
WHITE PAPER

# DATA SOLIDARITY

December 2022

Barbara Prainsack*
Research Platform Governance of Digital Practices, and Department of Political Science, University of Vienna, Austria *Corresponding author: barbara.prainsack@univie.ac.at

Seliem El-Sayed
Department of Political Science, University of Vienna, Austria

Nikolaus Forgó
Research Platform Governance of Digital Practices, and Department of Innovation and Digitalisation in Law, University of Vienna, Austria

Lukasz Szoszkiewicz
Faculty of Law and Administration, Adam Mickiewicz University, Poznań, Poland

Philipp Baumer
Department of Political Science, University of Vienna, Austria

# Table of Contents

1. Preface 2

2. The goals and pillars of data solidarity 4

3. How is data solidarity different from other concepts? 10

4. Data solidarity in existing legal frameworks 20

5. Policy instruments 30

5.1 Pillar I: Facilitating data use that creates significant public value 30

5.2 Pillar II: Better harm mitigation 33

5.3 Pillar III: Sharing the profits of corporate data use with the public 36

6. Other ways to enable and improve data solidarity 41

7. Summary and recommendations 45

8. Future research agenda 48

References and appendices 49

# 1. Preface

The issues emerging from datafication and data use in the digital era cannot be addressed effectively with the tools of the paper age. The risks and benefits of digital practices accrue for a much wider range of people than the primary data subjects whom current frameworks seek to protect. In addition, digital practices are embedded in stark power asymmetries. The domain of health, for example, has seen a structural shift in the last two decades. Technology companies have joined health care providers, pharmaceutical companies, manufacturers, and regulators as key players in the field (Prainsack 2020). In fact, they often perform several roles simultaneously, producing the devices that are used for patient monitoring as well as the software to do so.

Technology companies have become important players in research as well; the founders of Google, Baidu, Meta, or Xiaomi – to name only some of the most powerful – have become forces to reckon with in the funding landscape across the globe. They run research institutes and provide research funding in areas that they consider important, thus setting and controlling research agendas – without the public accountability requirements that traditional research funders face (see, for example, McGoey 2015; see also Sharon 2021). Finally, philanthropies and foundations founded by tech entrepreneurs also steer and shape research by setting up institutes and providing endowed chairs. While corporate funding for research is nothing new, the arrangements that we currently see are unprecedented in terms of the scale and force of private funding (Popkin 2019). Notably, they now also extend to the ethics of technologies (Phan et al. 2022). They also play an important role in determining what is researched and how, and what gets published (Van Dijck, Poell and De Waal 2018; Bero 2019).

Against this backdrop, the call by the report of the Lancet and Financial Times Commission on Governing Health Futures 2030 for digital technologies in health and health care to be driven by public purpose, and not by profit, could not be timelier (Kickbusch et al. 2021: 1728). One key element of such a new approach is to place stronger emphasis on collective control, responsibility and oversight. Not in the sense that 'public interest' is supposed to overrule individual rights - but what is needed is an approach that accommodates the ways in which justice, equity, or privacy are both individual rights and collective goods at the same time (see also Roessler and Mokrisinska 2013). Sometimes a person's individual rights and interests conflict with collective goods, and good mechanisms must be in place to ensure that individual rights are not sacrificed on the altar of alleged public interest. There are, however, many situations in digital societies where individual and collective interests are not opposed in principle. Inequities in digital societies, for example, do not only hurt the people who are affected by them in direct and immediate ways, but society as a whole. Similarly, everyone benefits from inclusive deliberations about what parts of people's bodies and lives should be exempt from datafication, the existence of effective mechanisms to support people who experienced harm from data use or the fair taxation of commercial activities in digital societies. Pretending that the challenges of the digital era can be solved solely by giving individuals more control over their data is not an emancipatory

stance. Transferring all responsibility to the individual does not consider the existing asymmetries between the citizen and corporate or state actors. These asymmetries help powerful commercial players more than anyone whilst also impede collective actions.

The concept of solidarity, applied to data governance, offers an approach to address the issues raised above. Solidarity-based data governance (in short: data solidarity) seeks to increase collective control, oversight and ownership over digital data and resources. In today's societies, digital technologies and practices are entrenched in every domain of practice. Even people who are not heavy users of digital technologies contribute to the benefits that emerge from digital data and practice. They do so when data about their bodies and behaviours are captured by public institutions and companies, and as members of societies that make available the technical, social and knowledge infrastructures necessary for the generation and analysis of digital data. In short, in digital societies, all people contribute to the benefits resulting from digital data and practice. Similarly, everyone bears risks – not only that their privacy will be infringed, but also that they or other people will be discriminated against, profiled, or otherwise harmed as a result of data analytics and other data practices in fields as diverse as policing, administration and insurance. Against this backdrop, approaches that seek to increase the control of individuals over the use of their data remain important, but they are not sufficient to address the issues emerging from political and economic constellations. Data solidarity's core premise is that the benefits and the risks of digital practices need to be borne by societies collectively.

The structure of this White Paper is as follows: After sketching our understanding of data solidarity and what a governance framework based on it should entail (Section 2), we discuss how data solidarity is different from related concepts (Section 3). We then give an overview of manifestations of data solidarity in existing legal frameworks (Section 4). Following this, we elaborate on policy instruments that can realise the proposed solidarity-based data governance framework (Section 5). We then discuss other ways to enable and improve data solidarity (Section 6). We end by providing specific recommendations to policymakers and other actors (Section 7) and presenting a brief research agenda for the immediate and near future (Section 8).

## 2. The goals and pillars of solidarity-based data governance

**What is solidarity?**

Prainsack and Buyx (2011, 2017) defined solidarity as a practice that reflects people's commitments to supporting others with whom they recognise similarity in a relevant respect. The similarities with others that people recognise are not necessarily 'objectively' existing properties. Instead, they are characteristics that we have learned to attribute to ourselves and to others. They are also the categories that are used in public discourse to classify people. Solidarity takes place when people act upon what connects them to others despite all the things in which they inevitably differ – including the asymmetries of power, resources and agency that shape people's lives. In this manner, solidarity does not deny difference but takes place when people see things that they have in common and act upon them, rather than foregrounding their differences. While solidarity is not a replacement for justice, when solidaristic practice of individual people or groups scales up into institutionalised forms of solidarity, it can help to counteract injustices.

**Why does data governance need solidarity?**

Solidarity harnesses the fact that many people are willing to support others with whom they recognise similarities. They are willing to accept certain risks if other people benefit. Applied to data, this means (at least) three things. First, whenever and wherever people accept such risks, there is a collective responsibility to ensure that data use that creates significant public value receives public support. Second, people who experience harms from data use also receive public support and are not left to fend for themselves. And third, commercial profits made based on data use are shared with the public.

The way we govern digital practices today is often not conducive to the goals of solidarity-based data governance (data solidarity). On the contrary, it has made possible very problematic concentrations of data and power that are not only a threat to democracy but arguably also to human wellbeing. One reason for this is that governments have often enabled, rather than limited, the expansion of the power of multinational technology companies. Another one is that data governance instruments so far have mostly focused on increasing the control of individuals over the use of their personal data. Let us explain why and how the latter is a problem.

Since 1970, when the world's first data protection law became effective in the state of Hessen in Germany, data protection legislation has proliferated all over the world. To date, as many as 137 countries have adopted data protection and privacy legislation (UNCTAD 2021). In many of these laws, informed consent is seen as an expression of informational self-determination. It has been the single most important instrument to justify the processing of personal data. But in today's world, this strong focus on individual control is problematic. First, on a conceptual level, the very notion of informational self-determination - which was conceptualised by Germany's Constitutional Court as early as 1983 and has been one of the

cornerstones of data protection law - treats data as if it were individual property. Personal data processing, however, should not be placed within a property-like regime (see Section 6). Another problem with relying too much on individual-level control over data is that it turns people's right to control their data into a de facto obligation. It thus places those people who, for whatever reason, cannot fully exercise their individual-level control right at a disadvantage. An additional issue is 'decision fatigue' (Vohs et al. 2008): When people are asked to consent to data use several times a day and in scenarios in which they want or need quick access to a service, they stop paying attention; we cannot give meaningful consent when we are using a mobile app on the go. As some studies have demonstrated, decision fatigue – or 'privacy fatigue', as Choi and colleagues call it – can have a more substantial impact on user behaviour and decisions than substantive privacy concerns (Choi et al. 2018). The nature and length of the terms of service of many online tools can exacerbate this problem (Obar and Oeldorf-Hirsch 2020). Individuals cannot reasonably be expected to read and digest these. Moreover, in some cases, access to important services is not possible when people do not consent to data processing by commercial providers. This is the well-known but little addressed problem of 'dark patterns' – strategies to manipulate users into selecting options that serve the interest of service providers, not users (Gray et al. 2021). In this sense, the historical focus on individual control over data can been seen to impede collective action against structural injustices and to increase power asymmetries (Prainsack 2020).

Another reason for strengthening collective responsibility and oversight is that in the digital age the benefits and risks of data use do not only affect the people from whom the data come – primary data subjects – but extend to much larger groups – secondary and tertiary data subjects (Taylor 2012; PERVADE 2017). For example, when specific patterns are found in a data set that suggest that people who buy sports goods online have a higher risk of defaulting on their mortgage, this information can be applied to people other than the original group whose data was used to find the pattern in the first place (Dixon and Gellman 2014). While this mechanism is not new – actuarial science in insurance and criminal sentencing in the United States have used this for decades (see, for example, Monahan and Skeem 2016) – the spread of digital practices has made this a much more widespread concern.

