enhanced-data-access.htm>

⁸⁹ Kateryna Heseleva, Vincent Jerald Ramos, and Alan Ichilevici de Oliveira, 'Towards a Multilateral Consensus on Data Governance,' *G20 Insights* (blog), May 29, 2020, https://www.g20-insights.org/policy_briefs/towards-a-multilateral-consensus-on-data-governance/.

⁹⁰ World Trade Organization, 'WTO Electronic Commerce Negotiations: Consolidated Negotiating Text' (World Trade Organization, December 2020), https://www.bilaterals.org/IMG/pdf/wto_plurilateral_ecommerce_draft_consolidated_text.pdf.

⁹¹ Christian Fuchs, 'Digital Prosumption Labour on Social Media in the Context of the Capitalist Regime of Time,' *Time & Society* 23, no. 1 (March 1, 2014): 97–123, <https://doi.org/10.1177/0961463X13502117>.

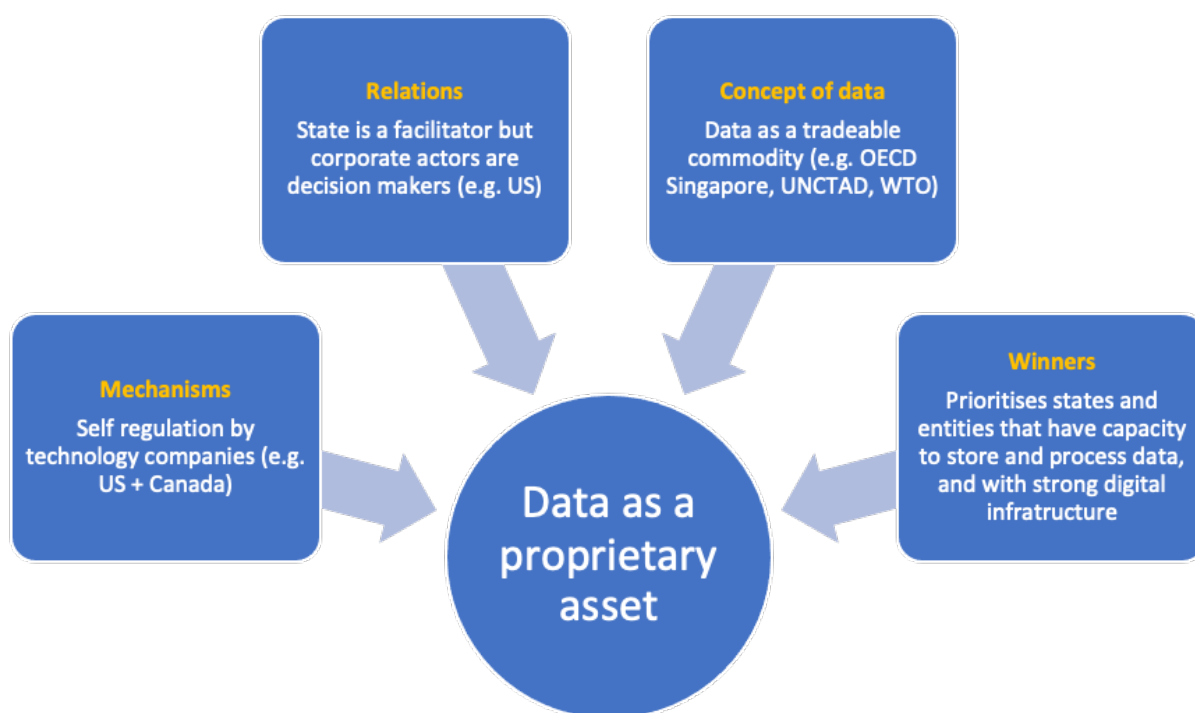
⁹² Ryan Burns, 'Rethinking Big Data in Digital Humanitarianism: Practices, Epistemologies, and Social Relations,' *GeoJournal* 80, no. 4 (August 1, 2015): 477–90, <https://doi.org/10.1007/s10708-014-9599-x>.

⁹³ Luciano Floridi, 'On Human Dignity as a Foundation for the Right to Privacy,' *Philosophy & Technology* 29, no. 4 (December 1, 2016): 307–12, <https://doi.org/10.1007/s13347-016-0220-8>.

when we talk of owning data, controlling it or keeping track of it: this is indeed possible, but only for actors with the storage and processing power to keep and channel data according to their own wishes. These actors tend to be large technology firms, government bodies or other organisations with legal traction on data such as insurance firms or analytics platforms.

Current state legislative frameworks around the world - which are likely to form the basis for international agreements⁹⁴ - are heavily shaped by this asset perspective. Principles of governance not explicitly targeted at data may also have the effect of making data mainly a market good. One example is the EU's establishment of the innovation principle,⁹⁵ which indicates that the path should be smoothed as much as possible for the development of new technologies, including AI. Where there is a conflict with existing regulation and governance, accommodation and flexibility will explicitly be sought. This changes the playing field for AI developers, as their interests are prioritised along with those of citizens and states in ways which will have impacts on how governance architectures are built, and practices emerge, around this new field. This principle has not, at the time of writing, been tested on digital technologies, making AI possibly the first digital domain where the principle will be applied. Meanwhile the US and many other markets have a tacit version of this innovation principle in place, where legislation around technology is aimed at maximising growth and minimising restrictions on the trade of data.

Figure 2 - Features of data as a proprietary model



Source: Authors' own work

⁹⁴ We do not discount the importance of principles developed by international organisations, such as UNESCO's AI principles, in informing lawmaking and policy on AI, but expect that national and regional law will determine what is articulated in trade/adequacy arrangements between countries and regions.

⁹⁵ European Commission, 'Ensuring EU Legislation Supports Innovation,' European Commission, accessed April 28, 2022, https://ec.europa.eu/info/research-and-innovation/law-and-regulations/innovation-friendly-legislation_en.

3.1.3. Personal versus non-personal data

The distinction between personal and non-personal data has grown in parallel with thinking about privacy over more than a century, and is an assumption so embedded in most data governance discussions that it tends to become invisible. It is worth surfacing, however, first because it is the conceptual basis for many claims embedded in models and instruments of data governance, many of which become dysfunctional without it - for example that data should be governed in line with human rights which themselves are based on a liberal ideal of the rational, empowered individual - and second because we are currently seeing a destabilisation of the personal/non-personal distinction just at the point where data governance has come to rely most heavily on it (see section 2d above on the circumstances under which this definition becomes fluid). If there is no clear distinction any more between personal and non-personal data, particularly given the new profiling, inference and categorisation techniques that are available from AI/ML research, then preserving this distinction as a core assumption for data governance is a choice, rather than a necessity.

The twin concerns of individual and group interests in data present continual challenges to both innovation-friendly policy aims and asset-based approaches to governing data because they potentially disrupt the ability to broker and reuse data by bringing into play issues such as consent, but also red lines forbidding uses of data which have negative human impacts. Currently the market-oriented data governance system, both globally and locally, relies on the notion that we can segment data into sensitive personal data (where significant negative impacts on humans are likely if reuse is not constrained); personal data (where data relates to and has impacts on people, but can still be shared and used under most conditions); and non-personal data, which does not relate to people and is therefore, barring property claims, tradeable and reusable.

With the coming of more complex data analytics over the 2000s, the categorisation of what is sensitive, what is personal and what is not personal in terms of data has become eroded and the borders between these are increasingly fuzzy. Profiling techniques make it possible to proxy for sensitive attributes that could not be used directly, and the dialogue between collective profiling and individual impacts is a common technique in both public and private-sector models.

It is important to pin down what 'personal data' is because otherwise we cannot connect to historical claims about why data matters and how it is connected to people. It is also, however, impossible to pin down 'personal data' because it demands that we leave out the collective aspects of data's effects.

3.2. Alternatives to dominant data governance models

In this section, we explore some of the alternative data governance models that aim to challenge the assumptions of the dominant model. These models are mainly proofs of concept or pilots that continue to be developed, but this characterisation has the objective of defining their basic characteristics with which to support alternative perspectives on data governance. First, the *public data trust* concept contests the idea of data as a commodity to be exploited for profit by promoting that public institutions must govern data as a public resource. Second are *data collaboratives* in which the public and private companies collaborate to generate public benefits by using an independent third party that governs the resource. Third, *data commons* suggest that some kinds of data can usefully be defined as 'common pool resources'. Fourth, *data cooperatives* oppose a dominant model centred on monopolistic data controllers by suggesting a model in which data subjects govern data collectively. Fifth, the *indigenous data sovereignty* movement defines data as a resource of indigenous communities and challenges the models that frame data governance as a matter of colonial nation-states. Sixth, *personal data sovereignty* promotes the idea that data

subjects ought not to organise collectively, but rather look for technical possibilities to control their data individually in exchange for benefits or personal autonomy.

3.2.1. Public data trusts

The public data trust is an alternative data governance model in which a public institution is in charge of managing data. In this model, the public actor accesses, aggregates and uses the data collected from different sources. The main agreement between data subjects and public actors is a trust relationship that depends on public engagement using consultations, strong accountability mechanisms and collective benefits.⁹⁶ The main value of this model is the public interest that could translate into the use of public data for policy making, social innovation and to address social challenges.⁹⁷

This model is a reaction against the dominant model that defines data as a commodity that should be exploited for maximum profit. In this case, the data is defined as a public interest infrastructure.⁹⁸ Therefore, it is also possible that public institutions share data with third parties with strong agreements in place to ensure public benefits and strong accountability. Likewise, it is possible to require data that be provided to the public trust even if it was privately collected to socialise its value.⁹⁹

The public data trust continues to be represented by a series of pilots rather than a consolidated model. The model was popularised by the efforts of the Digital Plan launched by Barcelona in 2018. The city government organised a digital transformation roadmap to treat data as a key infrastructure for reaching more democratic decisions, improving public services and empowering people.¹⁰⁰ Likewise, the city implemented 'data sovereignty' clauses in public service contracts to enable data access by public service providers. The local government also aimed to construct a cryptographic alternative to allow citizens to share data with the trust and, at the same time, open the data to local companies, cooperatives and social sector companies. Another example is the DECODE initiative that includes the governments of Amsterdam and Barcelona that constructed three pilots to contribute to data governance based in public entities. The initiative aims to include open standards for public data and the creation of privacy-preserving apps for digital democracy tools.¹⁰¹

3.2.2. Data collaboratives

The data collaborative could be framed as an alternative model rather than a way to resist the dominant model which, as we developed in the previous section, is mainly market-based approaches to data governance. The idea of a data collaborative is that private data collected by companies is pooled with public data through an independent third party. The data is governed by the third party and data access, sharing and use is restricted to the members of the partnership. The data pool is mainly used for projects that benefit the development of public policy and empower

⁹⁶ Marina Micheli et al., 'Emerging Models of Data Governance in the Age of Datafication,' *Big Data & Society* 7, no. 2 (July 2020): 205395172094808, <https://doi.org/10.1177/2053951720948087>.

⁹⁷ Bruno Carballa Smichowski, 'Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions,' *Intereconomics* 54, no. 4 (July 2019): 222–27, <https://doi.org/10.1007/s10272-019-0828-x>.

⁹⁸ Marina Micheli et al., 'Emerging Models of Data Governance in the Age of Datafication,' *Big Data & Society* 7, no. 2 (July 2020): 205395172094808, <https://doi.org/10.1177/2053951720948087>.

⁹⁹ Evgeny Morozov and Francesca Bria, *Rethinking the Smart City: Democratizing Urban Technology* (Rosa Luxemburg Stiftung, 2018).

¹⁰⁰ Ibid.

¹⁰¹ Theo Bass and Rosalyn Old, 'Common Knowledge: Citizen-Led Data Governance for Better Cities,' Text, DECODE, February 17, 2020, <https://www.decodeproject.eu/publications/common-knowledge-citizen-led-data-governance-better-cities>.

members by using data that was not previously accessible. The data collaboration generally takes the form of a public-private partnership that governs access to, sharing and use of data.¹⁰²

Data collaboratives are more developed than other models covered in this section because they reproduce the basic principles of the dominant model like the centrality of data controllers in governance. The NYU GovLab has created a database with more than 200 data cases of collaboratives, mainly from the health, transportation and humanitarian sectors.¹⁰³

3.2.1. Data (semi) commons

A vision continually interacting with that of the free data market is the data commons - an approach referenced by the EU's legal framework on data in the proposals on data altruism and sharing between different sectors discussed in section 5 below. The commons is a powerful imaginary in technology policy and data governance, though it has yet to be realised even in relation to the areas where it has been most discussed – these being primarily scientific research and health data. The commons - a vision of a system based on making data available through shared infrastructures in ways which make it less excludable by economic interests - potentially offers alternatives to pure market assumptions while still allowing for markets for data to exist and for innovation to be prioritised.

The idea of data commons of different kinds has its roots in the fundamental work done by Ostrom and others¹⁰⁴ on how to keep public goods available to the public. Applying this work to the digital realm has mainly resulted in the idea that some kinds of data can usefully be defined as 'common pool resources' - owned by a particular community, where that community can restrict access by outsiders. Gurumurthy and Chami,¹⁰⁵ for instance, quote Benkler (2016) on the notion of the open access commons, which 'guarantee[s] symmetric use privileges to an open general class of users, rather than assigning an asymmetric exclusion right to an individual or known class of individuals'. They argue for a 'semi-commons' approach where individuals, public agencies and legal persons such as corporations or data altruism intermediaries all have different and conditional rights over data, within the parameters of privacy rights.¹⁰⁶ They centre the idea of 'distributive integrity' as an update to Helen Nissenbaum's 'contextual integrity' norm for data governance,¹⁰⁷ where 'the local economy and its regenerative potential', rather than the interests of the most powerful market actors, determine how the commons should function.

This could mean that, for instance, particular digital data resources (however defined) could be governed as the common property of EU states and their residents, and varying degrees of access to this common resource could be granted by a central stewardship process. The notion of the commons, although it has roots in discussions of land and, in more recent classic work, fisheries, requires some reworking for digital resources because data requires both material and digital

¹⁰² Mozilla Insights, Jonathan van Geuns, and Ana Brandusescu, 'Shifting Power Through Data Governance' (Mozilla, September 2020).

¹⁰³ THEGOVLAB, 'Cases | Data Collaboratives,' accessed April 25, 2022, <http://datacollaboratives.org/explorer.html>.

¹⁰⁴ Elinor Ostrom and Charlotte Hess, 'Private and Common Property Rights,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 29, 2007), <https://doi.org/10.2139/ssrn.1936062>.

¹⁰⁵ Anita Gurumurthy and Nandini Chami, 'Governing the Resource of Data: To What End and for Whom?,' Working Paper (IT for change, 2022), <https://itforchange.net/sites/default/files/1741/WP23-Governing-the-Resource-of-Data-AG-NC.pdf>.

¹⁰⁶ This model is currently being explored in the Decode platform, based in Barcelona (<https://ec.europa.eu/research-and-innovation/en/projects/success-stories/all/enabling-citizens-take-control-their-own-data>).

¹⁰⁷ Helen Nissenbaum, 'Contextual Integrity Up and Down the Data Food Chain,' *Theoretical Inquiries in Law* 20, no. 1 (2019): 221–56, <https://doi.org/10.1515/til-2019-0008>.

architectures in order to be managed and shared, creating more complex routes through which community members' access to resources can be shaped.

It is important to note that there are many different versions of the commons approach in relation to data, ranging from open-access commons where no restrictions are placed on others' use of the resource, to limited commons based on, for example, scientific or health-sector usership. As discussed in section 5 below, the EU's data governance architecture as currently under construction is gaining some features of a common-pool-resource philosophy, particularly in the cases of proposed legislative instruments to channel data from business to government and vice versa, or from either of these to society in general. The proposed Data Act and Data Governance Act both aim to tackle cases in which 'the social value of pooled data is higher than the sum of private values of segmented data'¹⁰⁸ (i.e. segmented by being proprietary to different actors), by forcing the creation of a common pool resource structure under certain conditions in order to allow businesses to share in each other's data resources.

There is no single notion of 'the data commons', given the balancing between priorities outlined at the start of this section. Elinor Ostrom warns anyone interested in institutionalising commons-oriented approaches that 'no institutions generate better outcomes for the resource and for the users under all conditions',¹⁰⁹ something that should be taken seriously as part of this balancing process. Her warning implies that, as is currently under discussion in the EU, such an approach needs to be combined with contextual and domain-relevant instruments to make sure that the resource behaves in the interests of the expected beneficiaries.