Against this backdrop, increasing collective control over data use is particularly important. As noted, data solidarity does not deny that individual control is and remains important in many contexts. Individual control, however, does not go far enough, and it cannot solve some of the most pressing structural challenges. It is not only that collective governance can help to reduce individual 'decision fatigue' by engaging communities in privacy management (and ideally also much more upstream in decisions about what aspects of our lives should – or should not – be datafied in the first place). Collective action in the form of legislation can also help address the asymmetries of power and resources. It can help to ensure that the benefits and the risks that emerge from digital practices and data are shared collectively.

## Summary box: What is data solidarity?

In addition to ensuring that individuals have meaningful control over how their data is used, the digital era urgently needs better instruments and approaches for collective control as well. This is where data solidarity comes in. By focusing on situations and settings where people act upon what they consider to be their similarities with others, and not on their differences, solidarity gets us out of the unproductive dichotomy between individual interest and collective good. When people enact solidarity or when institutional arrangements, policies and norms enshrine solidarity, this does not mean that individuals are supposed to deny their own needs for the sake of others. Solidarity does not mean that collective needs overrule individual ones. When we accept that people's needs, interests, and identities are shaped also by their relationships to their human, natural and artefactual environments, then the relationship between self-interest and other-directed concern is not based on a trade-off, but complementary. In this context, solidarity highlights and supports practices, instruments and structures within which people recognise and act upon the needs and interests of others. It supports that such practices solidify into institutional and legal arrangements.

**Data uses, not data types, shape risks and benefits**

That data can be more easily collected, integrated, and shared in the digital era than in the paper age brings manifold opportunities for new and important insights, but also new risks to people. Traditionally, the key distinction within most regulatory frameworks across the world has been between personal and non-personal data. Moreover, within the category of personal data, certain types of data – such as data on health, sexual orientation, race, or union membership – have been considered particularly sensitive, due to the harm that they can do if they fall into the wrong hands. Many data protection frameworks (such as Article 9 of the GDPR) have special protections in place for this category of sensitive data. Today, any data set – if linked with suitable kinds and quantities of other data sets – can lead to the identification of the person data came from, so that the distinction between personal and non-personal data becomes more and more blurred. Recent studies indicate that individuals are re-identifiable even in country-size location datasets (Farzanehfar et al. 2021). Data can also be used to make inferences about certain groups of people. Seemingly innocuous pieces of information about a person, such as whether she buys sports goods online, or whether she has a telephone landline, could lead to discrimination against this person if it had been found, in other studies or data sets, that these characteristics correlate with supposedly risky, unhealthy, or otherwise costly behaviour. Even data that does not come from humans at all can cause harm to people. For example, if environmental data sets showed higher levels of toxins in a specific area, then people there could face discrimination from insurers based on this information.

For these reasons, it seems misguided to merely speak of specific data types that need special protection. Instead, it is certain types of data use, and the underlying

motivations and goals, that are more or less likely to harm people. From a data solidarity perspective, the most important distinction is between data use that generates significant public value without posing unacceptable risks to anyone and data use that does the opposite. (Some of the authors of this White Papers are currently involved in developing an online tool for a structured assessment of such public value; see El-Sayed & Prainsack 2022).

**Different types of data use**

Data use should be actively supported when it creates significant public value, meaning that it can be plausibly assumed that it will have clear benefits either for many people, for society as a whole or for future generations, and that no person or group is likely to experience significant and undue harm. Moreover, public value will regularly be more pronounced if the benefits are likely to materialise for underprivileged groups, due to the overall lower baseline and potential size of impact (Prainsack and Buyx 2016: 497). In this context it is important to adopt a broad and flexible understanding of underprivileged groups. For instance, adolescents – who are usually not protected under anti-discrimination laws focusing on gender or minorities – are particularly at risk of harm 'in the context of their ability to manage privacy, commercial targeting, reinforcement of gender norms and stereotypes, and hate and abuse in online environments' (Kickbusch et al. 2021: 1733).

Considering the expected public value that specific instances of data use create and the risks that they pose for individuals and groups, data solidarity distinguishes between four categories of data use. The first (Type A, Table 1) consists of data use that is likely to create significant public value and does not pose unacceptably high risks (light blue box). An example is the re-use of previously collected and de-identified urine samples for disease research in a hospital, or the analysis of sales data from a supermarket chain in different urban areas to understand variations in dietary patterns across socio-economic strata by a public health body seeking to improve school meals. Such data use should receive support – by easing regulatory requirements where necessary, or by making available public funding to facilitate such data use.

The second category of data use (Type B, blue box) is unlikely to create significant public value but also unlikely to do harm. An example is the use of customer purchase data by a clothing company to tailor advertisements to groups with different style preferences. If these types of data use create corporate profits, it should be ensured that some of these return to the public domain.

In the third category (Type C, light grey box) is data use that is likely to create significant public value while also posing high risks. For example, if an NGO carried out research on former organised crime members to learn how barriers for people reporting on financial crimes could be reduced, this may yield great public value, but it may also put people at great risk. Another example would be a public transport company that uses fine-grained geolocation data to learn when to increase the frequency of trains. While this may create significant public value, this may create serious harm to low-income groups without an alternative means of trans-

port; for example, if the company also uses the same insights to make trains at peak demand times more expensive. In general, data use in this category should not go ahead, unless the data use strategy can be adapted in such a way that the risks can be reduced to an acceptable level. In the latter case, the data use would become an instance of Type B data use.

The fourth category comprises data use (Type D, grey box) that is unlikely to create significant public value, and that also poses high risks of harm. An example would be a social media company exploring how it can instigate people posting hateful content to be more active on their site as they create more traffic. Such data use should be outlawed with fines severe enough to deter even powerful corporations from breaching the law.

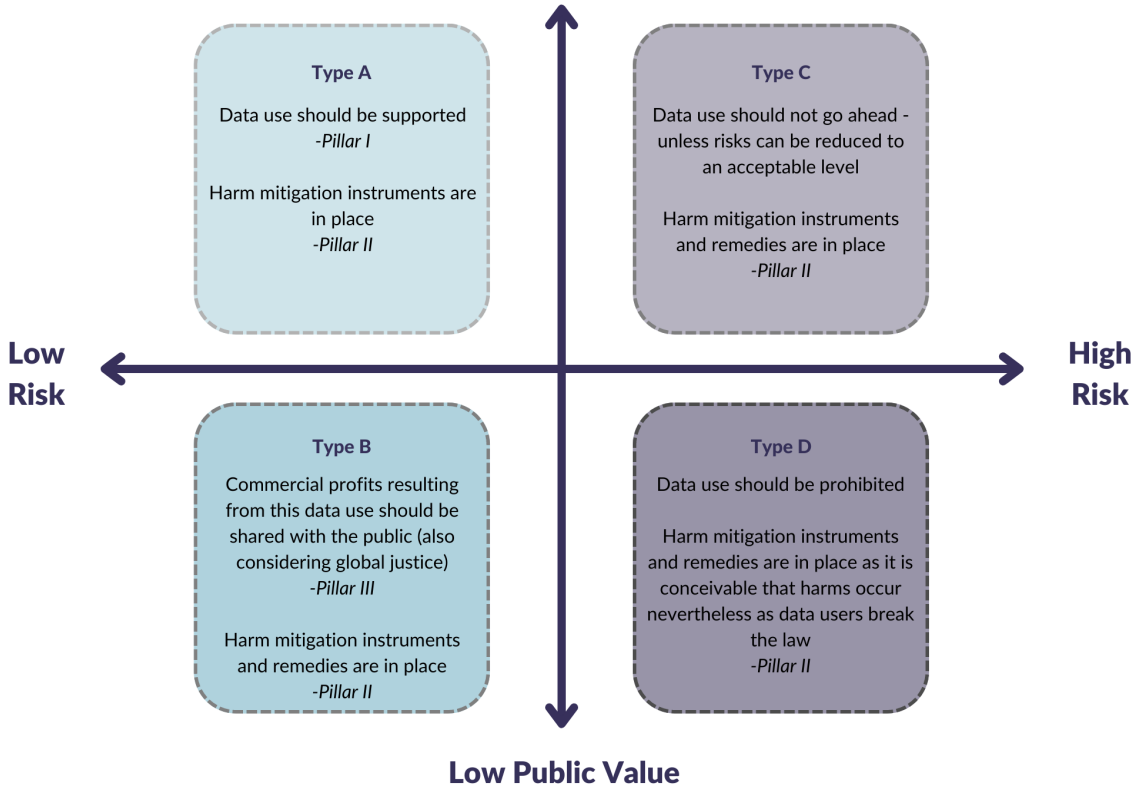


*Table 1: Overview of different types of data use*

The operationalisation of this model in policy and practice requires decisions about who should make assessments and against what criteria. It will be essential to develop rules and norms to ensure that powerful corporate players who exploit the law to their advantage would not be able to claim that everything they do creates significant public value without posing grave risks. At the same time, small and medium-sized data users (individual citizens, small businesses, or non-profits) should not be burdened with additional bureaucracy. One possible solution is to impose different evidentiary burdens on different types of entities. The burden of proof for some types of institutions should be lowered to make up for a potential lack of resources in staff and financial resources (for example, small and medi-

um-sized companies or public institutions). In cases that involve large businesses, the evidentiary standard would require more detailed documentation.

**Three pillars for a solidarity-based data governance framework**

The framework for solidarity-based data governance proposed in this White Paper has three main pillars, with policy instruments geared to the scenarios and categories of data use explained in the previous section (see Table 1 and Illustration 1).