Gurumurthy and Chami¹¹⁰ examine how the notion of the commons is employed by the Indian proposal for non-personal data governance, which states that such data 'are a nation's or community's collective resources as arising from their natural and/or social spaces, and should be governed as such'.¹¹¹ They argue that data stewardship models advocated by the private sector and its representatives as a way to establish a commons for the public good actually tend to both legitimise the corporate claim to ownership of data with public value, and to reinforce inequality of access to – and power over – data by centring guardianship with the most powerful interests.¹¹²

3.2.2. Data cooperatives

Data cooperatives are alternative data governance models in which data subjects are the main actors that organise the use, sharing and access to data through the collective organisation of particular groups or communities.¹¹³ The main value of this model is the public interest that

¹⁰⁸ Bertin Martens and Bo Zhao, 'Data Access and Regime Competition: A Case Study of Car Data Sharing in China,' *Big Data & Society* 8, no. 2 (October 27, 2021): 1–11, <https://doi.org/10.1177/20539517211046374>.

¹⁰⁹ *ibid* p.1

¹¹⁰ Anita Gurumurthy and Nandini Chami, 'Governing the Resource of Data: To What End and for Whom?,' Working Paper (IT for change, 2022), <https://itforchange.net/sites/default/files/1741/WP23-Governing-the-Resource-of-Data-AG-NC.pdf>.

¹¹¹ Committee of Experts on Non-Personal Data Governance Framework. (2020, December 16). Report by the Committee of Experts on Non-Personal Data Governance Framework. Ministry of Electronics and Information Technology, Government of India. https://static.mygov.in/rest/s3fpublic/mygov_160922880751553221.pdf

¹¹² One example of this process is how the Global Data Access Framework, led by UN Global Pulse, has been positioned as providing a basis for data sharing for meeting the SDGs. See <https://www.unglobalpulse.org/blog/>: 'The GDAF also sees itself as contributing towards digital public goods by aiming ' to transform data discoverability by establishing a global platform for data sharing that utilizes various protocols for exchanging data. The GDAF will rely upon a state-of-the-art reference architecture that will be developed through a collaborative multi-stakeholder effort and will enable data to be discovered by AI systems more easily. Through this mechanism, the GDAF intends to enable entrepreneurs, academics, and governments to unlock the potential for big data and AI for good.'

¹¹³ Mozilla Insights, Jonathan van Geuns, and Ana Brandusescu, 'Shifting Power Through Data Governance' (Mozilla, September 2020), <https://assets.mofoprod.net/network/documents/ShiftingPower.pdf>.

translates into collective protection and empowerment of the community.¹¹⁴ In this case, the members of the community voluntarily provide data to the pool. The community members also participate in the construction of the governance model based on shared values including norms and procedures to limit the use, sharing and access to collective data.¹¹⁵

This perspective emerged as a response to two trends of the hegemonic data governance model. On the one hand, it was a response to the narrative of defining data as a market commodity. Data cooperatives define data as a commons that should be managed by and for the community.¹¹⁶ On the other hand, data cooperatives aim to challenge data governance models based on top-down solutions through decentralised bottom-up governance models. Likewise, data cooperatives aim for transparency and openness around the value of data as a reaction to the ownership and secrecy that rules in the dominant model.¹¹⁷ The objective of this shift is to address the power imbalances that are fundamental in the hegemonic data governance model in which data owners exercise control over data subjects.¹¹⁸

Data cooperatives are diverse, so their interests range from privacy and labour to social concerns like autonomy and sustainability. One of the most common data cooperative models aims to challenge the dominance of platforms and share the profits of their activities fairly among members. This category could include initiatives like FairBNB, where hosts and local communities share the profits of the house-renting services,¹¹⁹ platform cooperatives for gig workers of delivery and transportation apps like Co-Op Ride¹²⁰ and TURPI,¹²¹ and cooperatives of creators like Resonate for musicians¹²² and Stocksy for photographers.¹²³

3.2.3. Indigenous data sovereignty

Data sovereignty, rather than being a model per se, is a principle used to portray different ways of governing data. The general idea is that data is an infrastructure that should be managed according to the norms and values of data subjects. This idea is inspired by proposals for technological sovereignty framed in the context of the interests of national states that are recognising that digital technologies are fundamental assets for which they should not depend on foreign powers.¹²⁴ Even though the original proposal relates to nation states limiting their dependence on other states, the idea has been reinterpreted by people with the aim of promoting self-determination and rebalancing the current power relationship between data subjects and data controllers.

¹¹⁴ Ibid.

¹¹⁵ Chih-Hsing Ho and Tyng-Ruey Chuang, 'Governance of Communal Data Sharing,' in *Good Data* (Amsterdam: Institute of Network Cultures, 2019), 202–13, [https://www.ea.sinica.edu.tw/UploadFile/files/2019%20Governance%20of%20Communal%20Data%20Sharing\(1\).pdf](https://www.ea.sinica.edu.tw/UploadFile/files/2019%20Governance%20of%20Communal%20Data%20Sharing(1).pdf).

¹¹⁶ Jan J. Zygmuntowski, Laura Zoboli, and Paul F. Nemitz, 'Embedding European Values in Data Governance: A Case for Public Data Commons,' *Internet Policy Review* 10, no. 3 (September 30, 2021), <https://policyreview.info/articles/analysis/embedding-european-values-data-governance-case-public-data-commons>.

¹¹⁷ Mozilla Insights, van Geuns, and Brandusescu, 'Shifting Power Through Data Governance.'

¹¹⁸ Zygmuntowski, Zoboli, and Nemitz, 'Embedding European Values in Data Governance.'

¹¹⁹ See: <https://booking.fairbnb.coop/en/>

¹²⁰ Hansa Carney, 'The Co-Op Ride App Could Be Revolutionary for the Gig Economy,' *Kulture Hub* (blog), June 16, 2021, <https://kulturehub.com/co-op-ride-app-gig-economy-problems/>.

¹²¹ Carolina Loza Leon, 'Migrants Faced the Worst of the Gig Economy, so They Made Their Own Delivery App,' *Rest of World*, February 2, 2022, <https://restofworld.org/2022/turpi-delivery-app-ecuador/>.

¹²² See: <https://resonate.is/>

¹²³ See: <https://www.stocksy.com/>

¹²⁴ Stéphane Couture and Sophie Toupin, 'What Does the Concept of 'Sovereignty' Mean in Digital, Network and Technological Sovereignty?,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, January 22, 2018), <https://doi.org/10.2139/ssrn.3107272>.

Data from indigenous communities has long been used historically by colonial powers to exercise control over the lives of indigenous people.¹²⁵ The debate about data sovereignty in this context is part of a long-standing fight by communities over their resources. In the current data governance model, indigenous communities are denied autonomy or are forced to rely on outsiders to access and use data about themselves. Therefore, the indigenous data sovereignty movement promotes the idea of managing information about their peoples, territories, lifeways and natural resources according to their laws, practices and values.¹²⁶

Indigenous data sovereignty is based on the UN Declaration on the Rights of Indigenous Peoples (UNDRIP), which is an international instrument adopted by the United Nations in 2007 that defines the minimum requirements for the wellbeing of the indigenous communities.¹²⁷ In Articles 18 and 19, the document underlines that indigenous peoples have the right to participate in the matters that affect them through their means and procedures. In this context, indigenous data sovereignty is an opportunity to put in practice those articles to exercise control of data from communities.¹²⁸

As noted above, indigenous data sovereignty approaches are a movement rather than a full-blown governance model at present. The movement has taken some shape in different parts of the world, but one of the furthest developed propositions is the Maori data governance mode, which is constructed around a series of principles and the struggles of the Maori communities to participate in policies around the integration of administrative data resources in New Zealand.¹²⁹

The Maori communities are framing their fight to exercise sovereignty over data as part of their effort to obtain control over a variety of resources connected to their livelihoods.¹³⁰ The Maori data sovereignty movement constructed principles for data governance that considers control of Maori communities over their data. The principles for Maori data governance are (1) authority over their data, (2) maintaining relationships of their data with the indigenous context, (3) obligations for balancing individual rights with collective interests and accountability, (4) collective benefit building Maori capacities, (5) reciprocity in terms of consent and respect of the Maori culture, and (6) guardianship over data.¹³¹

The Maori data governance movement led to an agreement being reached with New Zealand's statistics office. The agreement aims to construct a Maori-government work programme to include the Maori worldview into the decisions taken across the public data ecosystem in a process of co-design. The progress on expanding this collaborative model into other areas of governance has been complicated but the efforts around administrative data show promising change to construct a collaborative vision for data governance.¹³²

¹²⁵ Raymond Lovett et al., 'Good Data Practices for Indigenous Data Sovereignty and Governance,' in *Good Data*, 2019, 26–36.

¹²⁶ Taylor and Kukutai, *Indigenous Data Sovereignty*.

¹²⁷ Ibid.

¹²⁸ Lovett et al., 'Good Data Practices for Indigenous Data Sovereignty and Governance.'

¹²⁹ Maggie Walter et al., *Indigenous Data Sovereignty and Policy* (Routledge, 2020).

¹³⁰ Ibid.

¹³¹ Te Mana Raraunga, 'Principles of Māori Data Sovereignty,' Te Mana Raraunga Maori Data Sovereignty Network, October 2018, <https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5bda208b4ae237cd89ee16e9/1541021836126/TMR+Ma%CC%84ori+Data+Sovereignty+Principles+Oct+2018.pdf>.

¹³² Maggie Walter et al., *Indigenous Data Sovereignty and Policy* (Routledge, 2020).

3.2.4. Personal data sovereignty




Personal data sovereignty is a model in which data subjects are defined as market agents that aim to control the access, use and sharing of their data. The idea is that data subjects have more choices over their data inside the market to receive more benefits. Therefore, the objectives of personal data sovereignty are to promote greater control over personal data and deliver better services centred on user control over data. The model is dependent on new intermediaries of the data economy that seek trust from individuals to share and transfer personal data.¹³³ This perspective supposes that data subjects are free actors that can rebalance power relationships. Likewise, the model defines personal data as a commodity that should be governed by data users in exchange for better services.¹³⁴

Personal data sovereignty is increasingly being translated from a political ideology into a set of different technologies and technical tools aimed at realising greater control over one's own data. The advent of cryptographic technologies like blockchain, and the libertarian values motivating its promoters, has inspired advocates to pursue decentralised technical systems to achieve 'self-sovereign identity' by which users can assert and trust claims about one's personal data without needing to rely on traditional intermediaries like state agencies. While excitement for these technologies continues to grow, especially with the emergence of so-called Web3, it remains to be seen whether the architectures underpinning personal data sovereignty will achieve the vision of 'trustless trust'¹³⁵ and prove to be as liberatory as its promoters hope, especially as most providers will be for-profit private ventures in the current market-based paradigm.¹³⁶

3.2.5. How these models interact with the dominant narratives over data

In table 3 below, we summarise some of the key aspects of the different models across the following criteria: *purpose*, which is related to the objectives underpinning the existence of a particular data governance model; *beneficiaries*, being the parties and stakeholders for whom a particular model has been designed; *value*, which describes the cultures that each model seeks to promote; and finally *tools*, which are the mechanisms that help to operationalise a particular model.

Table 3 - Features and characteristics of alternative data governance models

Name	Purpose	Beneficiaries	Value	Tools
Public data trusts 	Public interest	Public entities	Accountability, participation, policy-making	Consultation, democratic accountability
Data Cooperatives 	Collective benefit	Civic cooperatives	Shared values of the group	Voluntary participation, decentralised decision making
Personal data sovereignty 	Increased choice, individual autonomy	Business, data subjects	Commoditisation, individual choice	Transfer and exchange, trust mechanisms, data portability

¹³³ Mozilla Insights, van Geuns, and Brandusescu, 'Shifting Power Through Data Governance.'

¹³⁴ Zygmontowski, Zoboli, and Nemitz, 'Embedding European Values in Data Governance.'

¹³⁵ Kevin Werbach, 'Summary: Blockchain, The Rise of Trustless Trust?' *Wharton Public Policy Initiative*, September 20, 2019, https://repository.upenn.edu/pennwhartonppi_bschool/3.

¹³⁶ 'My First Impressions of Web3,' Moxie Marlinspike, July 1, 2022, <https://moxie.org/2022/01/07/web3-first-impressions.html>.

Indigenous data sovereignty 	Collective autonomy	Indigenous communities	Ownership, control	Connecting management of data to cultural, and social values, norms
Data collaboratives 	Participation	Public entities, business	Control of data to members of partnership Innovation	Third party management, partnerships, contracts
Data Commons 	Collective benefit	Sectoral/interest groups	Access to data to create public goods	Sectoral agreements, public infrastructures

Source: Authors' own work

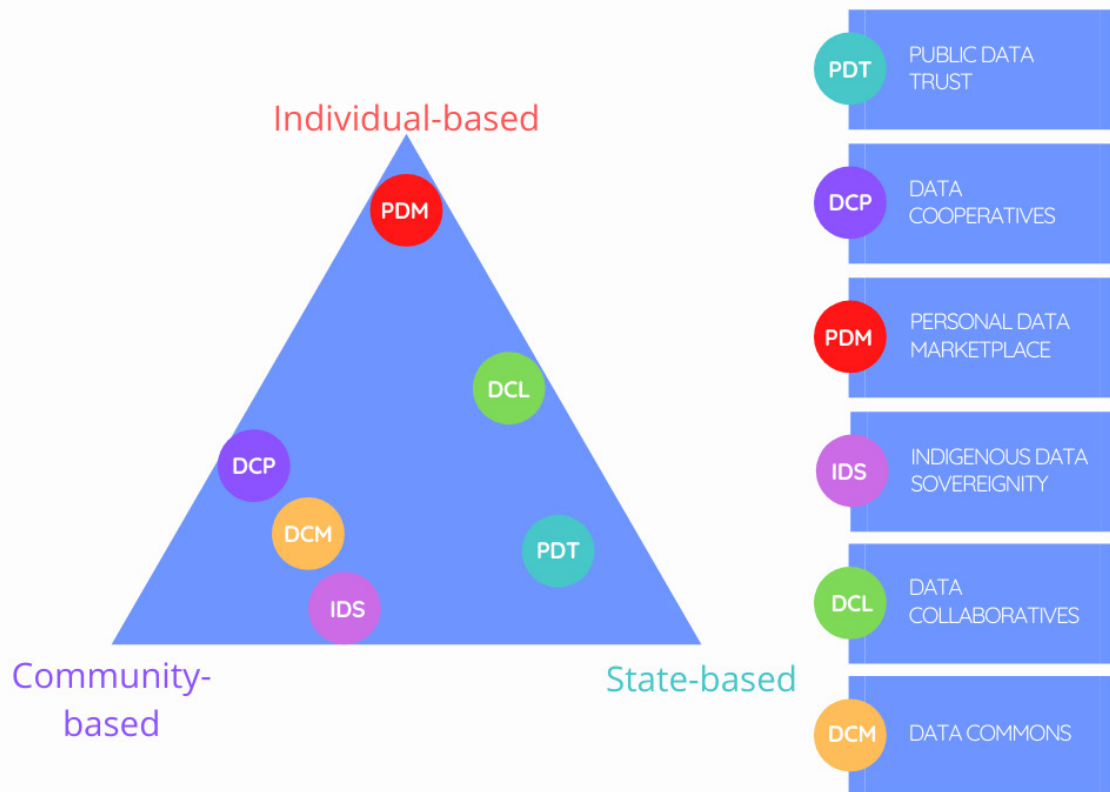
The different examples give us insights into the ways in which data governance has differing motivating purposes. This includes understanding how these resistant/alternative models do not just offer a new institutional apparatus within which to govern data, but a fundamental reimagination of what the purpose of data governance should be.¹³⁷

Data governance models can also be grouped according to the actor that governs the data (see figure 3 below). First, the individual-based initiatives reproduce the idea of *personal data as a commodity*, but are interested in giving more control to data subjects over their data. This model reproduces some ideas of the dominant model: personal data is treated as individual property that requires consent to be exploited. These initiatives consider that the market is the best mechanism to govern data, so it must flow unrestricted to make possible the exchange of the commodity.¹³⁸

¹³⁷ Bruno Carballa Smichowski, 'Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions,' *Intereconomics* 54, no. 4 (July 2019): 222–27, <https://doi.org/10.1007/s10272-019-0828-x>.

¹³⁸ Anita Gurumurthy and Nandini Chami, 'Governing the Resource of Data: To What End and for Whom?,' Working Paper (IT for change, 2022), <https://itforchange.net/sites/default/files/1741/WP23-Governing-the-Resource-of-Data-AG-NC.pdf>.