Pillar I seeks to facilitate data use that is likely to create significant value and unlikely to pose unacceptable risks of harm (Type A of data use in Table 1). Pillar II is devoted to preventing and mitigating harm across all types of data use. While societies need to do their best to prevent harm from occurring in the first place (for example, by prohibiting data uses that are likely to lead to significant harm), it is of key importance that when harm occurs nevertheless, people and communities receive adequate support. For this reason, across all types of data use, harm mitigation instruments should be available to all affected people. These instruments need to be easy and fast to access, and they must not depend on any laws being broken, or on the harmed party being able to prove whose act or omission caused the harm (see Section 5.2). Pillar III aims to strengthen mechanisms of benefit sharing to ensure that some of the profits that emerge from commercial data use come back to the public domain that has enabled the data use via the data work of individuals and public infrastructures, etc. The latter point draws attention to the broader political economy that digital practices are embedded in (see Section 5).

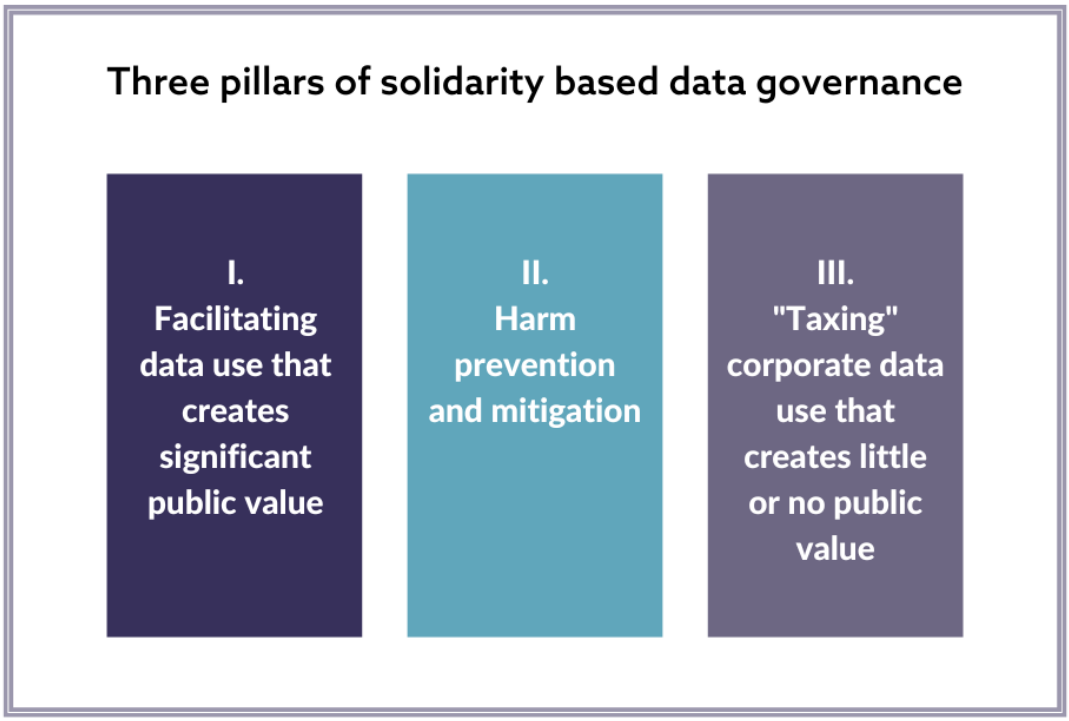


*Illustration 1. Three pillars for a solidarity-based data governance framework*

## 3. How is data solidarity different from related concepts?

To illustrate how data solidarity builds upon, complements, or deviates from other relevant concepts, this section outlines the relationship between data solidarity and these other notions.

**Data sharing**

At times, data solidarity is conflated with data sharing. Data sharing typically refers to people or organisations making data available to third parties, regardless of what purpose the data is used for and to whose benefit. Such an equation of data solidarity with data sharing is incorrect. While some instances of data sharing can be an expression or a result of data solidarity, others are not. In other words, not every instance of data sharing serves the purpose of data solidarity. At times, data solidarity can express itself in refraining from making data available to specific entities (for example, powerful private or public corporations). It can also express itself in refraining from recording some aspects of people's bodies and lives in data in the first place.

Moreover, the notion of data sharing is also very broad. A wide range of diverse activities – from research institutions making interpreted results of genomic research available to participants to customers allowing online companies to use information on how they use their services – all fall under data sharing. This conceptual openness has several advantages. A disadvantage, however, is that it makes it impossible to distinguish between instances of data sharing that benefit people and societies, and those that serve merely the maximisation of commercial profits. For data solidarity, in contrast, this distinction is crucial.

**Digital and data commons**

Commons are social institutions for governing common-pool resources. They are governed by the principles of values of fairness, equality, justice, and sustainability (Bollier and Helfrich 2019). They are characterised by 'a plurality of people (a community) sharing resources and governing them and their own relations and (re)production processes through horizontal doing in common, commoning' (De Angelis 2017: 10). Commons have a long tradition in law and history in fields ranging from agricultural land use to forests to educational resources. There has been some debate about the extent to which commons regimes can be applied to intangible resources such as digital data (see, for example, Prainsack 2019). Some authors, however, use the notion of commons in a much wider sense. They subsume under the label of digital commons those resources including data, information, culture and knowledge that have been 'created and/or maintained online' (Dulong de Rosnay and Stalder 2020) and that are for public use. Such a wide understanding of commons tends to equate commons regimes, where people jointly own the resource and decide over its use, with open access regimes, where anyone can use the resource as they please and nobody can be excluded (Prainsack 2019).

The digital commons movement was particularly strong in the aftermath of the

banking crisis of 2008, but also when access to scientific literature became increasingly expensive, around the year 2010. All over the world, new volunteer organisations, novel digital formats of participation, non-commercial infrastructures to share resources and new social spaces emerged (Fuster Morell 2011). Their intention was to create an alternative to the hegemonic power structures and markets in a networked society (Dulong de Rosnay and Stalder 2020). The success of these initiatives has been compromised by the commercialisation of many open knowledge resources, and also because actors in research-rich contexts could afford 'opening up' their data and resources more easily, and also make better use of data and resources that others had opened up. These issues illustrate once more the limitations of 'commons' concepts that are, in fact, open access regimes, which tend to unfold a Matthew Effect: If everything is up to be taken, then those who already have more power and resources can take more of the open resource.

Data solidarity supports commons as a way for people – at local, regional, national, or transnational levels – to jointly own and govern resources (see Section 5.3). It does not support commons that are open access regimes.

**Data altruism**

The concepts of altruism and solidarity are rooted in different – and incommensurable – understandings of how people act in the world. Altruism assumes that people act either in a self-interested manner (selfishly) or in the interest of others (altruistically). Human practice thus appears as a zero-sum game where the extent to which people pursue their own interests comes at the cost of helping others, and vice versa. Solidarity, in contrast, assumes that people's needs, interests and identities result from their relationships to their human, natural and artefactual environments. Doing something for others does not necessarily take anything away from the one doing it. On the contrary, acting in solidarity with others can empower both the giver and the receiver. While individual self-interest and what is good for others can be at odds, this is not inevitably so. Another difference between altruism and solidarity is that solidarity is relational and bi-directional – that is, it takes place between people – whereas as altruism is unidirectional, going only from a giver to a receiver.

Regarding data altruism, the 2020 proposal for the European Data Governance Act (COM [2020] 767 final; DGA) is of particular relevance. In this proposal, the European Commission defines data altruism as 'data voluntarily made available by individuals or companies for the common good'. It considers it a core concept to facilitate the collection and the processing of personal and non-personal data for public interests. Data altruism, according to the proposal, means 'the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services' (Article 2 [10] DGA).

The draft act suggests that organisations that foster data altruism should be able to become accredited as data altruism entities. To do so, an organisation would

need to operate on a not-for-profit basis and be independent from any entity that operates on a for-profit basis. It would also need to be able to ensure that its activities related to data altruism take place through a legally independent structure, separate from other activities it has undertaken (Article 16).

The draft act also includes a common European data altruism consent form. This form should 'bring additional legal certainty to granting and withdrawing of consent, in particular in the context of scientific research and statistical use of data made available on an altruistic basis' (Recital 39 DGA). The ultimate goal of the initiative is to 'attain a higher level of trust in these services, without unnecessarily restricting these activities, and help develop an internal market for the exchange of such data' (European Commission 2020).

Most of the current literature on data altruism discusses the European Commission's legal terminology and explores what a legal framework that effectively encourages entities to donate their data could look like (Kruesz and Zopf 2021). The European Commission's perspective on 'altruism' seems to be underpinned by specific assumptions about how data subjects can be motivated to make their data available 'for free'. It remains to be seen whether the additional gain of 'certified trustworthiness' of registered data altruism organisations and consent form templates will create sufficient incentives for data sharing.

In the wider legal, policy and ethical literature, there is also no common understanding of what data altruism means. Some authors use it as a generic term to refer to a set of values and practices that is closely linked to the concepts of data donation and data sharing. One attempt to define data altruism was made by the TEHDAS (Towards European Health Data Space) project within the framework of the DGA (see above). For these authors, an example of data altruism is what they call data donations, whereby donors transfer intangible property and the right to use proprietary information to a receiving entity without consideration of what the donors will receive in return (TEHDAS 2021: 8f). These authors also posit that the will to make personal data available altruistically is dependent on the existence of legally binding and transparent mechanisms that ensure the protection of the donor's privacy when processing sensitive personal data (Raj et al. 2020: 2f.). The entities to which people make their data available need to be accountable and trustworthy to the public. When a person feels protected from the consequences of data misuse, the authors argue, they are more likely to hand over their data altruistically.

The recently proposed act on a European Health Data Space (European Commission 2022b) dedicates its Article 40 to data altruism in health. It does nothing more, however, than just referencing the definitions and requirements of the draft DGA.

**Data justice**

Linnet Taylor defines data justice as a concept for national and supranational law-making processes that ensures fairness in the way people are made visible, represented and treated as data producers (Taylor 2017: 1 f), as well as fairness

in the way in which they are subject to data-based decision making in the public and private sector. According to Taylor, legal frameworks that stop at sanctioning visible violations of data protection and civil rights cannot provide justice in cases where violations are invisible to affected individuals or groups (Taylor 2017: 4 f). Most socially deprived and marginalised groups face a particularly high risk of discrimination due to this lack of visibility (Taylor 2017: 5). A legal framework that enforces an idea of data justice and that is capable of mitigating individual or common risks of privacy violations, according to Taylor, must reach three central goals: It must provide individuals with the legal capacity to know about the collection and the use of their personal data, it must enable a detachment of personal data from automated commodification on global data markets while encouraging the analysis of big data for common good, and it must counteract technical conditions that might lead to intentional or unintentional discrimination, such as while applying algorithms to support decision making by public authorities or private companies*. Such rules would allow a shift from individual responsibilities for data protection towards a general public responsibility.

Drawing on the conceptual foundations of different types of justice and the work of Johnson (2016), Heeks and Renken (2018) developed a multi-perspective framework for evaluating conditions of (in)justice that arise from data practices in the field of sustainable development. The first is a procedural perspective that is primarily concerned with the means by which data is extracted, stored and processed. Individual data subjects should have some degree of control over how their data is made available to those who use the data (Heeks and Renken 2018: 92-93). Similar to Taylor's (2017) second pillar of data justice, Heeks and Renken argue that individuals who are potentially affected by data processing should have a certain degree of control by being able to opt in or out, or more broadly by being able to control the terms of their engagement. This can be realised in terms of a need for consent to the entire chain of data processing, or it can be extended to require all processes being performed in a consistent and unbiased way and to provide individuals with legal means to effectively challenge errors (Heeks and Renken 2018: 93).

Second, a rights-based data justice perspective echoes Taylor's (2017) first pillar of data justice, which addresses the question of who is made visible through what types of data and to whom. In this view, (in)justice plays out not only in the way in which people are granted the right to withhold certain information about themselves but also in the way in which they can make themselves seen (that is, when it comes to the distribution of public goods based on certain eligibility criteria). Those who are made visible should be able to determine what is known about them and whether they want to give away their data for a particular purpose at all. The right to data inclusion (in terms of an extension of the freedom of expression), in contrast, implies that everyone has a right to be represented in a data set, even if one's opinions or characteristics constitute outliers to a collective average (Heeks and Renken 2018: 94).

* In the EU, Article 22 of the GDPR is probably the most prominent example to achieve these goals.

Third, an instrumental data justice perspective is concerned with the (in)justice produced by the outcomes of data practices. With this approach, data rights for their own sake are meaningless. Instead, we should focus on the purposes and effects of data use and whether they comply with, or advance, societal core values found in constitutions or in wider notions of social justice (Heeks and Renken 2018: 92).

Besides these three classical justice perspectives, Heeks and Renken (2018, adapting Young 1990) make the case for a structural data justice perspective, which critically attends to the power relations that data governance is embedded in. Here, data governance is assumed to reflect an extension of political struggles into a digital sphere. Data injustice is the result of structural conditions that shape wider societal relations of which data extraction and processing has become one part. This fourth approach is especially concerned with how access to certain types of data are distributed between the Global North and South, for example, or how individuals' legal status impacts a government's use of their data (Heeks and Renken 2018: 96).

Let us now turn to data solidarity and its relation to data justice. Braun and Hummel (2022) argue that data solidarity is a necessary, catalytic element for data justice. They hold that whenever injustices in the realm of data-driven practices arise and lead to individuals or groups being discriminated against or marginalised, there is a need for shared practices of solidarity to address these injustices. In this sense, data justice depends on shared practices of hearing the voices of others (in particular, those who are marginalised) and treating their concerns as collective problems. Braun and Hummel believe it necessary that people engage in shared practices of attending to and acting upon the concerns of others for sustainable real-world arrangements of data justice.

Moreover, data solidarity can play an important role in discovering and identifying injustice (Hummel and Braun 2020). Often people will become aware that they have been subjected to injustice only when they consider their experience in comparison to the experiences of others. Likewise, observers might be alerted to potential injustices experienced by certain individuals or groups on the basis of observing the expressions and the formation of movements of solidarity. Thus, it is important that practices of solidarity within a society are recognised as also epistemically valuable in identifying and addressing injustice (and thus indirectly reinforcing the content of justice).

Having said this, solidarity is not 'automatically' directed towards justice. Dedicating ourselves to the ends of certain individuals or groups does not always lead to arrangements that are just. For solidarity to unfold its catalytic potential towards a more just society, it needs to be inclusive rather than exclusive, spanning a wide range of people instead of being limited to exclusive groups of people bound together by characteristics that are unachievable for most.

**Data donation**

Legal definitions of the concept of donation, leading back to ancient Roman law, focus on the following element: The owner of a thing transfers it to another person or entity without expecting anything in return. The latter aspect – that something is given without demanding or even expecting anything in return – tells us two things about donations. First, they are outside of the commercial domain, meaning that no economic profit motive is attached to a transfer. Second, donations are not reciprocal in a direct and linear manner – even if they, like all other gifts, are embedded in networks of mutual moral and social obligations (Carrier 1991). Instead, donations are indirectly reciprocal. In addition, donations are also rivalrous and consumable. If a person donates money to a hospital, then they cannot give the same money to disaster relief. If someone donates a kidney to one person, they cannot donate the same kidney to somebody else. Traditionally, donations have entailed that there is a consumable thing that is transferred from one entity to another. Can the same be said for data, and especially digital data?

One of us has argued that digital data is best described with the term simultaneity (Prainsack 2019). Simultaneity captures the characteristic of digital data of being able to be in more places than one at the same time, in leaving traces even when they are 'deleted', and of being able to be copied and used by several people at the same time, independent of what the others are doing. This raises the question of whether we can transfer something that is simultaneous – can a person 'donate' their phone records to a research project or their DNA information to a biobank? In the case of post-mortem data use, there is a clear case to be made for the use of the term 'donation' over 'sharing' (Krutzinna and Floridi 2019), because sharing implies that both the entity who shares the data and the party with whom it is shared can access and use the data. If a person shares their car with someone else, both parties can use it. If they share an apartment, they can both live in it. In the case of post-mortem data use (resting upon the informed consent of the donor prior to their death), such a kind of sharing is not possible as the donor is no longer able to jointly use anything. But the use of the term donation is more contentious beyond the context of post-mortem data. It is especially confusing when it is used in a context where people give something to others that they continue to have access to – such as when a person allows data from their mood diary to be used for research.

Whether or not a specific instance of data donation is also solidaristic depends on the motivation of the donor and on the purposes and contexts that the data is used for (see Section 2). In brief, while most types of data donation are solidaristic practices, not all instances of data solidarity are data donations.

### Digital (and data) sovereignty

Digital sovereignty (sometimes also called data sovereignty) is not a clearly defined concept. Very different issues come together under this term and diverse questions are discussed under its remit: What form of control do data subjects need over their data for sovereign handling of data, and how can such control be implemented? How can states or international organisations be in control of the material and immaterial digital infrastructures? The notion of data sovereignty can

refer to control over data on the side of different actors, from individual consumers, digital firms, to entire societies and countries, sometimes leading to conflicting claims of data sovereignty at these levels. Based on their different approaches to key issues of data control, Gao suggested to classify the data sovereignty regimes of three major jurisdictions in the world – the US, China, and the EU – as firm sovereignty, state sovereignty and individual sovereignty respectively (Gao 2021). Data sovereignty is also prominently debated in the context of the design of IT architecture and/or laws governing data processing.

In virtually all accounts, digital sovereignty aims at increasing the control and power of collective actors over data, whereby the type of power in question – unlike in classical concepts of sovereignty – is not brute force or arbitrary power, but closely linked to goals of inclusiveness, deliberation and the fundamental rights of the persons concerned. This, of course, raises the question whether or not the term data sovereignty is a misnomer (McDonald 2022).

One important field of development and debate on the concept of digital sovereignty is Indigenous data and knowledge. Information related to traditional, Indigenous and local knowledge on, for example, river systems, fisheries management, disaster prevention and exploitation of natural resources is being digitised by the public and private research entities that execute access and ownership of data. In many cases, the rights of Indigenous communities remain limited and reinforce 'knowledge colonialism' (Saxinger and First Nation of Na-Cho Nyak Dun 2018). Scientists' misconceptions of knowledge sharing have been addressed in 2007 and again in 2014 by the Canadian-based First Nations Information Governance Committee and the First Nations Regional Health Survey, based on first documents developed in 1998 (First Nations Centre 2007; First Nations Information Governance Centre 2014: 2). They developed the OCAP (Ownership, Control, Access, and Possession) guideline tool for local communities to facilitate research on their land. This policy, which is also meant to serve as a guide for researchers, demands doing away with '[s]ocial research [that] was used as 'an instrument of oppression, imperialism and colonialism' (First Nations Information Governance Centre 2014: 3). Noting that '[n]egative experiences have led First Nations to feel distrustful and reluctant to participate in social research. Past research practices were often disrespectful, damaging and stigmatizing to First Nation people' (First Nations Information Governance Centre 2014: 3). OCAP also insists that First Nations should assert their authority over all research concerning their communities: 'This includes the right to make decisions about what, why, how and by whom information is collected, as well as how it will be used and shared' (First Nations Information Governance Centre 2014: 4).