Figure 3 - Interactions between different data governance models



Source: Authors' own work

Second, other models expect that a public institution governs data as a collective resource to generate public benefits. For this, we have two examples. On the one hand, we have initiatives that expect that public institutions construct the arrangements for governing data. The *public data trust* is an example of such a model that aims to socialise the benefits of data and govern the resource through the public policy mechanisms of accountability and participation. However, this model could face problems of financial sustainability due to the investments required to maintain such infrastructures.¹³⁹

On the other hand, *data collaboratives* leverage public-private partnerships as a way to govern data as a resource. Generally, these partnerships trust in a third party or independent trustee to create the rules. Public institutions are actors that pool data and have voting powers. This model reproduces the idea of data as private property that some companies could own and share with public institutions through an independent body. The model fails to recognise the power relationships inside the collaborative in which private companies have an enormous capacity to exercise control over the decisions of the independent body, limiting possibilities for public accountability and allowing private parties to access public institutions' data.

The third group is the community-based models. Through the examples of data commons, indigenous data sovereignty and data cooperatives, there is an emphasis on thinking in terms of

¹³⁹ Amay Korjan and Vinay Narayan, 'Socializing Data Value: Reflections on the State of Play' (IT for change, July 2021), <https://itforchange.net/sites/default/files/1948/Socializing-Data-Value-Reflections-on-the-State-of-Play-2021.pdf>.

how best to represent the interest of communities and groups.¹⁴⁰ Doing so implies that there is greater acknowledgment not just of values that are important to the group, but also a contextual application of rules and regulations so that it is meaningful. Further, there is also an acknowledgment that within communities there are power differentials, and that identities that underpin the creation and existence of communities need to be accounted for.

In the case of *indigenous data governance*, the identity and historical claims of a community are the basis for claiming sovereignty over data. Therefore, the community approaches for data governance models will be diverse.¹⁴¹ And in the case of the *data cooperatives*, there is an enormous diversity of groups constructed by different interests and a great amount of fluidity, but members share the collective effort of gaining autonomy and collective decision-making over their data. The groups that claim autonomy over data to challenge the dominance of some private actors through collective governance could include gig workers, artists, developers, smaller companies, etc. Data cooperatives by design advocate for voluntary participation, recognising that there is a fluidity to how communities exist, and how interests may change over time.

3.3. Points of interaction between fundamental visions

The systems and approaches described above almost always overlap. Data can be a market asset, an object of state control and sovereignty, and a component of individual identity all at the same time, and thus be covered by multiple legal or governance frameworks simultaneously. Moreover, the four ways of approaching data governance detailed above also interact and coexist. Both market and commons approaches usually take the personal/non-personal distinction into account. For market systems, giving individuals some rights over personal data potentially provides a warning system for when unacceptable uses are taking place which might lead to the disruption of the efficient functioning of the market by human rights claims, for instance. Conversely, rights over personal data are automatically limited in the 'strategic state asset' model when state security comes into question. Examples of this include the use of citizens' private messages as a state intelligence asset by the US government, uncovered by the Snowden revelations in 2014¹⁴² and the use of biometric identification systems to track and control Uighurs in China's Xingjiang Province¹⁴³, or in a less egregious example of state intervention, to control population movements during COVID-19 outbreaks.

As legal researchers point out,¹⁴⁴ any commons-based governance system also has an accompanying system of 'rules-in-use' which determine how conflicts between different interests are resolved on the practical level, much of which centres around the 'interdependence between knowledge flows aimed at creative production and personal information flows'.¹⁴⁵ A commons

¹⁴⁰ Anita Gurumurthy and Nandini Chami, 'Governing the Resource of Data: To What End and for Whom?', Working Paper (IT for change, 2022),

¹⁴¹ John Taylor and Tahu Kukutai, *Indigenous Data Sovereignty* (ANU Press), accessed April 26, 2022, <https://doi.org/10.22459/CAEPR38.11.2016>.

¹⁴² David Lyon, 'The Snowden Stakes: Challenges for Understanding Surveillance Today,' *Surveillance & Society* 13, no. 2 (July 2, 2015): 139–52, <https://doi.org/10.24908/ss.v13i2.5363>.

¹⁴³ Danielle Cave et al., 'Mapping China's Tech Giants' (Australian Strategic Policy Institute, April 18, 2019), <http://www.aspi.org.au/report/mapping-chinas-tech-giants>.

¹⁴⁴ Madelyn Rose Sanfilippo, Brett M. Frischmann, and Katherine J. Strandburg, eds., *Governing Privacy in Knowledge Commons*, Cambridge Studies on Governing Knowledge Commons (Cambridge: Cambridge University Press, 2021), <https://doi.org/10.1017/9781108749978>, 5

¹⁴⁵ *ibid.*, 8

approach, however, is potentially in tension with the 'privacy as contextual integrity' norm,¹⁴⁶ i.e. prioritising people's expectation that data will flow in ways that align with their own expectations. Gurumurthy and Chami argue that a nuanced version of commons governance has the potential to take account of this problem: in their analysis of data stewardship models they demonstrate that these unfairly empower technology firms, making it likely that data will be used in ways not in line with people's expectations or preferences.

This suggests that, under a framework that attempts to make data a common resource while also taking account of contextual privacy, definitions of personal and sensitive data would have to be broadened in order to make it possible to guard people's individual interests in data. A commons approach therefore has implications for most countries' current rules-in-use, which are very much oriented toward the default of making data available to market actors across contextual boundaries (although often in forms which are de-identified or pseudonymised). Although the personal/non-personal data distinction works broadly to define areas in which we wish to respect the connection between data and individuals, and those where other needs take precedence (such as those of innovation and business or government), it is worth considering that the cases outlined above (see section 2a) in which we may find that we care about de-identified data as much as personal data, and wish to treat it with greater care than our current frameworks mandate. This kind of problem fits within an overall commons approach, but suggests that mechanisms for the contextual adaptation and reworking of rules-in-use need to be provided, in order to prevent the system becoming formed in the image of the market at the expense of people.

At the level of existing digital technology governance frameworks, one key component of the EU's approach is that it is risk-based. The risk-based precepts contained in the GDPR and the draft AI regulation require that developers and deployers of digital technologies assess their products, models or services and report to authorities the extent to which they will require scrutiny and protections to make them comply with human rights. Within this, data can be governed according to different principles, whether those of fair competition in the marketplace, fair trading and consumer law, privacy, the protection of personal data, and others. These principles sometimes overlap, but often work independently so that a given application may be governed by multiple instruments at the same time. What characterises the system overall, however, is a pro-market, pro-innovation set of assumptions.

3.4. How broader models are incorporated in current data governance

The way in which the overarching conceptual framings set out above play out in the actual instruments and approaches chosen to enact data governance varies by country and region. The personal/non-personal data categorisation is perhaps the most prevalent globally, largely because in many of the countries still establishing data protection legislation, the need for adequacy agreements leads to a framing that follows the EU's. In the highest-income countries, where rules are now being set out for AI as well as data governance beyond personal data protection and digital markets regulation, the challenge is more complex and requires thinking beyond a single taxonomy of data as personal/sensitive and non-personal, and instead to think about the broader effects of data use and balance the demands of different sectors and concerns. This section looks briefly at the kind of approaches we see emerging in relation to the definitions and related challenges laid out above.

¹⁴⁶ Helen Nissenbaum, 'Contextual Integrity Up and Down the Data Food Chain,' *Theoretical Inquiries in Law* 20, no. 1 (2019): 221–56, <https://doi.org/10.1515/til-2019-0008>.

Domain-specific commons: Data is often governed from outside the frame of 'data governance', and the same may occur to some extent with AI. Arrangements for data often follow ethical and practical norms in fields such as health, security and mobility, in ways defined by the history of those sectors, and which then overlap with the provisions of data protection or consumer law. We see historically and normatively informed processes in the health sector, for instance, around the definition of data as research versus operational data, and patient records versus organisational data, even though the two often speak to each other in terms of use and management. As such, health organisations often handle some version of a data commons where patient data may exist in the form of a fiduciary relationship between doctor and patient, but may simultaneously constitute public health data and become part of a bounded commons arrangement with public health authorities, local and national (for instance, this occurs in many countries with vaccination records, which have different functions over time). Similarly, at different levels of law enforcement data sharing occurs which may resemble a commons approach, governed by different rules within that commons according to whether the data is personal data, whether it is sensitive, whether it is covered by law enforcement rules (such as the EU's Law Enforcement Directive) or general data protection provisions (such as the GDPR), and whether it is being handled by local, national or international authorities. Data on an individual may, for instance, be collected by local police but then become incorporated into a national-level crime investigation, or become part of the records of Europol on a particular issue, and will be covered by different legal provisions, often simultaneously, in each database.¹⁴⁷

Sectoral approaches to data governance: Many countries have evolved a patchwork set of regulations about data over time, as different sectors have developed. The US, Indonesia and many other countries have rules governing commercial data processing activities that are separate from those governing, for example, health or law enforcement. In the EU the sectoral approach can be seen in the division of law enforcement data protection and general data protection. The sectoral approach often leads to, or reinforces, the data-as-asset framing because it focuses attention on (mis)uses of data by the most powerful actors, rather than providing a comprehensive discussion of risks and objectives in the data economy.

Comprehensive legislation on data: This requires picking an overall model, which in the vast majority of cases worldwide results in the categorisation of data into personal and non-personal, and providing protections for personal data while allowing non-personal data to be traded freely, often depending on both sectoral provisions and national strategic concerns. This is compatible with the data-as-strategic asset vision and can be seen, for instance, in India's ongoing formation of a data governance regime based on framing non-personal data as a tool of public and economic policy, and also in China's data governance framework, perhaps the system which most explicitly seeks to define the state's interests in both personal and non-personal data, with (limited) rights over personal data as a tradeoff for the public.¹⁴⁸

The precautionary principle: The most common framing for a comprehensive, market-oriented approach is a risk-based one, placing those who collect and process data in the position of judging its possible harms. In contrast, when applied to data governance the precautionary principle is one main alternative to the risk-based model. It is often discussed as being potentially based on human rights or other sector-specific principles (such as the humanitarian principle of doing no harm, or the life sciences principle of respecting people's autonomy and integrity). This is not the only way

¹⁴⁷ See, for example, the order given by the European Data Protection Supervisor to Europol where they determined that data should not be kept as crime data unless it was determined separately to be so on the level of Europol itself: https://edps.europa.eu/press-publications/press-news/press-releases/2022/edps-orders-europol-erase-data-concerning_en

¹⁴⁸ Rogier Creemers, 'China's Emerging Data Protection Framework,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 16, 2021), <https://doi.org/10.2139/ssrn.3964684>.

to use it, however: for example the Chinese instruments on data governance arguably combine a consent-based model for consumer data with a precautionary approach on data defined as relevant to social cohesion and state security. The precautionary principle is most aligned with the vision of data as a strategic asset and with the notion of personal and non-personal data, given that it may result in the prohibition of particular uses of data - something less possible once data is in a commons infrastructure. It is also less compatible with what might be termed the strong market approach to data, which is predicated on the assumption that data must flow and, in principle, value must be extracted.

Private regulation and self-regulation by data proprietors: The market-based model for data governance is often accompanied by private regulation such as standardisation, voluntary certification and self-regulatory provisions which aim to guide how data proprietors should handle or share data under given circumstances. This self-regulation component has received huge emphasis over the last decade in the EU and US in particular, and has become a lobbying focus for technology firms who would rather be subject to self-regulation provisions than hard law. The data ethics apparatus that grew up over the 2010s in the private sector, but also to some extent the public sector, has been criticised as an 'escape from regulation'¹⁴⁹. It has been criticised as not fit for purpose as a way of effectively governing data and AI on the basis that AI development, in the words of Mittelstadt¹⁵⁰, is a research area that lacks 'common aims and fiduciary duties, professional history and norms, proven methods to translate principles into practice, and robust legal and professional accountability mechanisms'. In the absence of these, Mittelstadt argues, there can be no useful consensus on high-level principles because there is in fact 'deep political and normative disagreement'. One of the chief proponents of a data ethics approach as a tool of (corporate) data governance, Luciano Floridi, wrote in 2021 that 'the time has come to acknowledge that, much as it might have been worth trying, self-regulation did not work.'¹⁵¹

Cooperative regulation: In their approach to data governance, the UK government has advocated for the need to have capacity across different departments in government. Regulatory functions in the UK also combine different technical expertise. For instance the Competition and Markets Authority (CMA) investigates anti-competitive behaviours, the Information Commissioner's Office is responsible for protecting information rights in the UK, while the Office of Statistics Regulation ensures compliance with official statistical standards.¹⁵² The National Audit Office came out with a detailed report outlining the importance of not just specific domain expertise but also cross-department accountability and governance, to ensure that there were clear responsibilities, but also effective communication.¹⁵³ This sort of regulation is premised on the idea of cooperation existing between different departments, and building competencies that do not just reside with one unit, but with several.

¹⁴⁹ Ben Wagner, 'Ethics As An Escape From Regulation. From 'Ethics-Washing' To Ethics-Shopping?', in *Ethics As An Escape From Regulation. From 'Ethics-Washing' To Ethics-Shopping?* (Amsterdam University Press, 2018), 84–89, <https://doi.org/10.1515/9789048550180-016>.

¹⁵⁰ Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI,' *Nature Machine Intelligence* 1, no. 11 (November 2019): 501–7, <https://doi.org/10.1038/s42256-019-0114-4>.

¹⁵¹ Luciano Floridi, 'The End of an Era: From Self-Regulation to Hard Law for the Digital Industry,' *Philosophy & Technology* 34, no. 4 (December 1, 2021): 619–22, <https://doi.org/10.1007/s13347-021-00493-0>.

¹⁵² The Royal Society, 'The UK Data Governance Landscape,' June 2020, <https://royalsociety.org/-/media/policy/projects/data-governance/uk-data-governance-explainer.pdf>.

¹⁵³ Comptroller and Auditor General, 'Challenges in Using Data across Government' (National Audit Office, June 21, 2019), <https://www.nao.org.uk/wp-content/uploads/2019/06/Challenges-in-using-data-across-government-Summary.pdf>.

4. What are the key principles and features of 'good' data governance, and what is our basis for judging this in the EU?

The EU's data governance approach currently draws on a range of overarching principles, sourced from a set of policy instruments, for defining what constitutes good data governance. Several of these principles potentially conflict, so that the question that emerges is one of prioritising and balancing. Data governance in this context becomes a matter of balancing these principles in order to determine how different objectives should interact, how public values feed into governance frameworks in practice, how civil society is allowed to participate in governance, and how people's interests should be represented in governance. Key instruments in play include the European Strategy for Data,¹⁵⁴ whose core proposal is that 'the EU can become a leading role model for a society empowered by data to make better decisions – in business and the public sector.' They also include the upcoming European Declaration on Digital Rights and Principles for the Digital Decade,¹⁵⁵ which focuses on connecting and aligning digital governance with core areas of human rights as established by EU law, EU values (dignity, freedom, democracy, equality, rule of law and human rights)¹⁵⁶ and the European Pillar of Social Rights.

On the economic front, other underlying principles the EU draws on in its vision for data governance include the innovation principle,¹⁵⁷ which 'ensure[s] that all new EU policy or regulations support innovation and that the regulatory framework in Europe is innovation-friendly,' and the principle of the digital single market, which in turn aligns with the core principle of free intra-European trade which forms the historical foundation of the Union. These values relating to the economic growth and wellbeing of EU Member States potentially stand in tension with the rights- and justice-based orientation of the other core value statements to do with digital strategy, for instance in the statement that the innovation principle is legislatively as important as the precautionary principle, which underlies much thinking about digital rights and data protection. For the Council of Europe, for instance, the precautionary principle is centred in the 2019 'Guidelines on artificial intelligence and data protection' issued by the Directorate General of Human Rights and Rule of Law, which proposes ways to align AI regulation with the internationally adopted Convention 108 on data protection: 'AI developers, manufacturers and service providers should assess the possible adverse consequences of AI applications on human rights and fundamental freedoms, and, considering these consequences, adopt a precautionary approach based on appropriate risk prevention and mitigation measures.'¹⁵⁸

How, then, should the EU address the tension between its mission to create and sustain free trade and economic growth for its Member States, and its mission to establish rights and the rule of law in relation to digital technologies? What should be the non-negotiable criteria and who should set them? This tension, reflecting as it does the tension on the global level between free digital markets and the prevention of harm and promotion of fundamental public goods, speaks to very different,

¹⁵⁴ European Commission, 'A European Strategy for Data,' Pub. L. No. COM(2020) 66 final (2020), <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>.

¹⁵⁵ European Commission, 'Declaration on European Digital Rights and Principles | Shaping Europe's Digital Future,' January 26, 2022, <https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles>.

¹⁵⁶ European Commission, 'The EU Values' European Commission, accessed April 28, 2022, <https://ec.europa.eu/component-library/eu/about/eu-values/>.

¹⁵⁷ 'Ensuring EU Legislation Supports Innovation,' Text, European Commission - European Commission, accessed April 28, 2022, https://ec.europa.eu/info/research-and-innovation/law-and-regulations/innovation-friendly-legislation_en.

¹⁵⁸ Consultative committee of the convention for the protection of individuals with regard to automatic processing of personal data, 'Guidelines on artificial intelligence and data protection, T-PD(2019)01' (Directorate General of Human Rights and Rule of Law, January 25, 2019), <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>.

and potentially incompatible, standards for what is 'good governance'. One way to arbitrate between these differing visions is to address them through a justice lens, which asserts that there are tests we can apply to statements about good governance and related models, to see whether they align with the core assumptions that make governance functional for people, and place it at the service of the public rather than in the interests of the most powerful and privileged.

In 2017, the Global Data Justice project (the authors of this report, based in the Netherlands) asserted three fundamental pillars of 'good' data governance.¹⁵⁹ These pillars offer a tool for understanding how public values can be incorporated in governance frameworks. The first – that people should have control over their own visibility – touches on both privacy and representation. It asserts that people should be represented through their data when it is in their interests (for example, when registering to vote, receiving benefits and entitlements, and in contexts where public health or safety is at stake). However, they should also be able to withdraw their participation in systems which render them visible in exploitative ways, for instance extractive practices of commercial profiling, unwarranted surveillance by government or commercial firms, and systems (such as commercial national identification infrastructures) that channel public goods into private wealth.

The second pillar of good data governance is that people should have autonomy over whether they adopt particular technologies or not, and in turn that they should control their terms of engagement with data markets. This implies establishing means of collecting and using data that do not serve the business models of big technology firms, and in turn that non-commercial institutions and organisations, especially civil society organisations which legitimately represent the interests of groups and communities, must play a central role in data governance.

The third relates to harm, namely the notion that although they should be able to make claims, people should not be made responsible for protecting themselves from exploitation through datafied (or automated) systems, given that these are generally invisible to those they affect. Where people cannot be aware of the paths their data is travelling, or the basis on which decisions are being made about them, it becomes the responsibility of the government to ensure that the public is protected, both individually and collectively. This responsibility can then, in the case of the private sector, be imposed on data handlers in the form of regulation that creates accountability - but the key principle is that accountability should be produced through, and aligned with, public and ideally democratic architectures, rather than commercial firms self-regulating according to codes or standards they define and enforce themselves.

These pillars are complementary to other statements about data governance in that they propose a normative underpinning for any governance model. If it has the effect of channelling power and profit toward the best-resourced and most powerful actors, whether governmental or corporate, a governance model is not in line with principles of social justice and requires reorienting. Equally, a system of data governance that does not countenance contestation - however pressing the public needs it addresses - requires further thinking and development. Perhaps most importantly, if a model proposes structures for accountability but these are technology-specific, rather than aligned with the existing architectures through which civil society asserts its agency, the model is unlikely to produce just outcomes.

This justice-based reasoning leads to a set of core benchmarks for good governance of data:

¹⁵⁹ Linnet Taylor, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally,' *Big Data & Society* 4, no. 2 (December 1, 2017): 2053951717736335, <https://doi.org/10.1177/2053951717736335>.

- i. **Preserving and strengthening public infrastructures and public goods:** does the governance model proposed channel benefits principally toward the public, and does it prioritise the integrity of public infrastructures and related public goods?¹⁶⁰
- ii. **Inclusiveness:** does the model envision rights and values as universally applicable, or does it prioritise rights-holders who are also citizens or geographically close to the institutions governing data? Does it conceptualise non-citizens and people in other regions of the world (for example, migrants entering or living in the EU, or people in other regions affected by EU firms' digital trading practices)¹⁶¹ as relevant communities for the values and protections it establishes? Does the understanding of rights and values account for the fact that people have different capacities to claim these rights and entitlements, and explicitly account for these social-political contexts?
- iii. **Contestability and accountability:** does the governance model assume that people may disagree, resist, refuse and bring claims which need forms of redress? Are there mechanisms to allow data and technology policy to respond to social and political change, and does the model offer ways of recognising and negotiating about dissent and contestation? Does the model align with principles and processes of democratic accountability through the institutional mechanisms it offers?
- iv. **Global responsibility:** does the model only look inward, for instance toward the EU Member States, or does it take into account norm-making effects on external actors in the global, political and economic landscape? The EU plays a de facto role in international standard-setting because other nations have to align with the norms conveyed by its legislative instruments in order to trade with the bloc. This regulatory convergence means that the norms contained in the EU's data governance model will have practical effects on decision making elsewhere (positive or negative) and therefore have a diplomatic character as well as an internal one.¹⁶² This means that the EU has a responsibility not to establish harmful norms around data and AI, and to consider the practices its data governance may make possible elsewhere.

The implications of these principles for good governance are explored below.

¹⁶⁰ One prominent example of this problem is the privatisation of South Africa's welfare payment distribution services in 2012 to a technology firm, CPS, which rendered many welfare recipients destitute by channelling deceptive offers from its subsidiary firms, a problem which then could not be effectively redressed because of the dependency that had been created on the digital provider's capacity: See: Roy Cokayne, 'We Could Be Stumbling into a Digital Welfare Dystopia: Open Secrets,' *TechCentral* (blog), December 7, 2021, <https://techcentral.co.za/we-could-be-stumbling-into-a-digital-welfare-dystopia-open-secrets/205646/>.

¹⁶¹ For instance, the US's data governance practices have been challenged by EU claimants for channelling data from EU users into the country's intelligence apparatus, although they cannot claim (digital or privacy) rights in the US: Kevin Cahill, 'Max Schrems: The Man Who Broke Safe Harbour,' *ComputerWeekly.Com*, June 10, 2015, <https://www.computerweekly.com/feature/Max-Schrems-The-man-who-broke-Safe-Harbour.r>

¹⁶² When WhatsApp announced a change in its privacy policy last year over sharing data with Facebook, Europeans were allowed an opportunity to opt out of it due to the GDPR - Ahmed Yasmin, 'Why WhatsApp Users in Europe Can Opt-out of New WhatsApp Privacy Policy but Users in India Cannot?,' *India Today*, April 21, 2021, <https://www.indiatoday.in/technology/news/story/why-whatsapp-users-in-europe-can-opt-out-of-new-whatsapp-privacy-policy-but-users-in-india-cannot-1793555-2021-04-21>. However this was not the case in different parts of the world. In India for instance, WhatsApp users moved in large numbers to other platforms such as Signal. Sindhu Hariharan, 'WhatsApp's Privacy Policy Pushes Users to Signal, Telegram,' *The Times of India*, January 9, 2021, <https://timesofindia.indiatimes.com/business/india-business/whatsapp-privacy-policy-pushes-users-to-rivals/articleshow/80178485.cms>. The Indian Government also wrote to WhatsApp to remove its discriminatory policy citing the fact the Indian customers did not have the opportunity to opt out and using the European case as a benchmark.

4.1.1. Preserving and strengthening public infrastructures and public goods

This issue goes to the equitable distribution of resources, but also to the preservation of the functionality of public (digital) infrastructures which in turn provide public goods. We should also consider here the creation of what has been termed 'public value'. In terms of the first of these two, these infrastructure-related goods include the social safety net, scientific knowledge,¹⁶³ public education and healthcare, access to justice, electoral processes and law enforcement. Where these become delegated to non-state actors in the process of digitisation, rather than building state capacity and, with it, the apparatus of accountability between state and people, these public goods are potentially at risk and should be centred in discussions about process and regulation of such transfers of responsibility.¹⁶⁴ The term 'public value' refers to the creation of non-economic forms of value for society.¹⁶⁵ It has many possible interpretations, but is different from the concept of the 'public interest' and instead covers the process of creating definable value, for instance through the creation of public goods such as education, social welfare systems, public transport or public health.

Establishing a data governance model therefore also represents a statement of the public interest: it requires a position on the extent to which public services should be provided through public infrastructures and what 'public' should mean in practice for instance in terms of transparency policies, monitoring mechanisms and avenues for the redress of grievances.¹⁶⁶ Establishing a data governance model by extension also informs architectures: it determines who has the right to construct infrastructures that directly impact the public, whether and how these should be made accountable to the public, and how publicly owned systems should interact with privately owned ones, and to whose benefit.¹⁶⁷ Engaging with this notion of 'publicness' also means to discern how the treatment of data as a commodity results in subjecting it to skewed market power where the dominance of a few players sets the agenda for others within the data economy. Building a level playing field requires ensuring not just that certain functions retain a public character, but also that there remains a capacity for other actors such as civil society to challenge the publicness of functions. Thereby, if they fall short in their standards such as in terms of accountability there remains a capacity to respond and monitor.

¹⁶³ For example, the Rector of the University of Amsterdam has called for a 'digital university act' to maintain the independence of universities as producers of public goods. She calls for publicly run digital infrastructures, open access publication, for universities to control the digital tools they use, and for platforms to provide access to data for research purposes. See: Karen Maex: 'Protect independent and public knowledge.' Speech January 8th 2021 https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiavscj5_D4AhVL_aQKHQzEA64QFnoECAUQAQ&url=https%3A%2F%2Fwww.uva.nl%2Fbinaries%2Fcontent%2Fassets%2Fuva%2Fnl%2Fover-de-uva%2Fspeech-karen-maex---dies-2021.pdf&usq=AOvVaw2V5_UZK8444_8hFUU2XU9s

¹⁶⁴ See the work of the Global Data Justice project on pandemic-related commercialisation of public goods: <https://globaldatajustice.org/sphere-trans/>

¹⁶⁵ For further explanation, see for example Timo Meynhardt, 'Public Value Inside: What Is Public Value Creation?', *International Journal of Public Administration* 32, no. 3–4 (March 19, 2009): 192–219, <https://doi.org/10.1080/01900690902732632>.

¹⁶⁶ Mittelstadt makes the case that professionals conducting public tasks (including lawyers and doctors) have particular accountability architectures based on fiduciary duties and ethics which are not present in technology firms: 'Commitment to public service: professions involve a public declaration to provide a service to society or for the public good, making use of specialized, often privileged expertise that takes precedence over individual gain.' Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI,' *Nature Machine Intelligence* 1, no. 11 (November 2019): 501–7, <https://doi.org/10.1038/s42256-019-0114-4>.

¹⁶⁷ For more on this, see: Linnet Taylor, 'Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector,' *Philosophy & Technology* 34, no. 4 (December 1, 2021): 897–922, <https://doi.org/10.1007/s13347-020-00441-4>.

This imperative to balance the power and needs of the public against those of the private sector requires that norms, structures of democratic representation and laws have to be placed in dialogue with each other. Attempting to govern data and AI, for instance, primarily through private standards bodies and certification by private actors would tend to privilege a neoliberal vision of the public interest which is subject to market-driven interests, but would not be easily updated to deal with new applications and systems, and which often fail to take account of democratic process. Finding a way to extend this process to technology governance is one of the priorities for the coming decades since without it, any governance model cannot be tested against the notion of the public interest and - given the prevailing power of technology giants and their interest in shaping governance - will suffer challenges to its legitimacy.

4.1.2. Inclusiveness

An inclusive perspective on data governance demands that any model centre justice and take into account the interests of all who will be affected by it, rather than simply a body of citizens in a particular state or region. This broad definition of whose data matters therefore includes, for instance, non-citizens, people who are suspected of crime or considered security risks, and people who are beneficiaries of social welfare programs - all groups whose agency is commonly disregarded when it comes to developing and deploying data analytics and AI. Data regulation tends to cite both human rights and market efficiency, the former of which aligns well with a social justice-oriented governance model. For the latter, this is harder because it is based on the notion of rational, fundamentally equally equipped participants in the market. In contrast, a social justice perspective starts from the assumption that all societies are unequal playing fields where people's circumstances give them different degrees of agency and ability to claim their rights. A good governance model, therefore, should conceptualise *who* matters and *how*, rather than only what data matters.

A social justice-oriented data governance model suggests additions to our way of conceptualising vulnerability in relation to technology. Beyond the notion that people are more easily harmed if they have certain minority or sensitive attributes, a social justice view would add another facet to these considerations: that of unequally distributed power and agency. In this view, data technologies and AI are often constructed in exploitative ways, and will amplify these inequalities, creating vulnerabilities where they may not inherently exist.

An inclusive vision of protection from harm is also an important component of an inclusive data governance regime. The contexts in which data causes harm have changed radically over time, from the analogue and early digital privacy harms of the 20th century to a set of harms including facial recognition, pseudoscience such as physiognomic AI,¹⁶⁸ and collective and more political threats such as deepfakes and online health-related misinformation. Data governance has not always kept up with them, particularly where the identification of harm is made dependent on definitions from other domains - for example misinformation as currently conceptualised by the proposed Digital Services Act covers almost exclusively child sexual images and content advocating for terrorism, requiring revision of other instruments in order to apply it to pandemic-related misinformation - something which has posed a large-scale threat to the public over the course of the COVID-19 pandemic.

Data governance implicates every data that can be treated as an economic commodity, a surveillance tool to protect national security, or linked to privacy and identity rights.¹⁶⁹ As long as

¹⁶⁸ Luke Stark and Jevan Hutson, 'Physiognomic Artificial Intelligence,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, September 20, 2021), <https://doi.org/10.2139/ssrn.3927300>.

¹⁶⁹ Rebekah Dowd, *The Birth of Digital Human Rights*, 2022, <https://link.springer.com/book/10.1007/978-3-030-82969-8>.

the data creates social or economic value or risks downstream, the interests of the individuals and groups that share relevant features with the data subjects and data producers should be factored into how data governance allocates privileges and duties among actors in the digital economy. The connection envisaged in the preceding paragraph has a lot to do with data relativity, i.e. how data connects people to one another (horizontal data relations) and how these connections produce social effects that are hardly recognised under existing data governance frameworks.¹⁷⁰

If a good data governance model should be inclusive, how do we determine the extent of inclusiveness, such as the variety of data governed by the model and the communities to include, in this case, citizens generally, also non-citizens, and citizens in their capacity as consumers, businesses and public authorities?

Public and private sector controlled data matters in this space. A public intervention that relies on data as an instrument for social policy should be of interest. With regard to social policy, it usually is a top-down approach, with public authorities deciding with little or no input from those who are expected to benefit from the interventions. A case in point are the Dutch municipalities attempting data-driven initiatives for social benefits services, youth care services, debt support, care for vulnerable people and early school dropout.¹⁷¹

The procedures in a data governance model respond to historical and cultural arrangements. The current dominant model of economic exploitation of data was constructed around the use of information infrastructures with colonial heritages.¹⁷² How data is collected, shared and used is part of a long-standing trajectory of observing, collecting and sorting data from colonial endeavours involving long-distance interaction and sponsored by powerful actors like the state or corporations.¹⁷³ Even though we try to disconnect the digital technologies from their history, the long-term continuities between these technologies and archival and statistical practices are present in the way in which data is collected, organised, shared and used. The prediction of exploitation and discrimination of certain groups would come from the legacies of exclusion rather than futuristic claims.

For example, the practices of colonial datafication in South Africa during British rule shaped the implementation of biometric technologies in the country after independence. The colonial state in South Africa has been described as a gatekeeper whose activity was focused on controlling trade and enforcing racial divisions. This construction facilitated the development of information infrastructures centred on knowing and governing colonial subjects to construct racial divisions, the rights that could be entitled, and the allocation of labour. Therefore, the colonial state was eager to construct architectures of identification that could help to put in practice the gatekeeper function of the state. Biometric identification technologies, mainly fingerprinting, were developed first in India and they were then brought to South Africa by the Indian Colonial Service. Both India and South Africa had a special place as twentieth-century laboratories for the empire for technologies of racial segregation. Following this colonial effort, the governments after the independence tried

¹⁷⁰ Salome Viljoen, 'A Relational Theory of Data Governance,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 11, 2020), <https://doi.org/10.2139/ssrn.3727562>.

¹⁷¹ Liesbet van Zoonen, 'Data Governance and Citizen Participation in the Digital Welfare State,' *Data & Policy* 2 (ed 2020), <https://doi.org/10.1017/dap.2020.10>.

¹⁷² Payal Arora, 'Decolonizing Privacy Studies,' *Television & New Media* 20, no. 4 (May 1, 2019): 366–78, <https://doi.org/10.1177/1527476418806092>.