Prominent voices for the recognition of Indigenous data sovereignty come from Aotearoa/New Zealand, Australia, Canada and the United States (Kukutai and Taylor 2016). Moreover, Article 4 of the UN Declaration on the Rights of Indigenous Peoples (2007) enshrines the right of Indigenous peoples to self-determination, including the right to autonomy and self-government. At the same time, Article 5 recognises Indigenous peoples' right to maintain and strengthen their distinct institutions while retaining their right to participate fully in the political, economic,

and cultural life of the state. Cultural life encompasses science and participation in scientific progress, which are indispensable for enjoying human rights on equal footing with others.

To the extent that data/digital sovereignty seeks to increase collective control and oversight over data, it shares a key concern with data solidarity. In particular, data solidarity is closely aligned with the ontologies and demands of Indigenous data sovereignty. In contrast to some understandings of data or digital sovereignty, however, that focus on the state controlling digital data and resources, data solidarity is not primarily focused on the state as a collective actor (also because states have been part of the problem that we are now trying to solve).

**Digital humanism**

Following the historical notion of Renaissance and Enlightenment humanism, today, humanism is often considered as a guiding principle for human interaction. In the context of digital practices, it is often used to refer to human-machine interaction that is anthropocentric, and that is respectful of values such as privacy, dignity, or solidarity (see Autili et al. 2019; Fry 2019; Mittelstadt et al. 2016; Werthner et al. 2022).

Scholars and practitioners in the field of digital humanism work on a range of questions such as: How can humanistic values be translated into tools for governing data practices in the public and private sector? What requirements must be set, for example, for the application of algorithmic decision making to avoid that the quest for technological efficiency and economic growth hurts fundamental human rights? These questions make the protection of human rights a crucial subject of data governance.

An example is the Vienna Manifesto on Digital Humanism (2019). This document, to which scholars and practitioners from a wide range of fields contributed, defines digital humanism as the linking of humanistic ideals with critical thoughts about technological progress. It treats digital humanism as an interdisciplinary approach to the complex interplay of technology and humankind for a better society and life, fully respecting universal human rights. Technical innovative power must be human-centred. The increasing transition of data-based human decision making towards complex automated systems is often driven by the assumption that such systems outperform human mathematical and analytical skills, especially regarding pattern recognition based on large-scale data. The transition to digital decision-making systems may reach a point at which human judgment has been crowded out by machine 'action' (Nowotny 2022: 320). Digital humanism opposes such an understanding of supposedly 'autarchic technological development' (Nida-Rümelin 2022: 74). Neidhart and colleagues see digital humanism as 'an approach to counteract negative effects of the digital transformation, [that] aims to foster the control and the design of digital infrastructure in accordance with human values and needs' (Neidhart et al. 2022: 335). In this context, digital humanism emphasises the exclusive human abilities to determine the purposes of computational problem-solving mechanisms by shaping the premises and

values being applied in them. When policymakers as well as users are aware of these abilities, this can encourage them to focus on the opportunities that digital technologies provide to improve human living conditions, given the premise of a legally binding normative demand to use technological innovation for this very purpose. In other words, digital humanism does not reject digitisation, but it uses 'the potentials of digitalization that relieve the burden of unnecessary knowledge and calculations in order to give people the possibility to concentrate on what is essential and contribute to a more humane and just future for humanity' (Neidhart et al. 2022: 335).

While data solidarity shares a lot with digital humanism, it is not limited to humanism in terms of the substantive values that it draws upon.

**Open Data and Open Science**

Open Data describes any data that can be freely accessed, (re)used and shared by anyone without restrictions. Open Data has been developing as a movement for decades (Kitchin 2014) and has gradually influenced public policies at the national and international levels (see, for example, the adoption of the International Open Data Charter). The primary focus is facilitating the disclosure of publicly held data in open repositories in a machine-readable format. To this end, countries develop Open Data portals and Application Programming Interfaces (APIs) that enable businesses and researchers to access and process administrative data efficiently. Open Data is considered a vehicle for developing science, technology, innovation, and the economy. In this context, it has also been recognised by the United Nations as an initiative that can significantly contribute to achieving the Sustainable Development Goals (Szoszkiewicz 2021).

Open Data remains closely related with Open Science. The latter combines various concepts, movements and practices aiming to make scientific knowledge openly available and accessible to everyone (SC-PCB-SPP/2021/OS/UROS, para. 6). Openness refers primarily to the absence of cost barriers, but it also extends to informal barriers related to the discoverability of data, digital skills and available resources. What is to be made openly available are mainly to the means and outputs of knowledge production, particularly academic publications and scientific data. The latest policy instruments also apply to algorithms, source codes, software, and workflows (OECD/LEGAL/0347).

Open Science has already influenced policies at the international and national levels. For example, the OECD (2021a) and UNESCO (2021a) have adopted recommendations in this matter. Although this concept encompasses various research outputs, data accessibility remains at the heart of the policies. Examples include recent developments in Latin America. Since 2016, Argentina has been facilitating the accessibility of scientific data by implementing a national strategy and regulatory initiatives. The Argentine Science and Technology Information Portal currently contains more than 100,000 open access publications and scientific data, including biological, genomic and satellite images. The data released encompasses scientific data produced by publicly funded projects, including research and

development projects co-financed with public funds (Luchilo et al. 2018: 11). In Colombia, the Biodiversity Information Systems launched in the early 2000s facilitates access to biodiversity data. The repository includes publicly and privately funded scientific data. Today, Colombian and other Latin American companies are among the most active depositors of biodiversity data to the Global Biodiversity Information Facility (Figueira et al. 2022).

Especially insofar as the facilitation of data use that creates significant public value is concerned, Open Science shares important goals with data solidarity. At the same time, as critical scholarship has emphasised, openness is not an end itself. If Open Science is not put in the service of substantive goals such as increasing equity and justice, it could have merely cosmetic effects and even contribute to the exacerbation of the gap between researchers and publics in resource-rich and resource-poor contexts (see Levin et al. 2016; Bezuidenhout et al. 2017; Nerlich et al. 2018).

# 4. Data solidarity in existing legal frameworks

**Solidarity and data solidarity at the United Nations (UN)**

The legal framework for international solidarity is derived from three general UN sources: the Charter of the United Nations, the Universal Declaration of Human Rights and international human rights treaties. After incorporating input from numerous consultations, the UN Independent Expert on human rights and international solidarity submitted a draft declaration on the right to international solidarity in 2017 (A/HRC/35/35). In its opening provision, international solidarity is defined as 'the expression of a spirit of unity among individuals, peoples, States and international organizations, encompassing the union of interests, purposes and actions and the recognition of different needs and rights to achieve common goals' (A/72/171, para. 33). Although data is not expressly mentioned in the draft declaration, its role has been discussed in the thematic report related to COVID-19. The Independent Expert has noted that the waiver from the specific provisions of the TRIPS Agreement between India and South Africa, including the protection of medicine test data constituted a positive expression of international solidarity (A/HRC/47/31, para. 42).

The concept of solidarity forms a basis for the work of the United Nations Educational, Scientific and Cultural Organization (UNESCO). The organisation treats solidarity as an essential way to address challenges in the domains of science and culture that cannot be effectively tackled at the level of individual states, such as the preservation of biodiversity, cultural heritage, or scientific progress. The constitution of UNESCO considers the 'intellectual and moral solidarity of mankind' as one of the foundations for peace and security. Solidarity is also evoked in the context of addressing asymmetries between people living in low-resource conditions and those that lead privileged lives, as well as the need to establish international institutions and instruments to address global challenges such as climate change or pandemics (see, for example, Elungu 2009). The concerns about the complex character of challenges leading to increasing inequalities worldwide remain valid in the digital era (Ragnedda and Gladkova 2020).

To address challenges posed by digital transformations, UNESCO recently adopted several recommendations including guidance on the obligations of member states related to data governance and international cooperation in this matter. The Recommendation on Open Science (2021a) highlights the obligation of member states to promote international cooperation in the context of open science, including universal access to the outputs of science, data included, to reduce digital, technological and knowledge gaps between states and communities (SCPCB-SPP/2021/OS/UROS, par. 22). The Recommendation on the Ethics of Artificial Intelligence (UNESCO 2021b) underlines the need to facilitate fair access to AI technologies to address challenges to the diversity and interconnectivity of cultures and ethical systems (SHS/BIO/PI/2021/1, preamble). In this context, UNESCO stresses the 'responsibility of solidarity with the least advanced' to ensure that everyone benefits from scientific progress (SHS/BIO/PI/2021/1, par. 28). This

responsibility is crucial in tackling global health risks and uncertainties (SHS/BIO/PI/2021/1, par. 121).

Both recommendations seek to strengthen the democratisation of knowledge and access to publicly funded data. They also encourage states to engage and collaborate with business actors to expand the scope of Open Science principles to the private sector (SC-PCB-SPP/2021/OS/UROS, par. 16g). Another recommendation highlights the unique status of scientific researchers and the obligations of states to incentivize the free circulation of data and to provide institutional and financial mechanisms to support them (39 C/Resolution 85, para. 16a(i)). All UNESCO instruments call for the dialogue between various knowledge systems, which could better utilise traditional knowledge of members of Indigenous groups and local communities while also recognising their rights as knowledge holders (SC-PCB-SPP/2021/OS/UROS, par. 11). This recommendation also resonates with the concept of Indigenous data sovereignty as introduced above (see Section 1.3).