¹⁷³ Elena Aronova, Christine von Oertzen, and David Sepkoski, 'Introduction: Historicizing Big Data,' *Osiris* 32, no. 1 (September 2017): 1–17, <https://doi.org/10.1086/693399>.

to construct the biometric database repeatedly, but it was not until the emergence of digital technologies when that was possible.¹⁷⁴

4.1.3. Contestability and accountability

Contestability and accountability are key to the functionality of a data governance model. They serve as the link between those governing data, those using it and those affected by it, and therefore also constitute a barometer which tells us whether the model is achieving its aims or not. Contestability and accountability are linked: if there are structures for accountability then people who are negatively affected will make this known, and conversely, it is only through claims that accountability is realised.

There are many forms of contestability relating to data and AI systems, some (but not all) of which connect to structures of democratic accountability. It is important to think beyond state-based democratic process, which is the ideal but does not apply to many instances where data governance has to provide accountability. Sometimes accountability is internal within companies or public bodies, for instance in relation to content moderation or unfair treatment. Often, this accountability needs to be linked to broader political and legal forms of accountability in order to have meaning. In other cases, contestation is weakly linked to accountability: for instance, when refugees and certain subjects of humanitarian action experience harms relating to data or AI, there is a gap in terms of meaningful accountability because in many crisis contexts international organisations are mostly accountable to states and much less so to the people they directly intervene on.

Common denominators for accountability and contestability processes in different contexts include transparency with regard to what systems and data are being used, who is making decisions within the system, and who is liable for any harm caused. The issue of contestability overlaps with that of inclusiveness - most often those who are not conventionally envisaged as subjects of data governance (such as migrants, suspected criminals and children) are also those unable to bring claims when harmed. For instance, migrants subjected to intrusive or excessive processes of border control could in theory claim that the technologies in question are illegitimate, but do not have the standing to do so.

For meaningful accountability to be feasible, several features are necessary. First, data governance should be treated as a problem that affects collectives as well as the individuals who are currently those entitled to make rights claims. Conceptualising the public, as well as individuals, in relation to data collection, use and the distribution of benefits would necessitate governance that included the possibility of institutional responses to data production and exploitation.¹⁷⁵ Second, taking account of the temporal dimension of data governance, it matters that infrastructures and processes exist for identifying problems and recognising claims at different stages of an intervention or application.

Third, the system must provide ways to equalise power differentials around data and AI interventions. Unless people are legally, structurally and institutionally empowered to dissent or not to consent to a given application of technology in the first place, any consent obtained from them is invalid. The forms of consent currently used in relation to commercial data collection systems, for example, are mostly invalid on the basis that people cannot be properly informed about the possible

¹⁷⁴ Keith Breckenridge, *Biometric State: The Global Politics of Identification and Surveillance in South Africa, 1850 to the Present* (Cambridge: Cambridge University Press, 2014), <https://doi.org/10.1017/CBO9781139939546>.

¹⁷⁵ Salome Viljoen, 'A Relational Theory of Data Governance,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 11, 2020), <https://doi.org/10.2139/ssrn.3727562>.

results of their agreement.¹⁷⁶ Theorists of the social contract distinguish between situations where people consent to an intervention unwillingly out of necessity of constraint ('convention consent'), versus situations where people offer 'endorsement consent' and willingly cooperate because they agree that the intervention is in the public interest.¹⁷⁷

An example of unwilling consent in data governance can be seen in the draft AI Act's description of the way people are expected to consent to physiognomic AI (AI developed to draw emotion or personality inferences from people's facial features, including 'emotion recognition' models). This group of techniques is a common type of AI application, but one that is not in line with scientific evidence, and is categorised by the Act as a practice that should be regulated mainly through transparency and informed consent. In the words of the draft Act, 'This allows persons to make informed choices or step back from a given situation.' (p.15). However, the private-sector contexts in which physiognomic AI is most often used - recruitment, examination monitoring and others - are situations where people by definition cannot 'step back' without the loss of a potential job, of a qualification they have worked for, or other opportunity. This constitutes forced consent and fails the contestability test. Similarly, the use of biometric analysis or 'emotion detection' AI on forced migrants who need to cross a border for their own safety, or on students who need to take an exam in order to gain a qualification (and often to keep a visa), also fail the contestability test.

A final aspect of contestability, which may appear quite radical within the dominant paradigm, but which nonetheless merits critical consideration, is the possibilities for resisting the power of markets and private actors to shape data governance priorities. Among other concerns, this means challenging the market-based framing for policy relevance in data governance debates, questioning why certain courses of action are deemed irrelevant, impractical or infeasible by market actors and fundamentally questioning the role of the market (or markets, as it were) in setting the vision and terms for future data governance. Current modes of contestation do not provide sufficiently material ways to challenge the power of private actors other than through accusations of anticompetitive practices or consumer rights violations. A data justice perspective demands that we broaden our conceptualisation of contestability beyond claims based on competition and consumer protection regulation.

4.1.4. Global responsibility

Global responsibility in terms of data governance in our understanding is not a responsibility to offer templates and transplants of how to build data governance frameworks for different regions around the world. Rather it is an acknowledgment that the decisions taken by the EU have a lifecycle and effects that go beyond governing within its jurisdiction. This effect - already articulated by the EU in relation to its digital strategy¹⁷⁸ - is not just on account of extra-territorial impacts of legislation but also on account of the fact that very quickly the EU regulatory approach often becomes the global standard for regulation elsewhere.

¹⁷⁶ See, for instance, the judgement of the Belgian data regulator on website consent platform IAB in 2022: Natasha Lomas, 'Behavioral Ad Industry Gets Hard Reform Deadline after IAB's TCF Found to Breach Europe's GDPR,' *TechCrunch*, February 2, 2022, <https://social.techcrunch.com/2022/02/02/iab-tcf-gdpr-breaches/>.

¹⁷⁷ Linnet Taylor, 'Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector,' *Philosophy & Technology* 34, no. 4 (December 1, 2021): 897–922, <https://doi.org/10.1007/s13347-020-00441-4>.

¹⁷⁸ 'Europe must now strengthen its digital sovereignty and set standards, rather than following those of others – with a clear focus on data, technology, and infrastructure.' https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en

The dominant model for data governance translates into a reproduction of the global inequality in which an identifiable group of powerful actors control data.¹⁷⁹ For example, trade agreements between the EU and other countries have historically been constructed in unfair terms. It is argued that these kinds of agreements allow firms to access data from the Global South to exploit it and, at the same time, externalise the privacy damages of aggressive data technology development.¹⁸⁰ For this reason, it is fundamental to keep in mind the effects of promoting data governance models in countries outside Europe – we argue that the models need to take a plural understanding of values that underpin them. For instance in the case of the value of rule of law, it is necessary to view it not just as a European value, but rather as an essentially contested concept that has different meanings in different regions.¹⁸¹

At one level, in a very narrow sense, the rule of law can be seen as essential to maintaining law and order, but at a broader level it can be understood as the basis to ensure human rights, secure democratic institutions and fulfil the progressive realisation of socio-economic rights.¹⁸² As a principle of good data governance, taking on a thick description of rule of law is important because it will allow for different manifestations of how regulation materialises, without dictating the parameters on what it should constitute. Whereas a thin description of the rule of law will take into account how institutions can ensure fairness, due process and accountability for the governing of data, a thick description will acknowledge the ways in which the governance of data takes place in practice. Doing so addresses the institutional as well as user-centred ways in which we need to think of governance.

This includes acknowledging the power imbalances between different parties within a data economy, recognising that in order to ensure equitable data governance one needs to incorporate a plurality of ways in which people have access, use and engage with data infrastructures and finally building mechanisms and institutions that can ensure that due process is also contextual, and is flexible to have relevance in different country contexts.¹⁸³

In doing so, when thinking of building a model that considers the global effects of its use, it becomes important to think beyond narrow institutional frameworks that are typical to a particular jurisdiction. It should rather examine the implications that such a model can have on a procedural, substantive and formal basis. With the application of the rule of law as a principle in this instance, this would mean: on a procedural basis, it provides, rules and standards through which data can be governed such as ensuring that there is a lack of arbitrariness and constraints on the use of discretion in terms of governing conflicts within a data economy; it can have substantive value, wherein it provides normative aspects such as the protection of liberty of people, ensuring they have agency and are protected against discrimination; and finally it can have formal dimensions including aspects of its generality and stability, in terms of the expectation with which the architecture of data governance rules are applied and materialised.¹⁸⁴

¹⁷⁹ Anita Gurumurthy and Nandini Chami, 'Governing the Resource of Data: To What End and for Whom?', Working Paper (IT for change, 2022), <https://itforchange.net/sites/default/files/1741/WP23-Governing-the-Resource-of-Data-AG-NC.pdf>.

¹⁸⁰ Sofia Scasserra, Carolina Martínez Elebi, and Nick Buxton, 'DIGITAL COLONIALISM: Analysis of Europe's Trade Agenda' (Transnational Institute, October 2021), <https://www.tni.org/en/publication/digital-colonialism>.

¹⁸¹ Martin Krygier, 'Four puzzles about the rule of law: why, what, where? And who cares?', *Nomos* 50 (2011): 64–104.

¹⁸² Wolfgang Merkel, 'Measuring the Quality of Rule of Law: Virtues, Perils, Results,' 21–47, accessed April 28, 2022, <https://doi.org/10.1017/CBO9781139175937.004>.

¹⁸³ Christian D'Cunha, 'A State in the Disguise of a Merchant': Tech Leviathans and the Rule of Law,' *European Law Journal* (2021): 1– 23, <https://onlinelibrary.wiley.com/doi/abs/10.1111/eulj.12399>; Brian Z. Tamanaha, 'The Rule of Law and Legal Pluralism in Development,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 1, 2011), <https://papers.ssrn.com/abstract=1886572>.

¹⁸⁴ *Ibid.*

Locating the effects of a model globally, by acknowledging that other regions may have different cultures and institutions, is important to ensure that these local specificities are supported rather than supplanted by data governance models that the EU proposes.

5. Review of EU policy context

5.1. Existing data strategy and regulation

After positioning in the discussion in the 2020 European Data Strategy and summarising the state of existing EU regulation relevant to data governance, we will review each of the key pending legislative files relating to data and AI governance, explaining how they come together into an embryonic data governance model for the EU.

The 2020 European Data Strategy is grounded in finding ways to ensure that society can make better decisions with greater agency over data. There is an emphasis on ensuring that Europe is competitive as a player in the data economy by developing its connectivity capacities, computing power and cybersecurity. At the same time it emphasises that there should remain strict protections and controls to ensure that the legal framework prioritises data protection, fundamental rights, safety and security.¹⁸⁵

In this strategic context, the existing regulatory framework in the EU pertinent to data governance could be described as piecemeal and fragmented because, while the focus on building markets is coherent, the different instruments involved create disjunctures in the way technological harms are conceptualised (i.e. through the lenses of data protection, competition and consumer protection). In turn, this fragments the ways in which claims can be made and redress sought. For instance, a platform that is violating privacy rights, building an unfair monopoly and facilitating unfair trading practices must be separately addressed on all three fronts, effectively by different interest groups and specialised intermediaries to help them make their claims. This makes it hard to conceptualise 'good governance' along the lines set out above because any legal or regulatory approach is, by design, incomplete. It thereby places an impetus on the right-seeker to be able to meander through the regulatory maze.

In this regard, seeing these developments as a way to constitutionalise digital spaces is important. This is because a constitutional law perspective offers not only a way to limit the power of different actors, whether public or private parties, performing public functions, and ensures that they remain accountable to the people. It also provides a language of rights, wherein people are safeguarded and protected in regard to their fundamental freedoms, but also have an opportunity to redress any grievance on account of excesses of public and private power.¹⁸⁶ In doing so, it provides a series of guarantees and values to ensure that there is an equilibrium and balance among different players.¹⁸⁷ Framing these discussions through digital constitutionalism therefore ensures that even as a flurry of new legislation addresses different aspects of practice in the data economy, testing legislation from a perspective in terms of how they create instruments to balance power, provide tools to challenge power, and frameworks to hold those with power accountable is critical.

Against that backdrop, among the key laws already in place in the EU that impact on data governance in different ways are the 2016 General Data Protection Regulation (GDPR), the 2016 Data Protection Law Enforcement Directive, the 2002/2009 Privacy and Electronic Communications Directive (i.e. the ePrivacy Directive), the 2018 Non-Personal Data Regulation, the 2003/2019 Directive on the Re-use of Public Sector Information (i.e. the Open Data Directive), and the 2000 e-

¹⁸⁵ European Commission, 'A European Strategy for Data, COM(2020) 66 Final' (European Commission, February 19, 2020), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>.

¹⁸⁶ Giovanni De Gregorio, The rise of digital constitutionalism in the European Union, *International Journal of Constitutional Law*, Volume 19, Issue 1, January 2021, Pages 41–70, <https://doi.org/10.1093/icon/moab001>

¹⁸⁷ Edoardo Celeste,, Digital Constitutionalism: Mapping the Constitutional Response to Digital Technology's Challenges (July 25, 2018). HIIG Discussion Paper Series No. 2018-02, Available at SSRN: <https://ssrn.com/abstract=3219905> or <http://dx.doi.org/10.2139/ssrn.3219905>

Commerce Directive, among others.¹⁸⁸ While several of these have previously been discussed in this report, here we will briefly summarise their regulatory scope, resulting governance architecture and notable gaps, before turning to discuss how the key pending legislative files aim to build on and strengthen these established rules.

The GDPR and Law Enforcement Directive both regulate the processing of personal data in the EU but in different contexts. The GDPR is a comprehensive regulation with the dual aim of ensuring fundamental rights while also stimulating protected data flows both within the EU and outside the bloc with countries deemed to have an essentially equivalent data protection legal framework in place. The nature of police and judicial activities in criminal matters necessitated a differentiated set of rules for the protection of personal data in law enforcement contexts, leading to the Law Enforcement Directive. National data protection authorities (DPAs) provide oversight and enforcement for both the GDPR and Law Enforcement Directive, as well as the ePrivacy Directive which supplements data protection rules with additional regulations for the confidentiality of information, treatment of data in transit, spam and the use of cookies. From a governance perspective, it is worth noting in particular the existence of the European Data Protection Board, which is an independent body whose purpose is to ensure consistent application of the EU's data protection rules and to promote cooperation among the EU DPAs – a governance design mimicked in the draft AI Act.

In terms of the governance of non-personal data, the Non-Personal Data Regulation and Open Data Directive are both relevant but have different aims. The Non-Personal Data Regulation seeks to remove obstacles to the free movement of non-personal data between different EU countries by:

- Prohibiting Member States from imposing requirements on where data should be localised (with limited exceptions);
- Establishing a cooperation mechanism to ensure that competent authorities can exercise their rights to access data being processed in another Member State; and
- Providing incentives for industry – with support from the European Commission – to develop self-regulatory codes of conduct on switching service providers and data portability.¹⁸⁹

Some have argued that the Non-Personal Data Regulation sets the basis for promoting the free flow of non-personal data as a 'fifth fundamental freedom of the EU', alongside the free movement of people, goods, services and capital.¹⁹⁰ The Open Data Directive, on the other hand, focuses on the economic aspects of reusing public sector information specifically. It encourages Member States to make as much information available for reuse as possible and addresses data held by public sector bodies at national, regional and local levels. This includes data held by ministries, state agencies, municipalities and organisations funded mostly by or under the control of public authorities.¹⁹¹

¹⁸⁸ Here we focus on these six existing legislations because they are most relevant to the pending legislative files. However, it is worth noting that EU regulations for cybersecurity and electronic identification and trust services could also be considered relevant to more expansive treatments of data governance, as could the machinery directive and other instruments which will define liability, as these all come together to shape mechanisms of accountability in relation to data and AI.

¹⁸⁹ European Commission, 'Communication from the Commission to the European Parliament and the Council: Guidance on the Regulation on a Framework for the Free Flow of Non-Personal Data in the European Union' Pub. L. No. COM(2019) 250 final (2019), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:250:FIN>.

¹⁹⁰ Kaspar Kala, 'Free Movement of Data as the 5th Fundamental Freedom of the European Union,' e-Estonia, October 6, 2017, <https://e-estonia.com/free-movement-of-data-as-the-5th-fundamental-freedom-of-the-european-union/>.

¹⁹¹ In terms of governance specifically: 'The means of redress should include the possibility of review by an impartial review body. That body could be an already existing national authority, such as the national competition authority, the supervisory authority established in accordance with Regulation (EU) 2016/679, the national access to documents authority or a national judicial authority. That body should be organised in accordance with the constitutional and legal

The e-Commerce Directive, a two-decades old regulation aimed at establishing an internal market for 'information society services' and limiting liability for service providers in certain circumstances, is of tangential interest to the current data governance discussion largely because the Digital Services Act seeks to strengthen its key regulatory principles by adding obligations to address notifications of content deemed to be illegal.

Among the existing data regulations outlined above, the enforcement of the GDPR has been a regular target of criticism. In addition to concerns about insufficient resourcing among enforcement authorities and claims that the 'one-stop-shop' enforcement system works against protecting individuals, data protection enforcement in the EU is also argued to suffer from a lack of transparency and cooperation between authorities, diverging legal interpretations, cultural conflicts, inconsistencies in prioritising which cases to pursue and, in some cases, a perceived reluctance among the authorities of certain Member States to enforce the law.¹⁹² Importantly in the context of future AI and data governance in the EU, GDPR enforcement is also argued to lack a collective dimension which is an increasingly essential component to the 'architecture of enforcement'.¹⁹³

This is the context in which the EU's policymaking about AI is situated. It combines underlying conceptual layers with regard to fundamental rights and economic cooperation, with strategy designed to inform the legislative files discussed in the following section. Taking a justice perspective, our analysis explores the extent to which these interests productively interact, where governance becomes skewed toward economic priorities at the expense of people, and how the plural interests of the EU's diverse populations can be recognised and represented through the process of governing technologies.

Before moving to introduce the legislative files, it is worth mentioning again that in 2022 the European Commission introduced a declaration on digital rights. Arguing that such digital rights are needed to ensure that people are placed at the centre of a digital transformation, and that their freedoms, security, safety and participation remain assured. The declaration emphasises the ways in which people need access to – but also protection from – digital technologies. It includes a wide variety of protections including the ability to access high speed internet and technologically-equipped education resources, while also having access to a safe digital environment for children, and information about the impacts of technology on the environment and health.¹⁹⁴

It is also worth considering the main contexts in which the EU makes policy about data, beyond the high-profile area of shaping digital markets. The places where policy about data becomes visible through legislative debates are areas where data has serious implications for human rights and the European economy, and often where states play a stronger role than the European Commission in determining the direction of policy. When thinking about data governance, we should also consider

systems of Member States. Recourse to that body should not pre-empt any means of redress otherwise available to applicants for re-use. It should however be distinct from the Member State mechanism laying down the criteria for charging above marginal costs. The means of redress should include the possibility of review of negative decisions but also of decisions which, although permitting re-use, could still affect applicants on other grounds, in particular by the charging rules applied. The review process should be swift, in accordance with the needs of a rapidly changing market.' European Parliament and European Council, 'Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information,' PE/28/2019/REV/1 § (2019), <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>.

¹⁹² Jukka Ruuhonen and Kalle Hjerpe, 'The GDPR enforcement fines at glance,' *Information Systems*, 106, 2022, <https://doi.org/10.1016/j.is.2021.101876>.

¹⁹³ René L. P. Mahieu and Jef Ausloos, 'Harnessing the collective potential of GDPR access rights: towards an ecology of transparency,' *Internet Policy Review*, July 6, 2020, <https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487>.

¹⁹⁴ European Commission, 'Commission Puts Forward Declaration on Digital Rights and Principles for Everyone in the EU,' Text, European Commission, January 26, 2022, https://ec.europa.eu/commission/presscorner/detail/en/IP_22_452.

these areas of facilitatory data policy as points where values are articulated both in relation to data, and through the uses of data allowed and disallowed. These include security and intelligence; cybersecurity; migration and border control; market competition (beyond the digital); and SMEs and innovation policy.

These areas are core foreign and economic policy domains for both the European Commission and Member States, and perhaps due to their diplomatic importance are also where EU values are cited least often. Thus the notion of data governance working 'for citizens' in these areas is not necessarily the central consideration, and benefits to the EU's population may be conceptualised indirectly. All this suggests that attention to EU values is important to (re)articulate when thinking about data governance, because it is frequently taking place in relation to core policy areas where power and economic priorities are the central considerations.

5.2. Relevant legislative files and topical debates

5.2.1. AI Act

The European Commission's proposed Artificial intelligence Act 2021 is part of the EU's strategy to introduce a harmonised risk-based approach to determine which AI systems should be regulated and how. The proposed Act takes an ex-ante approach to regulating AI, where actors and their decisions at the development and initial deployment stage are the main focus, and a risk calculation is required at this point, by these actors. The proposal recognises three categories of risks (i) unacceptable risks, (ii) high risks, (iii) limited risks and (iv) minimal risks. The Act applies to AI systems that create high risks to the health, safety and fundamental rights of individuals. Article 10 of the AI Act provides for the governance of training, validation and testing data sets, using the appropriate data governance and management practises to ensure that the data in high-risk AI systems satisfy the quality criteria laid down in paragraphs 2 to 5 of the Article.

The main debates emerging around the AI Act are set out below with an accompanying outline of the issues involved, relating them to the principles set out in section 4 above:

5.2.1.1. Capturing the diverse lifecycle of AI systems

This first problem turns on the argument that AI is not a product that is developed for a single purpose and has a single use with a single user.¹⁹⁵ Instead it is more usefully seen as a set of interacting systems with different lifecycles, which may be used by different actors for different purposes. For example, a developer may create a facial recognition system, but different institutions and companies may then license it for purposes as diverse as identifying war victims,¹⁹⁶ policing cities,¹⁹⁷ or proctoring examinations in high schools.¹⁹⁸ This presents a problem both for conceptualising risk (see below) and for enforcing standards and finding the right triggers for oversight and regulation. It also runs the risk of stimulating regulatory entrepreneurship (an

¹⁹⁵ This section draws directly on the arguments of Lilian Edwards: Lilian Edwards, 'Regulating AI in Europe: Four Problems and Four Solutions' (Ada Lovelace Institute, March 2022), <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf>.

¹⁹⁶ Euronews and Reuters, 'Ukraine Begins Using Facial Recognition to Identify Russians and Dead,' *Euronews*, March 14, 2022, sec. next_biztech-news, <https://www.euronews.com/next/2022/03/14/ukraine-begins-using-controversial-clearview-ai-s-facial-recognition-to-id-russians-and-th>.

¹⁹⁷ Tate Ryan-Mosley, 'The NYPD Used a Controversial Facial Recognition Tool. Here's What You Need to Know,' MIT Technology Review, September 4, 2021, <https://www.technologyreview.com/2021/04/09/1022240/clearview-ai-nypd-emails/>.

¹⁹⁸ Nakeema Stefflbauer, 'How EU Students Are Being Forced into a Surveillance Nightmare,' *www.euractiv.com*, March 4, 2021, <https://www.euractiv.com/section/digital/opinion/how-eu-students-are-being-forced-into-a-surveillance-nightmare/>.

increasingly common tactic amongst platforms, where those deploying technology in new ways wait for public protest and regulatory response rather than assessing potential harms beforehand, hoping that they can then push for regulatory space based on its already being on the market).¹⁹⁹

This problem relates to our benchmark for good governance around contestability: if AI can only be contested at the point where it is developed or purchased, this creates a problem with society and regulation's ability to respond to unforeseen and emergent negative effects of the technology.

5.2.1.2. Defining fake AI

The Act does not distinguish between systems that are scientifically evidence-based and systems that constitute pseudoscience. For instance, the Act contains definitions for systems 'used to detect emotions', while this is considered by experts in the social and computing sciences to be impossible to do reliably, and at its origin is based on long-disproven (un)scientific theories.²⁰⁰ This kind of system has been defined as 'fake AI' and 'snake oil' by experts in the EU and US. By offering a way to calculate the risk of such systems, the Act (if passed) would support their validity and effectively normalise their use.

This approach creates an accountability paradox: any contestation addresses the operation of such systems, effectively lending them a level of pseudo-credibility. It is not the accuracy or transparency of such systems that is in doubt, but the foundational principles on which they are based, which are not in line with scientific evidence.

5.2.1.3. Defining the role of the public in shaping and contesting AI systems

The AI Act in its draft form addresses AI within the paradigm of product liability and does not include mechanisms for complaint or enforcement if a system causes harm to individuals, nor does it conceptualise how harm may come to light, other than through the deployer's reporting responsibilities. Neither individuals nor interest groups (such as consumer associations or rights groups) are given the right to raise complaints or to bring cases for harms caused by AI systems. Furthermore, there is no provision for individuals or groups to challenge regulators' decisions on AI systems. All of these mean that there are few pathways for harms to come to light and for systems to be changed based on the problems identified post-deployment.

Again, this is clearly a problem with the missing conceptualisation in the draft Act of public contestation in relation to AI. For certain groups subject to AI systems, there is also no way in which they can legitimately make a claim - migrants and refugees are one such category, because if a system used in assessing their claims is wrongly applied or should not be applied at all, they are deported and thus lose the ability to make claims.

5.2.1.4. Capturing risk with regard to AI

The draft Act is not able to give a clear definition of risk given that this is dependent on the use of a system or model, and that such use may change across its lifecycle. Sectoral authorities have called the draft Act's way of defining high versus medium or low risk AI 'arbitrary'.²⁰¹ Connected to this, defining risk requires defining the population that is placed at risk by a given system, and the draft

¹⁹⁹ Elizabeth Pollman and Jordan M. Barry, 'Regulatory Entrepreneurship,' *Southern California Law Review* 90, no. 3 (2017 2016): 383–448.

²⁰⁰ See Douglas Heaven, 'Why Faces Don't Always Tell the Truth about Feelings,' *Nature* 578, no. 7796 (February 26, 2020): 502–4, <https://doi.org/10.1038/d41586-020-00507-5>. See also Luke Stark and Jevan Hutson, 'Physiognomic Artificial Intelligence,' SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, September 20, 2021), <https://doi.org/10.2139/ssrn.3927300>.

²⁰¹ For example, in the health sector: Janneke van Oirscho and Gaby Ooms, 'Interpreting the EU Artificial Intelligence Act for the Health Sector' (Amsterdam: Health Action International, February 2022), <https://haiweb.org/publication/interpreting-the-eu-artificial-intelligence-act-for-the-health-sector/>.

Act noticeably does not offer a clear definition of vulnerability. Instead, it draws on a mixture of high-risk situations, innate characteristics or attributes, and created vulnerabilities which come with a particular use of a general-purpose form of AI. For instance, it cites a range of innate vulnerabilities such as age or physical or mental disability; but it also conceptualises vulnerability as created through the intent of a designer or deployer of an AI system, for instance through the use of subliminal techniques, or as created through dependence on authorities in a given situation, for example when claiming asylum.

This does not align with other ways of defining vulnerability in relation to technology, for instance in the GDPR which defines 'sensitive' data categories and groups who are inherently vulnerable, such as children. This risks asking regulators to think in very different ways about problems which often overlap, i.e. those of automation and those of data protection. It does become problematic, however, in relation to the 'defining fake AI' problem outlined above. For instance, the Act defines migrants and refugees as needing particular protection given a combination of vulnerability and dependence on authorities for recognition of their fundamental rights, but in particular relation to 'polygraphs and similar tools or to detect the emotional state of a natural person' (recital 39).

Such technologies pose very different problems if we consider them as fake, compared to if we consider them as producing potentially accurate results that should then be used with caution because the subjects are in a high-risk situation. It also does not conceptualise cumulative harms,²⁰² neglecting a long-recognised way in which many problems with technology emerge.

This problem relates to inclusiveness in the vision of regulating AI, where it is necessary to think about the ways in which such technologies will impact on different groups differently, and contestation and accountability, where those who are not recognised - collectives, for example, or categories not recognised in traditional discrimination legislation, cannot effectively make claims.

5.2.1.5. Meaningfully protecting fundamental rights

Two main criticisms have arisen in relation to the draft Act's stated aim of protecting fundamental rights.²⁰³ First, that it does not incorporate any notion of those actually affected by the technology (either end-users, or those on whom the technology is used), and therefore does not align with the way fundamental rights are either conceptualised or used to make claims. Second, that it does not include provisions for any ex-ante rights impact assessment, so that it does not have a requirement for developers or deployers of AI systems to consider fundamental rights as a real constraint. Furthermore, assessing fundamental rights impacts requires conceptualising the mechanisms through which those impacts may materialise, so that the task of both regulators and AI deployers becomes to understand which rights are affected, and how.

This problem relates to our benchmarks of contestability, accountability, and also global responsibility: not making it possible to ground claims in the framework of fundamental rights is problematic because it treats people as consumers addressing a product, rather than the public faced with systems which impact on almost every area of fundamental rights. It is also a threat to the EU's claim that it will regulate AI in line with fundamental rights, and to the bloc's responsibilities on human rights as an international actor.

5.2.1.6. Protection beyond a fundamental rights framing

Beyond adequately addressing fundamental rights, the draft AI Act also fails to conceptualise how to address threats to the common interest and on the collective level caused by AI. Examples of

²⁰² Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach,' *Computer Law Review International* 22, no. 4 (August 1, 2021): 97–112, <https://doi.org/10.9785/crl-2021-220402>.

²⁰³ Edwards, 'Regulating AI in Europe: Four Problems and Four Solutions.:'; Ibid.

these might include the environmental damage caused by large computing infrastructures,²⁰⁴ impacts on groups and collectives caused by algorithmic groupings²⁰⁵ which are not covered by the individual focus of the fundamental rights approach, and emergent impacts that go beyond the protected characteristics defined in discrimination law (for example, if a system is used to cause discriminatory effects based on proxies created through data mining, for example a particular consumer profile that is common to people with low incomes or a particular ethnic heritage).²⁰⁶

This problem relates to our criteria for good governance on the dimensions of contestability, accountability, and global responsibility. Claims in relation to AI harms do not fit easily with established structures of accountability. They will frequently be transnational in nature, they may relate to opaque processes of automated decision-making, and the subject of claims may involve intersectional and complex forms of discrimination, the creation of new groupings, or to a whole set of civil and political rights (including environmental rights) that have not yet been brought into focus at scale in relation to AI.²⁰⁷ All this requires a shift in the approach to oversight: emergent harms may escape recognition unless oversight includes both sectoral and local knowledge and expertise, and unless there are mechanisms and resources for making and channelling claims at the local as well as national level.

5.2.2. Data Governance Act

The Data Governance Act provides the basis for the reuse of certain types of data held by public sector bodies such as commercially confidential data, data subject to statistical confidentiality, data subject to intellectual property rights and trade secrets and personal data in public databases.²⁰⁸ It however does not oblige public sector bodies to allow for the reuse of data.

The purpose behind this legislation is to create a competitive landscape to allow for data sharing, while at the same time ensuring trust between different players in the data economy.

Given the nature of the data involved, the draft legislation proposes to ensure that reuse can only take place once it considers the implications for the principle of non-discrimination, data protection and fundamental rights of users.

The draft legislation also lays down the framework for data intermediation services which entails, 'a service, which aims to establish commercial relationships for the purpose of data sharing between an undetermined number of data subjects and data holders, on the one hand, and data users on the other hand, through technical, legal or other means, including for the exercise of data subjects' rights in relation to personal data.'²⁰⁹

The draft legislation provides for a concept of data altruism, which is the voluntary sharing of data for a public purpose, and offers a framework for the voluntary registration of entities that will provide such services. It also ensures that there are restrictions in place for the transfer of non-

²⁰⁴ University of Washington. 'Large computer language models carry environmental, social risks.' ScienceDaily. www.sciencedaily.com/releases/2021/03/210310150408.htm (accessed April 28, 2022).

²⁰⁵ Taylor, Floridi, and van der Sloot, *Group Privacy*.

²⁰⁶ Edwards, 'Regulating AI in Europe: Four Problems and Four Solutions. '; Ibid.

²⁰⁷ Nancy Fraser (2008). 'Abnormal justice.' *Critical Inquiry*, 34(3), 39. <https://doi.org/10.1086/589478>

²⁰⁸ European Parliament and Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' Pub. L. No. COM(2021) 206 final, 2021/0106 (2021), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

²⁰⁹ Ibid.

personal data to third countries, such that there are equivalent protections for intellectual property as well as trade secrets, and that there is effective judicial redress.

There is also a provision for a right to seek effective judicial remedy – any individual and/or legal entity is given the right to lodge a complaint with the competent authority against the intermediation service providers or an entity registered in the register for data altruism organisations. Even if an individual or legal entity is not convinced or aggrieved with the decision/order of the competent authority, they can take the legal recourse and exercise their right to effective judicial remedy to review the decision by an impartial expert body.

Finally, the draft legislation sets up a data innovation board for the purposes of implementing, advising and training public sector bodies. The board will also be responsible for creating best practices across different sectors and ensuring standardisation, for instance, on matters of security and access procedures. In our discussion so far on thinking about data governance from a social justice standpoint, there are few considerations that have emerged that require further discussion.