The principle of international solidarity has been operationalised within the international human rights law framework. UNESCO published a catalogue of cultural rights enshrined in the International Covenant on Economic, Social and Cultural Rights of 1966, which remains the primary source of legally binding human rights obligations for states worldwide. The covenant formulates expressly at least two such obligations: an obligation on 'the conservation, the development and the diffusion of science and culture' (Article 15, par. 2) and an obligation to encourage and develop international contacts and cooperation in the scientific and cultural fields (Article 15, par. 4). There was little guidance on the normative content of these provisions for a long time (Chapman 2009).

In 2020, the UN Committee on Economic, Social and Cultural Rights adopted General Comment no 25, aiming to address this gap. Although not legally binding, it contains an authoritative interpretation of the covenant and guides states in its implementation (Azaria 2020; CCPR/C/GC/33, par. 13). General Comment no 25 intends to clarify the relationship between scientific progress and the enjoyment of economic, social and cultural rights. To this end, the committee interpreted several obligations relevant to data solidarity. In line with UNESCO, the committee recognised that publicly funded research findings and data should be accessible to the public, and that states should promote open science (E/C.12/GC/25, par. 16). Ensuring equitable and open access to data also includes an obligation to empower citizens by delivering basic knowledge of science within the education system (E/C.12/GC/25, par. 49). Only then can individuals and communities reap the benefits of data and act as rights-holders instead of merely being the objects of protection. The committee critically assessed existing power asymmetries, in particular data exclusivity regimes included in TRIPS-plus agreements (E/C.12/GC/25, par. 61; WHO 2017) and excessive capacity of some business actors to exploit Big Data (E/C.12/GC/25, par. 76). To this end, the committee encouraged states to 'regulate the ownership and control of data according to human rights principles' (E/C.12/GC/25, par. 76). When considering the periodic report of the Czech Republic in 2022, the committee also specified that the data should be released in compliance with the FAIR principles of data management, requiring that data should

be findable, accessible, interoperable, and reusable (E/C.12/CZE/CO/3, par. 51c).

All these elements suggest that the committee perceives the ability to use some types of data (at least publicly funded data) as creating public value that should outweigh particular intellectual property rights. At the same time, its approach to privately funded data is more nuanced. The committee explicitly stated that the intellectual property framework should never lead to denial or restriction of the right to enjoy the benefits of scientific progress stipulated in Article 15 of the covenant (E/C.12/GC/24, par. 24). The committee also provides several practical examples that would constitute a violation of the covenant, including the restriction of access to essential medicines or productive resources such as seeds. In this context, the intellectual property protection of privately funded data would need to be assessed on a case-by-case basis, considering such elements as the impact on the enjoyment of human rights (for example, data on COVID-19 and its impact on the right to health), the financial burden of data production or the impossibility to reproduce certain research data. The latter could be of specific relevance as many of the data sets are unique and cannot be reproduced (Wolkovich et al 2012).

**European Union**

Within the European Union, solidarity plays an important role in primary law (the EU treaties). In particular, solidarity is the heading of Title IV of the Charter of Fundamental Rights, which deals with collective rights such as fair and just working conditions, health care and consumer protection. In sources of secondary law (Laws issued by institutions of the EU), however, references to solidarity are scarce. There is virtually no mention of solidarity in connection with digital practices, even though the EU has been very active in shaping digital infrastructures, particularly since the start of the European Commission headed by Ursula von der Leyen in 2019. 'A Europe fit for the digital age' is one of its six top priorities and a cascade of legislation has been proposed in the last few years. Most proposals are currently under negotiation between the European Commission, the European Council, the European Parliament, and other stakeholders.

There is no shared concept of data solidarity enshrined in these different pieces of legislation. They also target different domains of practice and policy (some of them economic, others driven by human rights protection), with many of them trying to mitigate risks or to change undesirable practices in digital societies. The term solidarity is not used in the DGA, the Data Act, the Artificial Intelligence (AI) Act, the European Health Data Space, the Digital Services Act and the Digital Markets Act. Instead, solidarity regulation is delegated from general legal provisions for executive measures to be undertaken by a competent authority. The authority is the agent that shall bring solidarity into life. Many of the proposed acts have in common that they do not regulate aspects of data solidarity directly but foresee the establishment of new authorities that will monitor and steer data usage scenarios in the spirit of solidarity. For example, the monitoring and supervising of compliance is the main task of the authority foreseen in Article 21 of the DGA (European Commission 2020). According to the European Health Data Space regulation, member states must set up a digital health authority (Article 10) that has to

cooperate with the European Health Data Space Board (which is, in turn, regulated under Chapter 6 of the draft). The AI Act (COM/2021/206 final) foresees the creation of a 'national supervisory authority' to which each member state will assign the responsibility for the implementation and application of this regulation, for coordinating the activities entrusted to that member state, for acting as the single contact point for the European Commission and for representing the member state at the European Artificial Intelligence Board (Article 3 (42)). In particular, providers of high-risk AI systems would have to provide that authority with all the information and documentation necessary (Article 23). Similarly, the European Data Act (COM(2022) 68 final) foresees that each member state will designate one or more competent authorities as responsible for the application and enforcement of this regulation (Article 31). The Digital Services Act and Digital Markets Act also delegate application and enforcement to competent authorities.

The most relevant example for this approach, the DGA, seeks to support the exchange of personal and non-personal data by taking measures that are supposed to increase trust, and in particular by establishing competent bodies (Article 7) and single information points (Article 8) in each member state. In the words of the European Commission, this 'approach is designed to ensure that data sharing services function in an open and collaborative manner, while empowering natural and legal persons by giving them a better overview of and control over their data. A competent authority designated by the Member States will be responsible for monitoring compliance with the requirements attached to the provision of such services' (European Commission 2020). Competent bodies will, as a rule, have to decide on requests for data re-use within two months (Article 9).

**China**[*]

Before exposing elements of the three pillars introduced above in Chinese law, let us briefly look at the current state of data protection in law in China. In recent years, a series of laws, regulations, standards and specifications related to data protection have been issued in China.

First, at the national level, the National Security Law, the Network Security Law (commonly known as the Cyber Security Law), the Data Security Law and the Personal Information Protection law have jointly built a basic 'legal fortress' for data protection. Second, at the local level, 18 provinces and cities have published relevant data regulations to promote the application of big data and data solidarity. Among them, Anhui, Fujian, Guizhou, Hainan, Heilongjiang, Jilin, Liaoning, Shandong, Shanxi, and Tianjin have issued regulations for big data, and Chongqing, Shanghai, Shenzhen, and Zhejiang have issued regulations for data processing more broadly. In addition, Guangxi, Henan, Jiangxi, Sichuan, and others have published drafts of relevant data regulations. The relevant data legislation in various regions is based on promoting data utilisation and industrial development.

As a next step, we will look at regulations that can facilitate the incorporation of

* This section is based on contributions by Shi Li and Henry Gao, whose help we gratefully acknowledge. The usual disclaimer applies.

the data solidarity framework.

*Pillar I: Facilitating data use that creates significant public value*

Open public data is a principle that is widely recognised in the fundamental data laws in China. For example, according to Article 18 of the 2016 Cyber Security Law, 'the state encourages the development of network data security protection and utilization technologies, promotes the opening of public data resources, and promotes technological innovation and economic and social development'. Similarly, the 2021 Data Security Law also includes the following two provisions:

Article 41: State organs shall follow the principles of impartiality, fairness, and convenience for the people, and promptly and accurately open government data in accordance with regulations, except those non-open as required by the law.
Article 42: The state formulates a government data open catalog, builds a unified, standardized, interconnected, safe and controllable government data open platform, and promotes the open use of government data.

In line with the principles under the two laws, China's Central People's Government also issued the Guiding Opinions of the State Council on Strengthening the Construction of Digital Government in June 2022, which included as one of the top priorities for the government to 'build an open and shared data resource system', with the detailed plan as follows:

Compile a public data open catalog and related responsibilities list, build a unified, standardized, interconnected, safe and controllable national public data open platform, open public data by classification, promote the development and utilization of public data resources in an orderly manner, and improve the use of public data in various industries and fields. The ability to promote economic and social development. Promote the 'unified acquisition and sharing' of social data, realize the sharing and sharing of data across regions, departments, and levels, and improve the efficiency of data resource use. Promote the integration and application of public data and social data, and promote the circulation and utilization of data. (State Council of the People's Republic of China. 2022)

While such a national public data open platform is yet to be established, several local governments have taken the lead in data sharing and opening. Shanghai is one of the vanguards of China's digital economy. The Shanghai Data Regulations took effect on 1 January 2022. They focus on three aspects: data rights protection, data circulation and utilisation, and data security. In terms of public data, the regulations aim to improve the public data resource system, formulating public data resource planning. Based on big data resource platforms through Shanghai's big data centre, the regulations try to establish a unified public data directory management system throughout the city, and to promote public data sharing and opening based on the big data resource platform.

Zhejiang Province plays a big role in China's digital economy as it is the home to

the digital giant Alibaba. Its Regulations on Public Data clearly define the scope of public data and the norms of platform construction and data collection. The regulations specify that the data collected and generated by Zhejiang's public authorities, organisations authorised by laws and public service operation units – such as for water supply, power supply, gas supply and public transport – in the process of performing their duties or providing public services are public data. According to the application requirements of the province, the data provided by Zhejiang's administrative agencies and relevant state departments such as those of taxation, customs and financial supervision fall under the remit of public data as defined in the regulations. The regulations point out that the sharing of such public data between public administration and service entities should be the rule, and not the exception.