5.2.2.1. Implementing data altruism

Data altruism is the idea that people will on their own accord provide information that can benefit the public good. This idea speaks to how people can share their reactions to different institutions and systems whether in relation to health or finance which can contribute to understanding of how these systems may impact different people differently. However, as Veil has argued the legislation imposes strict requirements on donors and altruistic organisations under the GDPR. These onerous requirements, rather than facilitating registration and participation, create several legal, technical and procedural hurdles which may stifle innovative activity.²¹⁰

5.2.2.2. Risks of for-profit data intermediaries

Regarding data intermediation services, one of the critiques of its construction is that rather than challenging the concentration of market power among a few actors, it is providing an alternative way in which business can be done through creating a platformised entity. Vogelesang argues²¹¹ that the intermediation framework requires having a central intermediary who will connect data subjects and holders on one hand and data re-users on the other, and that this may result in a centralisation of power. In this regard, several commentators argue for how the Data Governance Act should not be a missed opportunity to develop the idea of a commons, where intermediaries should work in non-commercial ways and in the public interest.²¹²

5.2.3. Data Act

The draft Data Act aims to promote a fairer digital economy by expanding on the data subject right to access the data that people produce (i.e. data portability). The idea behind the proposal is to regulate data access and use by creating a governance framework to generate incentives for data sharing and pooling. The objectives of the Data Act are:

- facilitating the access and use of data by consumers and businesses while preserving the incentives to generate value through data;

²¹⁰ Dataprotection LANDSCAPE, 'Data Altruism,' Dataprotection LANDSCAPE, accessed April 28, 2022, <https://dataprotection-landscape.com/law/data-altruism/>; Winfried Veil, 'Data Altruism: How the EU Is Screwing up a Good Idea,' *AlgorithmWatch* (blog), January 27, 2022, <https://algorithmwatch.org/en/eu-and-data-donations/>.

²¹¹ Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) Analysis of the Final Compromise Text in View to Agreement' (Council of the European Union, October 12, 2021), https://openfuture.eu/wp-content/uploads/2021/11/211130_DGA_Triologue_Provisional_Agreement.pdf.

²¹² Francesco Vogelesang, 'A Closer Look at Data Intermediaries and the Risk of Platformization,' Open Future, March 1, 2022, <https://openfuture.eu/blog/a-closer-look-at-data-intermediaries-and-the-risk-of-platformization/>.

- creating a framework to allow public sector bodies to access the data held by companies in *exceptional contexts*;
- promoting access to data processing services for all actors in the market;
- creating safeguards against unlawful data transfer without notification by cloud service providers for the broader digital sovereignty strategy of the European Union; and
- providing the legal basis for developing interoperability standards for data use and sharing.²¹³

The draft Data Act proposes the consolidation of access rights for users of products and related services to the data generated by these devices and services that are owned or leased (Article 4). This proposition translates into a requirement for manufacturers of those products and services to design them in such a way that the data is easily accessible and with transparency over the use and access of the data produced. This proposal includes an exemption from this obligation for trade secrets that says that the user cannot use the data obtained to develop a product that competes with the product that originated that data.

The proposal expands the right to share data with third parties excluding the gatekeepers designated by the Digital Markets Act. In this case, the data holders can request compensation for the data that is reasonable and non-discriminatory. The regulation also directly prohibits using data for profiling natural persons and making data available to another third party unless it is strictly necessary for providing the service. Likewise, the proposal excludes micro and small enterprises (less than 50 employees) from the obligations.

Another proposition of the draft Data Act is the framework for business-to-government data sharing. The Act limits this proposal by leaving it as a request designed for emergency situations,²¹⁴ rather than constructing a systematic approach to use data for the public benefit. Likewise, the public entity should demonstrate the need for the data to respond to a public emergency and that there are no other means to access that data.

The Data Act also promotes the creation of interoperability provisions in the European data governance framework. The measures point to establishing a competitive cloud market to promote European digital sovereignty, avoiding vendor lock-in scenarios and dependency on foreign services. These interoperability regulations will apply to personal and non-personal data. Likewise, the regulation introduces the concept of *common data spaces*, but the concept is not properly clarified in the draft Data Act.

The Act is an important step towards designing services that could help to reduce the power imbalances that are produced by dominant data controllers. However, according to the literature, the reach of the proposal could be limited considering the following issues:

- The exception for trade secrets is potentially an open door for abuse especially considering the unequal power relationships in the current digital economy between consumers and big commercial companies.²¹⁵ Furthermore, the exclusion of micro and small enterprises from the regulation also limits the possibility of creating a more balanced data economy.²¹⁶

²¹³ Council of the European Union, European Parliament, and European Commission, 'Harmonised Rules on Fair Access to and Use of Data (Data Act),' Pub. L. No. COM(2022) 68, 2022/0047 (2022), 38

Thomas Margoni and Charlotte Ducuing, 'Data Act Blog Post Series: Introduction,' CITIP blog, April 21, 2022, <https://www.law.kuleuven.be/citip/blog/data-act-blog-post-series-introduction/>.

²¹⁴ Council of the European Union, European Parliament, and European Commission, 'Harmonised Rules on Fair Access to and Use of Data (Data Act),' Pub. L. No. COM(2022) 68, 2022/0047 (2022).

²¹⁵ Inge Graef and Martin Husovec, 'Seven Things to Improve in the Data Act,' *SSRN Electronic Journal*, 2022, <https://doi.org/10.2139/ssrn.4051793>.

²¹⁶ Paul Keller and Francesco Vogelesang, 'Data Act-Access to Data' (Open Future, February 23, 2022).

- The limitation on using data shared under this regulation to 'develop competing products' restricts the possibility to develop new services. It could be used by big economic players to stifle competition in the digital market.²¹⁷
- The proposal for creating a framework for sharing data between business and public entities is limited by the idea of needing special situations and demonstrating that there are no other available means to access data. The idea should be to create a systemic approach that allows public bodies to identify and access datasets for the public interest. The proposal for an *exceptional* situation could be replaced by requirements for using data for the public interest.²¹⁸
- There is a need in the regulation for the establishment of a public body that oversees and supports the flow of data between business and public entities.²¹⁹

²¹⁷ Francesco Vogezang and Alek Tarkowski, 'Data Act-Business to Government Data Sharing' (Open Future, February 23, 2022); Inge Graef and Martin Husovec, 'Seven Things to Improve in the Data Act,' *SSRN Electronic Journal*, 2022, <https://doi.org/10.2139/ssrn.4051793>.

²¹⁸ Alek Tarkowski and Francesco Vogezang, 'Data Act-Interoperability and Data Sharing Services' (Open Future, February 23, 2022).

²¹⁹ Inge Graef and Martin Husovec, 'Seven Things to Improve in the Data Act,' *SSRN Electronic Journal*, 2022, <https://doi.org/10.2139/ssrn.4051793>.

6. Policy options and alternatives: a data justice analysis

In this section we will analyse policy options first in relation to the European Strategy for Data, and then in relation to the principal legislative files currently under development.

We include, where relevant, lessons from other regions and systems by looking at some models that could inspire possible futures for European data governance, including both large-scale and smaller instruments which provide the opportunity to influence the larger scale (e.g. public trusts). We also address the issue of sustainability - both in the sense of the environmental sustainability of computing systems involved in our current and future data economy, and in the sense of the social sustainability created by good governance architectures and effective oversight.

6.1. European Strategy for Data

The strategy articulates four central principles: 1) a cross-sectoral governance framework for data access and use; 2) investments in data, capabilities, infrastructures, and interoperability; 3) building competences and skills; and 4) establishing common European data spaces. It also outlines a proactive international approach to making data available for European businesses.

Overall, the legislative files scrutinised in this report show that there is work to be done to build out the normative components of the strategy. Currently, it relies on the assumption that if data flows more freely, social good will follow. The strategy does not articulate a vision of what is 'better decision-making' beyond describing greater efficiency in public services and business processes, and the opportunity to personalise services, particularly in healthcare. All these add up to an argument that data can increase efficiency in processes and control on the part of organisations and authorities - but are not accompanied by a corresponding articulation of the mechanisms that could make this increase in efficiency and control beneficial for society.

One way to increase the likelihood of socially beneficial effects from the data economy is to complement the proposed cross-sectoral governance framework with sectoral mechanisms for identifying emergent problems and opportunities on the civil society level. Sectoral approaches are important for recognising needs and realising mechanisms of redress: unless domain organisations such as unions, collectives and civil society organisations are positioned as both sources of good use cases and contributors to oversight, the framework will fail to create beneficial forms of efficiency and decision-making power.

A second complement to the strategy is to dedicate increased deliberation and resources to creating public digital infrastructures for communities and civil society groups as well as under-served public sector organisations such as schools and hospitals, so that they have infrastructural options beyond those of the biggest commercial technology providers. This would maximise the chance that data's social value can flow outward toward communities and interest groups as well as being available to those with the greatest resources and data management capacity, i.e. market actors. Unless such public infrastructures are structurally resourced and fostered, data will continue to serve mainly as a form of financial capital, and the vision of interoperability will not fulfil its potential. This counterweight to the market and infrastructural power of big tech is essential to fostering social value in relation to data, but it is only beneficial in a plural approach where the recognition and work of small social actors is facilitated on the same basis as that of SMEs. The logic that competition amongst cloud services, increased availability of data to AI development, and decentralised ways of managing data will per se lead to the creation of social value and public goods is convenient rather than true.

6.2. AI Act

6.2.1. AI as a public technology

The notion of AI as a product, upon which the draft AI Act is based, positions the public as consumers of products rather than (as in the GDPR, for example) affected by systems in ways which require control and oversight. In order to adequately conceptualise AI as a public, as well as a private, set of technologies, it is necessary to provide checks and balances at the design and procurement stage in order to assess whether a given system is in line with the public interest and how the public can influence this discussion. One example of this kind of influence is the stand taken by Dutch university leadership on public digital infrastructures;²²⁰ another is India's 'Janta Parliament' discussions where a coalition of civil society organisations has come together to organise civil society deliberation on issues such as enforcing civil liberties and shaping digital rights.²²¹ These claims on the part of public institutions and publics affected by particular technologies have the common feature that they link concerns about public institutions and services directly to civil and political rights, and that they bring cases of emergent effects and harms from technology that are not yet articulated in law or within the portfolio of regulators.

Second, mechanisms are needed to provide ways for individuals affected by AI meaningful routes for contesting its deployment where necessary. This is particularly important given that most people affected negatively by AI are not aware that it is AI causing the harm in question, because it is used within an intermediary such as a commercial firm (e.g. a bank) or the police. The lack of direct and visible points of connection between algorithmic technologies and the public necessitate new architectures of accountability beyond those provided by data protection or consumer protection law.

Third, mechanisms are needed to make it possible to identify and tackle problems that emerge from the deployment of an AI model or system during its entire lifecycle, including when it is reconceptualised or repurposed along the way. In order for AI systems to evolve through different use cases without harmful effects, the Act could include mechanisms for both individual and collective complaints to be brought directly against deployers of AI, and could operate in relation to public and administrative law wherever appropriate, rather than within the framing of product liability and consumer rights.

6.2.2. Accountability

One key issue for accountability in relation to AI is assessing impacts in an ex-ante and systematic way. This is a problem of structuring legal frameworks to provide mechanisms for contestation, as noted above; expanding the landscape of rights considered relevant to AI in order to take account of emergent and use-related harms that are sector-specific; and providing oversight and enforcement appropriate to the scale of the challenge.

²²⁰ See Maex, note 161, section 4.

²²¹ For the session on digital rights, see: Janta Parliament, 'Janta Parliament Session on Technology and Surveillance,' Janta Parliament, August 18, 2020, <https://jantaparliament.wordpress.com/2020/08/23/press-release-technology-and-surveillance-session/>.

One way to institute meaningful ex-ante checks is through Human Rights Impact Assessments. These have also been proposed as a tool for the oversight of algorithms in the US²²² and in Canada,²²³ where, for example, scholars have proposed HRIA's for algorithms related to labour markets that look at the multiple related rights of: work, equality and nondiscrimination, privacy, free expression and free association. Prohibitions based on human rights could also be explored in the area of physiognomic AI, where in the US scholars make a powerful argument for a legal prohibition²²⁴ based on both moral and functional considerations.

AI regulation could include the notion of emergent harms by including post-deployment monitoring and (re)assessment in order to take account of the divergent paths AI systems and models often take after their initial conception, and how they pose different problems to rights, regulation and oversight depending on the use. The notion of collective risk and harm could also be incorporated into EU law in a way that allows for claims to be made about rights violations that occur through algorithmic groupings. This in turn may require the establishment of auxiliary oversight on the sectoral or local (provincial, municipal) level to triage claims, identify emergent harms, and channel claims appropriately to regulatory bodies, both sectoral or digital.

These proposed mechanisms go beyond classic formulations of rights in relation to digital systems. Data and AI governance could instead assume a) that AI systems' human rights impacts will be emergent as much as they are predictable, because they will depend substantially on the decisions made during deployment, and b) that the EU establish this as a question of global responsibility. If the EU's legislative package does not conceptualise civil society as a resource for governing and shaping AI, this may establish an undesirably low standard for AI regulation in other countries. The adoption of such a standard will in turn have effects on the EU's ability to protect Europeans from AI systems developed and deployed outside the EU.

In terms of the global responsibility, the EU could take the lead in defining emotion-recognition AI as pseudoscience and prohibiting it in all areas, rather than normalising it through a calculation of risk. This would require red lines, rather than a judgement of risk. This is particularly important in relation to AI that is used on the public, and even more so in cases where there is no meaningful possibility of complaint or redress, for instance with AI used in relation to investigating the claims of migrants and refugees, for policing marginalised communities or in fraud detection. This poses a particular problem for the product safety paradigm adopted by the framers of the draft AI Act, because it avoids the issue of whether a system does what it is claimed to do in the first place, instead asking whether it is safe when deployed.

6.2.3. Defining vulnerability

The Act could aim for a clear set of categories with regard to vulnerability in order to avoid limiting the definition by reference to a too-narrow set of possibilities. One category relates to those who are predictably vulnerable to the effects of AI systems, for example children, the elderly and those who have historically suffered discrimination based on attributes such as gender, ethnicity or disability. A second category relates to foreseeable but not attribute-based vulnerability created by the conceptualisation or deployment of a particular system, for example the public at large with

²²² Emanuel Moss et al., 'Assembling Accountability: Algorithmic Impact Assessment for the Public Interest' (Data & Society, June 29, 2021), <https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/>.

²²³ Josephine Yam and Joshua August Skorburg, 'From Human Resources to Human Rights: Impact Assessments for Hiring Algorithms,' *Ethics and Information Technology* 23, no. 4 (December 1, 2021): 611–23, <https://doi.org/10.1007/s10676-021-09599-7>.

²²⁴ Luke Stark & Jevan Hutson (2021). 'Physiognomic Artificial Intelligence'. Available at SSRN 3927300. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3927300

regard to facial recognition technologies used in law enforcement, or low-income groups or minorities with regard to policing and anti-fraud technologies. A third category relates to emergent vulnerabilities in relation to algorithmic sorting and categorisation, where people with any attributes or background may be disadvantaged by a grouping based on data mining, for example through data-driven credit rating.

6.2.4. Contestability

Contestability and accountability cannot easily be answered by existing mechanisms and oversight. In terms of contestability, the public can only make claims based on fundamental rights if they are clearly laid out as benchmarks for acceptable AI systems, and if an explicit mechanism is provided in the Act for making claims based on them. For accountability more broadly, the Act could avoid allocating the bulk of oversight responsibilities to Data Protection Authorities. This is manifestly insufficient for AI oversight because these authorities are already overstretched, and as importantly, because AI presents different problems from data protection in terms of scale and the types of rights that come into play which go far beyond privacy, data protection and discrimination.

There is also a need to conceptualise accountability with regard to those who cannot make claims, such as migrants and refugees who may be refused safety due to the misuse of AI. This problem is broader than the AI Act, and with a longer history, but deserves special attention when thinking about fundamental rights and AI because it is emblematic of the opacity of AI to the public in general. If this problem can be addressed for harder cases such as harms to migrants and refugees, it will go a long way to conceptualising how to address it in general. This requires standard-setting in terms of establishing internationally agreed procedures for fundamental rights impact assessment in relation to AI systems – something the EU is positioned to do.

In terms of global responsibility, as one of the first international actors to establish AI regulation, the EU has an interest in establishing the connection between that regulation and fundamental rights. Standard-setting also plays a role here: if AI technologies built outside the EU can be sold to EU deployers without explicit assessment of their impact on fundamental rights, and if, conversely, EU technologies can be sold for deployment elsewhere in the world without any guidance from the EU on potential rights violations that might result, the EU is not only missing a chance in terms of AI diplomacy, but is failing to behave as a responsible actor on the global stage.

6.2.5. Beyond the fundamental rights framing

In order to provide genuine contestability for AI systems in practice, it is necessary for the Act to address the ways in which fundamental rights are currently under-interpreted and under-realised. Without attention to collective harms, to civil and political harms, and to larger and longer-term harms such as the environmental burden of large-scale computational infrastructures and processes, it is hard to claim that rights are at the centre of EU policy on technology. This relates, as noted above, to the question of global responsibility. If the EU is to lead on the responsible use of AI, a necessary first step is to conceptualise the different dimensions along which responsible use could be defined.

6.3. Data Governance Act

6.3.1. Data altruism and digital public goods

In order for data altruism to work, there need to be ways to make the capacity for individuals and groups to participate less taxing and complex. Through introducing the idea of altruism, the legislation suggests ways in which people can contribute towards creating data for the public good. However, coupled with this is strengthening the notion of digital public goods. This includes introducing a comprehensive definition of digital public goods for the EU. It also includes increasing

institutional capacity to generate digital infrastructure that is not subject to private and proprietary concerns.²²⁵ It involves ensuring that standards for digital infrastructure are grounded in the principle of doing no harm²²⁶ and that there is a human rights impact assessment for digital infrastructure to ensure that it contributes to equitable development and access.²²⁷

6.3.2. Meaningful data collectivisation

As discussed above, the current articulation of the notion of data intermediaries may result in the unintended monetisation of data and the centralisation of power. For true inclusiveness, it is imperative to build more participatory decision making into data governance frameworks and enable consultation to understand people and their interests.²²⁸ This includes adopting a decentralised approach, which gives people collective autonomy around their interests and identity as well as strengthening the capacity of communities and groups to be able to build platforms/infrastructures which centre their interests and promote a fair digital market. In order to ensure that the re-use of data contributes towards the public good, it is also important to ascertain the implications of data relationality for people who are members of groups and communities and create mechanisms for the redress of collective harms.²²⁹

6.3.3. Democratising data and civil society agency

Democratising data may require challenging the definition of data as a commodity by promoting the notion of data as a commons. This would mean to develop the notion of data cooperatives such that there is an equitable distribution of gains in the data economy as well as decentralised governance instruments that ensure effective realisation of rights claims and resolutions of disputes within platforms. While the Data Governance Act develops the notion of European Data spaces, 'meaning an internal market for data in which data could be used irrespective of its physical storage location in the Union in compliance with applicable law, which, inter alia, could be pivotal for the rapid development of artificial intelligence technologies',²³⁰ civil society agency may need to be built in order to ensure that these spaces are representative. In addition to an adherence to FAIR principles, mechanisms could be introduced for civil society to make claims and articulate concerns. Civil society is insufficiently connected to both debates about data governance and processes and

²²⁵ The Secretary General's report on digital public goods connects the existence of digital public goods to the implementation of the sustainable development goals, arguing that creating such an infrastructure is critical to ensure that there is more equitable development and access, particularly as existing technical infrastructures are controlled under private and proprietary regimes. In addition to developing a working definition, the United Nations also recommends that there should be coordination in terms of developing a platform to share digital public goods and pool resources, and an understanding of the human rights implications that having such goods can have. United Nations Office of the Secretary-General's Envoy on Technology, 'Digital Cooperation: 'Digital Public Goods' Implementation Plan (SEP. 2020 – DEC. 2022, TOWARDS 2030),' April 15, 2021, https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/210415_Digital_Public_Goods_roundtable_workplan.pdf.

²²⁶ This includes ensuring that there are 'open-source software, open data, open artificial intelligence models, open standards and open content that adhere to privacy and other applicable international and domestic laws, standards and best practices and do no harm'. United Nations Secretary-General, 'Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation' (United Nations General Assembly, May 29, 2020), <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement>.

²²⁷ The United Nations Secretary General in a report from May 2020, articulated a roadmap for digital cooperation that aimed to optimise how digital technologies are used while reducing risk. The principles that are critical in this roadmap include the need to build human and institutional capacity, which protects human rights, and agency, fosters trust, security and stability, and enhances digital cooperation, such that there is an inclusive digital economy. Ibid.

²²⁸ See generally: https://nonalignedtech.net/index.php?title=Main_Page

²²⁹ Linnet Taylor, Luciano Floridi, and Bart van der Sloot, *Group Privacy*. Salome Viljoen, 'A Relational Theory of Data Governance'.

²³⁰ Recital 2, Data Governance Act. ([COM\(2020\)0767](https://eur-lex.europa.eu/eli/reg/2020/843/oj) – C9-0377/2020 – [2020/0340\(COD\)](https://eur-lex.europa.eu/eli/reg/2020/843/oj))

instruments relating to the data economy. Mechanisms for recognition and redress of both individual and societal effects of data technologies can be visualised as central for effective governance, rather than as a barometer to steer technology policy. A governance model that treats reporting of problems and unexpected consequences as a feature rather than a bug of a technologically-enabled society would look very different in terms of its architecture and mechanisms.

6.3.4. Governing for a sustainable data economy

While the DGA creates frameworks for data re-use when it is in the public interest, it is telling that there is little emphasis placed on aspects of embedding sustainable practices within the creation of new bodies of data intermediation and altruism. Focusing on sustainability will involve setting standards and limits for the environmental impacts of computing systems, and putting in place policy instruments and institutional architectures that enable different interests to be prioritised within these limits.

- Placing a premium on efficient computing would affect numerous large-scale systems which are currently main features of the digital landscape. For instance data centres' true net energy costs would have to be taken into consideration, rather than judging their sustainability by the amount of green energy they can access at the expense of other actors.²³¹
- A sustainability-based data governance system would require the government to arbitrate between the interests of the public and private sector on a continual basis, and the resulting international competition for energy resources would be likely to send the world's biggest platforms and service providers in search of more compliant authorities in lower-income countries and regions.²³²

6.4. Data Act

6.4.1. Resisting the commodification of data

The Data Act reproduces a long-standing tradition of defining data as a commodity for private exploitation. The regulation could instead aim to resist corporate-led ecosystems based on the extraction and commodification of data.²³³ This proposal could create the conditions for resisting the privatisation of data's value, rather than just enhancing access and sharing data for private profit. The Data Act is the opportunity to define the conditions for a data governance model based on the idea of data as a common resource that lives outside the private value creation framework to advance toward a public interest discourse around data.

²³¹ Similarly, types of systems such as the large language models on which online translation and search services increasingly rely would be considered unsustainable in terms of their energy costs, as would the growing blockchain-based sector of the tech economy. This would have implications for the current business model of big tech, which is based on infrastructural expansion and continual growth through the addition of new services and features based on the availability of ever-greater computing power. It would also have implications for the functionality of services and products across both the public and private sectors, which currently rely on these invisible but extensive infrastructures for much of their functionality. See, for e.g., Emily M. Bender et al. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.' *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021. <https://dl.acm.org/doi/abs/10.1145/3442188.3445922>

²³² European Commission, 'Commission Puts Forward Declaration on Digital Rights and Principles for Everyone in the EU,' European Commission, January 26, 2022, https://ec.europa.eu/commission/presscorner/detail/en/IP_22_452.

²³³ Sohel Sarkar and Amay Korjan, eds., *A Digital New Deal: Visions of Justice in a Post-Covid World* (IT for Change and Just Net Coalition, 2021).

6.4.2. Enabling public interest use of data

Even though the Data Act aims to create a framework for business-to-government data sharing, the instrument is far from constructing the conditions for a systemic approach for using privately-held data for the public interest. As we have seen, the consolidation of alternative data governance models depends on developing a framework that promotes an institution that oversees data for the public interest. In other words, this translates to challenging the idea of data as private property and promoting the use of data for the public interest. The systemic approach would consider the public agencies, educational institutions, media, and altruistic organisations as actors that use data for public benefit, thus challenging the limited idea of a government as the only actor that could use data for the public interest. The systemic approach means that rather than using an exceptional language for sharing data between businesses and governments, we need to create clear conditionalities and limitations for protecting vulnerable communities against the risk of data analytics.

6.4.3. Recognition of collective needs and rights around data

The map of actors that participate in accessing and sharing data in the Data Act is limited to businesses, consumers and public institutions. This discourse reproduces the tradition of the individual perspective from the GDPR. Public policy after the GDPR could overcome this vision and create ways to claim collective rights over data. Likewise, as we have shown, the collective identities around data governance could help to implement governance outside of the capitalist data extraction model. This kind of policy would not just aim to *fix the market* but to create the conditions for other ecosystems of data governance.²³⁴ The Data Act could be a space that makes possible the creation of new political frameworks and actions with collective data. The examples of data collaboratives and indigenous data sovereignty reveal the potentialities of imaging a multiplicity of actors that define data outside of its fictional commodity nature and move towards recognition, collective interest and rebalancing power relations in the data economy.

6.4.4. Interoperability for challenging the dominant data governance model

The interoperability framework could not respond only to the idea of private value creation but as a way to empower communities in vulnerable conditions. The proposals for interoperability in the Data Act promote competitiveness, markets and the freedom to choose for consumers. This definition of interoperability could lead to more sharing and accessing of data but based on the idea of private gain that leads to centralisation, exploitation of data and less autonomy for vulnerable communities. Interoperability, without challenging the commodification of data, could translate into the centralisation of data in companies, even European ones.²³⁵ The interoperability standards do not emerge in empty spaces, but rather in complex power relationships that, if not properly challenged, could translate into the consolidation of the dominant data governance model based on the commodification of data. Interoperability in the EU context thus needs to create conditions for promoting the public interest and the autonomy of communities through the construction of conditionalities for some datasets to flow to the private profit environment and to assure access for public interest institutions like public bodies, media, research institutions and altruistic organisations. The interoperability framework needs to produce the conditions for creating a semi-common perspective,²³⁶ cooperative data government and public data trust throughout the Union to rebalance the power relationships that shape the current digital economy.

²³⁴ Alek Tarkowski et al., 'Generative interoperability: building online public and civic space.' (Open Future, March 2022).

²³⁵ Ibid.

²³⁶ Ibid.

7. Discussion and conclusions

We find that AI is not a democratic class of technologies, either in terms of development or of use. This is because developing and deploying AI systems at scale is, and will remain for the foreseeable future, a privilege that is mainly accessible to the most powerful actors in society, whether commercial or public-sector. Furthermore, unless we apply substantial resources to counter it, it will continue to rely on large-scale commercial computing infrastructures that are built to channel the power to analyse and intervene to those with the most resources and capacity. As such, the central question this report has addressed is how to foster a positive vision of AI, as contributing to public goods and creating public value, through governance approaches that distribute power over AI systems and the data ecosystems they rely on, and that strongly incentivise good behaviour on the part of those developing and deploying those systems.

Starting from a definition of governance as arbitration between different interests with regard to public and private goods, we have offered a justice-based analysis of the legislative context with regard to artificial intelligence and contributing data technologies and argue that along with considerations of its economic benefits, a strategy as to how European AI generates *public value* could be central to the EU's policy aims. This would reframe the challenge of governing AI from maximising business value while preserving fundamental rights, to creating multiple forms of value that can serve differing and plural societal interests. We define four areas which data and AI governance could serve in order to be qualified as good: preserving and strengthening public infrastructures and public goods; inclusiveness; contestability and accountability, and global responsibility.

We then conducted an analysis of the core set of legislative files relating to AI, and explored how they could be aligned with these justice-based benchmarks for good governance. Our analysis shows that the draft legislative framework pertaining to AI and data would be complemented by these benchmarks for good governance and public value, and that evaluating them in this way shows avenues for further development to align them more explicitly with the public interest. We then identify and examine policy options based on these benchmarks, and on lessons that can be drawn from governance models, instruments and debates in other regions.

Overall, we highlight the importance of collective will and decision-making on the part of societal groups, combined with the normative orientation toward public value, as an important consideration for governing AI and data. The EU's investments in public infrastructure (named in its data strategy, and implied in the Data Governance Act and the Data Act), could be made to reflect plural understandings of how data generates value, especially in terms of both large and smaller-scale computing and data infrastructures. Plural thinking and input on digital infrastructures could help to support and build public goods within the EU, and to render those public goods resilient to capture by big tech.

We also argue that the current thinking on AI governance leaves civil society overexposed to exploitation and rights violations, and that multiple paths to civil society power over AI's development and deployment exist that are not yet being explored. This, we argue, is due to the framing of AI systems or models as single products rather than as components of a field of research, a set of dynamically evolving systems that will be used in different ways to generate different kinds of value over time.

A key measure to counteract this uneven power balance could be digital constitutionalism: an overarching set of aims regarding rights and the equitable distribution of power. This approach to lawmaking and regulation, by emphasising the structural balancing of power between different societal, governmental and business actors, and paying attention to the ability to make claims

where power is misused, has the potential to provide a holistic and substantive guide toward realising good governance.

Approaches that are potentially relevant to this problem include data trusts and data commons, which are predicated on establishing controls for data flows that would in turn give sectoral and interest groups understanding of how data is being used for AI applications in a given area. If data which flows between sectors or organisations under the provisions of the Data Act or the Data Governance Act becomes part of legally designed collectives and answerable to those interests, we find that the likelihood of positive shaping on the part of the public, and understanding of when uses of data are out of line with the public interest, would be higher.

Similarly, developing public infrastructures for AI and the data it requires will pay off in terms of the development of public-interest technologies. This logic is visible in the infrastructural needs of collectives who are using technology to preserve language as a component of the right to self-determination.²³⁷ When such systems become useful to a broader public, rather than local experiments, they come to rely on technological infrastructures from big tech firms and the community loses control over the development of the system. Enabling smaller ecosystems to survive within these infrastructures requires making it possible to centre governance and power in the hands of the local developer or community rather than passing it to the owner of the computing infrastructures within which it lives.

We also see potential benefits in structuring governance to foster civil society agency in governing data, namely through using sectoral and local organisations (such as associations, interest groups, municipalities and provinces) to monitor and source issues, positive and negative, with AI systems post-deployment and to feed that information into the oversight system in a continual and responsive way. Such a distributed, domain-related oversight infrastructure could also serve as a tool for identifying incremental harms of the kind most likely to occur with AI and data analytic systems, where a critical mass of complaints builds over time from individuals who are affected enough to make a claim. The oversight and enforcement apparatus for making claims about data harms is overloaded and often dysfunctional. A more efficient system could make use of sectoral bodies and interest groups for both monitoring the effects of new and existing digital technologies on the groups they serve, and for channelling claims when things go wrong.

Such a distributed system would also address the issue of democratising the process of oversight and enforcement with regard to data and AI: as powerful technologies increasingly used on the public in ways that are opaque to individuals, as we identify gaps in oversight and enforcement structures with a public-facing component, that can demonstrate democratic accountability and therefore that are more representative of society. It is a challenge to expect legal professionals to recognise emergent harms in diverse societies without a direct connection to civil society, and it is also a challenge to civil society to identify and seek redress in a situation where they do not see themselves represented in oversight bodies.

Finally, our analysis makes the case for the EU to further conceptualise what kind of public good it considers that data should be. The legal framework clearly gestures toward creating public value from data, but the mechanisms are unclear and vary across legislative instruments. Options for central aims include providing equitable development and access, creating and resourcing environmentally sustainable infrastructures and practices, centring on the FAIR principles in data policy as a requirement for global responsibility, and recognising that government is not the only actor who can use data in the public interest. Distributing and devolving the power to access, use

²³⁷ Karen Hao, 'A New Vision of Artificial Intelligence for the People,' MIT Technology Review, April 22, 2022, <https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/>.

and benefit from data is possible, as long as protection from harm is also addressed as a common good to be distributed equitably.

With a particular focus on artificial intelligence (AI), this study identifies and examines policy options for the EU's data governance framework that align with a data justice perspective. A data justice approach is one that centres on equity, recognition and representation of plural interests, and the creation and preservation of public goods as its principal goals. The analysis offers both an assessment of the EU data governance strategy overall and specific policy options for the AI act, the data governance act and the data act.

We propose four benchmarks for good data governance according to principles of justice: preserving and strengthening public infrastructures and public goods, inclusiveness, contestability and accountability, and global responsibility. Exploring examples of different governance models, we examine how these models and options intersect, and what lessons they offer for the EU case.

This is a publication of the Scientific Foresight Unit (STOA)
EPRS | European Parliamentary Research Service

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.