The term 'public data opening' as mentioned in these regulations refers to a public service that provides public data to natural persons, legal persons, or organisations without legal personality according to law. Public data 'opening' (in the sense of data provision) should follow the principles of fairness, high quality and convenience to the people and public data can be divided into unconditional open data, restricted open data, and prohibited open data. Public data should not be opened under any of the following circumstances:

1. If it endangers or may endanger national security after the opening.
2. If the public interest may be damaged after the opening.
3. If it involves personal information, trade secrets or confidential business information.
4. If the data acquisition agreement stipulates that it shall not be opened.
5. It shall not be opened to the public according to laws and regulations.

The public data specified in Item 3 may be included in the restricted open data or unconditional open data under any of the following circumstances:

1. If the public data involving personal information has been anonymised.
2. If the public data involving trade secrets or confidential business information has been desensitised.
3. If public data involving personal information, trade secrets or confidential business information has been authorised to be opened by a specific natural person, legal person or organisation without legal personality according to law.

The provincial department in charge of public data is tasked – in conjunction with the provincial departments of cyberspace, public security, economy and information technology – with formulating technical specifications for desensitisation and declassification of public data.

At the same time, to integrate public data platforms, Zhejiang's regulations put forward six requirements:

1. Building an integrated digital infrastructure.

2. Building an integrated and shared open channel through which public management and service institutions should share open data, rather than build new channels.
3. Building an integrated digital resource system as a whole to promote the intensive management of public data, applications, components and computing power, and to realise the multi-party circulation of data.
4. Building an integrated data resource system in the form of a unified provincial basic database and a provincial, city and county level thematic database.
5. Establishing an integrated data catalogue system.
6. Building an integrated data standard system to achieve unified standards for public data platform construction, public data processing and security management.

As the first national big data experimental zone in China, Guizhou is also the first province to legislate on big data. It has successively formulated and issued regulations such as the Regulations on Big data Security and the Regulations on Promoting the Development and Application of Big Data. The Regulations on Data Sharing and Opening came into force on 1 December 2020. They define government data, government data sharing and government data opening. All types of data created or received by administrative bodies while carrying out their legal obligations are referred to as government data, which is divided into three categories: unconditional sharing, conditional sharing, and non-sharing. Unconditional sharing applies to data that can be shared with all administrative bodies. Conditional sharing applies to data that can be shared fully or partially with pertinent administrative bodies. Non-sharing applies to data that cannot be shared with other administrative bodies.

Many provinces, including Tianjin, Hainan, Shanxi, Jilin and Anhui, have successively introduced relevant laws and regulations to promote the application of big data, which form the basis for data solidarity.

*Pillar II: Harm prevention and harm mitigation*

Although non-sharing provisions could be seen as ways to prevent harm, there are no instruments for harm mitigation in the sense of supporting people who have experienced harm beyond legal redress (see e.g. Chapter 6 of the Data Security Law, and Chapter 7 of the Personal Information Protection Law).

In terms of legal redress, the regulations of Shanghai specify that if the handling of personal information violates their provisions and infringes upon the rights and interests of many individuals, the people's procuratorate, the Municipal Commission for the protection of consumers' rights and interests and the organisations determined by the state Internet information department may bring a lawsuit to the people's court according to law.

In terms of policing data users, the regulations of Zhejiang, for example, specify

that the public management and service institution and the public data department should order a natural person, legal person, or organisation without legal personality to make corrections and temporarily close their authority to obtain relevant public data if they are under any of the following circumstances:

1. Using data beyond the scope agreed in the public data open utilisation agreement without consent.
2. Failing to take security measures in accordance with the public data open utilisation agreement and the security commitment.
3. Seriously violating the public data platform security management norms.
4. Otherwise seriously violating of the public data open utilisation agreement.

If the required corrections are not made, the authorities should terminate the opening of relevant public data.

The Regulations of Guizhou Province on Promoting the Development and Application of Big Data say that those who illegally collect and sell data involving national interests, public security and military scientific research and production in violation of the provisions shall be punished in accordance with the provisions of relevant laws and regulations. If the illegal collection or sale of data involving trade secrets - or personal privacy - does not constitute a crime, the illegal gains should be confiscated and a fine of not less than one time but not more than ten times the illegal gains should be imposed. If there are no illegal gains, a fine of not less than 10,000 yuan but not more than 100,000 yuan should be imposed.

*Pillar III: Sharing the profits of corporate data use with the public*

Currently China does not have regulations requiring the sharing of profits from corporate data use with the public. On the contrary, businesses focused on big data receive tax incentives for relocating to specific areas of China as well as start-up financial support. For example, under the Regulations of Guizhou Province on Promoting the Development and Application of Big Data, big data enterprises that comply with the provisions of the national tax preferential policies enjoy tax incentives, and high-level talent or employees of corporations using big data who have paid the required amount of personal income tax annually will receive tax cuts. However, this may change in the future. For example, in November 2021, the former Vice-Chairman of the National People's Congress Financial and Economic Committee Huang Qifan suggested to tax 20-30% of the profits of digital platforms (Sina.com 2021).

**India**[*]

There are several ways in which data solidarity could be seen as manifested in Indian law and regulation.

---

* This section is based on contributions by Ketan Modh, whose help we gratefully acknowledge.

*Pillar I: Facilitating data use that creates significant public Value*

The relevance of the idea of solidarity-based governance in India is demonstrated by the recent surge in draft policies that seek to exploit government and private databases for public value (see Prime Minister's Office 2022). These policies include the India Data Accessibility and Use Policy, a draft of which was released in February 2022 (Ministry of Electronics and Informational Technology 2022a); the Health ID number under the Ayushman Bharat Digital Mission, for which a revised Draft Health Data Management Policy was released in April 2022 (Ministry of Health and Family Welfare and National Health Authority 2022), and the National Data Governance Framework Policy (NGDFP), a revised draft of which was released in May 2022 (Ministry of Electronics and Informational Technology 2022b). Each of these seeks to create databases that can be used for public value generation, through better health care services, incentives for Indian AI-based tech startups or the promotion of AI research. For example, the NGDFP provides for the creation of a 'non-personal data based India Datasets program and addresses the methods and rules to ensure that non-personal data and anonymised data from both Government and Private entities are safely accessible by Research and Innovation eco-system.' (Ministry of Electronics and Informational Technology 2022b: 2). This would be managed through an entirely new institution, the India Data Management Office (IDMO).

*Pillar II: Harm prevention and harm mitigation*

While India does not have an omnibus, pan-sectoral data protection regulation akin to the EU's GDPR, one is being developed. Drafted initially as the Personal Data Protection Bill, 2019, it provides an excellent example of ex post support for those who have suffered harm through an ex ante law. It contains a broad definition of harms (see Section 4) arising out of the processing of personal and non-personal data, which includes: bodily or mental injury, identity theft, financial harm, discriminatory treatment, denial or withdrawal of service and restriction of speech through the chilling effect of surveillance. As a remedy, individual data subjects have the right to seek compensation under Section 64. In its review of the draft bill, the Joint Parliamentary Committee (JPC) in its report in 2021 not only approved this measure, but also broadened the definition of harm to include 'such other harm as may be prescribed' (Lok Sabha Secretariat 2021: 54-55).

Another aspect of mitigating harm in the draft law, which overlaps with data sovereignty, is the concept of data localisation. Specifically, the law prescribes several limitations on cross-border data transfers for sensitive personal data and 'critical personal data', where the former aligns with the GDPR's idea of sensitive data. At the same time, the latter is an entirely new concept. The JPC largely agreed with the principle of data localisation, enumerating several benefits that include safeguarding privacy, employment generation and bargaining power to encourage data-based innovation – thus overlapping with the first pillar of solidarity-based governance. The abovementioned draft policies released in 2022 do not conform to the idea of solidarity-based data governance in the sense that they associate risk with data types. Specifically, each policy seeks to also minimise risk by focusing on

non-personal data or data anonymisation standards, which are not yet prescribed.

*Pillar III: Sharing the profits of corporate data use with the public*

While the draft bill and policies in India discussed above seek to enforce the creation of public value through data, they do not impose sharing obligations on the private sector, instead choosing to encourage voluntary sharing. For example, one of the roles of the IDMO is to 'enable and build the India Datasets program, which will consist of non-personal and anonymised datasets from the Government entities that have collected data from Indian citizens or those in India. Private entities will be encouraged to share such data.' The Draft Health Data Management Policy for the Ayushman Bharat Digital Mission goes even further by providing for specific procedures through which data controllers (termed 'data fiduciaries') may share de-identified or anonymised data 'for the purpose of facilitating health and clinical research, academic research, archiving, statistical analysis, policy formulation, the development and promotion of diagnostic solutions and such other purposes' (p. 25).

## 5. Policy instruments for solidarity-based data governance

This section outlines what types of policy instruments are needed to complement existing laws and policies to realise and strengthen data solidarity through the proposed governance framework. As noted above, the key aim of data solidarity is to ensure that risks and benefits in digital societies are born collectively and fairly. Acknowledging that everyone is contributing to the profits that are made on the basis of digital data, it starts with more public support for those kinds of data use that are likely to create significant public value. It also seeks to prevent harm by outlawing those practices that do not create significant public value and pose unacceptable risks. When and where harms still occur, efficient harm mitigation strategies need to be in place to support affected people and groups. And finally, as noted above, in cases where private profits are accrued by digital data, some of these need to be shared with the people, communities and societies that enabled them in the first place.

### 5.1 Pillar I: Facilitating data use that creates significant public value

The first pillar of the proposed solidarity-based data governance framework aims to facilitate data use that creates significant public value without posing unacceptable risks. The criteria for such assessment are unfortunately not straightforward. Popular definitions, such as Talbot's '[p]ublic value is what the public values' (2006) provide only minimal guidance when applied to a specific data use in question. Even seemingly clear-cut cases of publicly valuable data use – for example, the development of an application to assist physicians in the diagnosis of a disease – may entail very high risks. For instance, in 2016 DeepMind, an Alphabet subsidiary, announced a collaboration with the Royal Free London NHS Foundation Trust. The goal was to develop an application to help physicians manage acute kidney injury (Powles and Hodson 2017). To develop the app, millions of patient records were transferred from hospital trusts to DeepMind. The purpose of the data transfer was limited to the development of the application. However, over time critics voiced concerns about the possibility of DeepMind training other models with the data. They also expressed worry about the parent company linking the data with other Google products (Powles and Hodson 2017). While many may view this data use as creating significant public value without posing unacceptable risks, others view it as creating more harm than good, also because patient privacy and agency were not adequately considered. Whether this risk was so high that it precludes DeepMind's work on the kidney disease app from being classified as publicly valuable would require a more detailed assessment. To enable such a structured assessment and an approximation of the public value created by a given data use, the authors of this White Paper are currently developing a first iteration of a tool for the structured assessment of public value.

*The Finish Data Permit Authority*

An example of an institution that aims to facilitate data use that creates significant public value is Findata – Finland's authority for deciding over access requests to social and health data. It is an independent authority located in the Finnish Insti-

tute for Health and Welfare, a research and development institute operating under the Ministry of Social Affairs and Health (2022). Findata is tasked with making decisions on data permits held by a number of specified controllers, including public and private service organisers of social and health care (see Section 5). Its role is collect, combine, pre-process and then provide data for secondary purposes. The purposes for which permits may be granted are specified and include scientific research, teaching and development and innovation activities.

It is those purposes that locate (the spirit of) Findata in the first pillar of the solidarity-based data governance framework. Findata is meant to facilitate data use that creates significant public value. The authority is supposed to serve as a single interface for the applicant. Findata describes itself as a 'one-stop shop for the secondary use of social and health data' (Findata 2021). Prior to the authority's commencement of operations in 2020, entities interested in obtaining health and social data were required to contact each individual data controller with an access request. This was a de facto impediment for research as it required large amounts of time and effort on the part of researchers, especially those working for non-profit or small organisations. Findata was set up to improve this situation: Researchers should no longer have to piece together the data sets that they need from different data controllers. Instead, they should be able to turn to Findata, which would collect the data from the different controllers, bundle and anonymise the data, and then make it available to the applicant together with the permit required for use. Accordingly, the administrative burden on the applicant would be significantly lightened.

There are currently a number of issues with the implementation of the Findata's goals, including long waiting times for permits and data, and the unintended effect of increasing, rather than decreasing, the cost for some researchers who now have to pay for access to platforms for secure data analysis. While these problems pertain to the implementation, there are also a few aspects in the design of Findata that could be improved from a data solidarity perspective. For example, Findata  does not distinguish between applicants coming from the for-profit and the non-profit sector. We believe that when assessing the public value of data use, the funding of the core activities of an applicant and the primary purpose of their operations should be taken into consideration. Moreover, the data and the associated permit can only be obtained in return for payment ranging from €250 to €3,000 depending on how extensive the data permit is. This could discourage some types of applicants – especially those without deep pockets – from using the data altogether. Ideally, there should be no cost barriers at all for research uses that are likely to create significant public value without posing unacceptable risks. Last, and most importantly, Findata – and any other entity that receives or provides public funding and support for data use – should be transparent regarding the criteria it uses for their public value assessment. As we argue in this White Paper, only data use that meets the public value criteria – meaning that they are likely to benefit many people without posing unacceptable risks to individuals or groups – should receive public funding and support (El-Sayed and Prainsack 2022). It would be problematic if entities such as Findata ended up facilitating data use that serves business interests more than anyone else.

*South Africa's Information Regulator*

Another example of a policy instrument that facilitates data use that creates significant public value is the Information Regulator of the Republic of South Africa (2021). It can grant exemptions from the applicability of certain provisions of the Protection of Personal Information Act if at least one of two conditions is met. The first is that 'the public interest in the processing outweighs, to a substantial degree, interference with the privacy of' the data subject. The second is that 'the processing involves a clear benefit to the data subject or a third party that outweighs, to a substantial degree, any interference with the privacy of the data subject'. (Section 37 (1) (a) (b)).

'Public interest' in the meaning of Section 37 (1) (a) includes "the interests of national security; the prevention, detection and prosecution of offences; important economic and financial interests of a public body; [...]; historical, statistical or research activity; or the special importance of the interest in freedom of expression". Notably, public interest exemptions are two-pronged. To be granted an exemption, data processing must first be in the public interest (4.2.3.1.1 Guidance Notes), and the public interest must pass a particular threshold. The guidance notes explicitly acknowledge that what constitutes public interest cannot always be answered in the abstract and instead 'should be assessed on a case-by-case basis'. Broadly, public interest ought to benefit the public at large (instead of merely a few individuals or entities) and 'should be accepted, imposed or pursued in the spirit of equality and justice' (4.2.3.1.1 Guidance Notes). The guidance notes do not provide substantive guidance on how to interpret the concept, but they explicate the grounds on which public interest exemptions can be granted. For the grounds of exemption relating to 'interests of national security', for example, they require that '(i) National security must reflect the resolve of South Africans, as individuals and as a nation, to live as equals, to live in peace and harmony, to be free from fear and want and to seek a better life'.

Applicants fill in an Exemption Application Form and submit it to the Information Regulator online, via e-mail, by post or by hand delivery. They need to provide information about themselves, and make a case that their data use would either create substantial public value or provide clear benefits to data subjects or third persons

The two-pronged requirement for applicants to show that their instance of data processing is not only 'in the public interest' but also passes a certain threshold demonstrates that individual privacy takes primacy. The Information Regulator may also impose conditions linked to a granted exemption, such as that the data may only be processed within a limited timeframe.

The Information Regulator acknowledges the diversity and context-specificity of the concept of public interest and refrains from defining it in a seemingly universal way. Instead, it lists conditions that indicate when an instance of data processing is more likely to be in the public interest or to generate clear benefits for data subjects. This approach allows for understanding the concept in a given context

over time and paves the way to include new insights and ideas about the meaning of public interest in the future. It could provide a template on which other similar policy instruments may be based.

### 5.2 Pillar II: Harm prevention and harm mitigation

While data solidarity calls for the full force of the law to prohibit data use that poses unacceptable risks, it also recognises that whenever data is used, including with great benefit for people and communities, there is a risk that individuals or groups are harmed. Data solidarity does not deny this fact, but instead tries to ensure that people who have experienced harm receive support.

Because harm can occur without any laws being broken, or without anyone knowing what or who exactly caused the harm, existing legal redress is not sufficient (McMahon et al. 2020). In the spirit of increasing collective responsibility for the benefits and harms emerging from data use, it is necessary to improve instruments not only for risk minimisation but also for harm mitigation. This section gives examples of what harm mitigation instruments could and should look like.

*Harm Mitigation Bodies – the example of Vienna's Care and Patient Advocacy (WPPA) in Austria*

In the EU under the regime of the GDPR as well as in some countries outside the EU, independent bodies that audit data processors, such as the Information Commissioner's Office in the United Kingdom or the Commission nationale de l'informatique et des libertés in France, can currently issue monetary penalty notices or initiate prosecution. The main remit of these bodies, however, is to monitor the conduct of data controllers and not to mitigate harm inflicted on data subjects. Harms to data subjects can be of a financial, social or emotional nature, each requiring a different type of action to be mitigated, ranging from financial support to the public acknowledgement that harm has occurred and the feeding back of information to data users on how such harms could be prevented in the future.

Against this backdrop, the establishment of Harm Mitigation Bodies (HMBs) has been suggested as a solution in any legal system in the world (McMahon et al. 2020). HMBs are independent bodies that review complaints from people who claim to have been harmed by data use. Secondary data subjects – that is, those who claim to have been harmed by the use of other people's data – would be eligible to appeal to HMBs. National or supranational law could mandate that every organisation above a particular size (a 'data-using corporation') be affiliated with an HMB. Unlike using metrics such as market share, which may be difficult to calculate and ambiguous for the purpose in question, using more straightforward parameters such as annual revenue, budget, and number of staff to determine an organisation's size would ensure comparability and uniformity. Large organisations, such as multinational companies, could set up their own HMB, while smaller enterprises could submit themselves to the purview of a HMB established at the national level.

HMBs are intended to complement existing legal remedies for people alleging to be harmed by digital data use in any country in the world. Four shortcomings or peculiarities of the applicable laws in most countries demand the need for a complementation of existing legal frameworks by HMBs. First, existing legal redress mechanisms typically protect primary data subjects only; that is, the person whose personal data was used to incur harm to this same person. This, however, is not always fit for purpose. Due to the pervasiveness of digital data, harms may be systemic and affect more than one individual. For instance, a person could be harmed by a predictive analytics system that makes assumptions about them based on insights that the system obtained from other people's data (McMahon et al. 2020: 163). This person would not currently have access to legal remedies. Second, existing legal frameworks in most countries, such as traditional tort-based remedies, require individuals to prove a causal link between the breach of law by the data controller and the harm incurred. As multiple copies of digital data sets exist, it can be extremely difficult or even impossible for individuals to find out and prove which of the data controllers is responsible for the alleged harm. Third, to give an example from the European Union, provided that a data controller has demonstrably taken insufficient precautions to prevent discrimination of persons leading to an establishment of a breach of GDPR, the deterrence effect of the imposable penalties might be insignificant. This means that data controllers could accept that they harm individuals even if there are penalties because paying the penalties is still cheaper for them than stopping the practices that are cause the harm. Fourth – and we continue with the example of EU law - although Article 80 of the GDPR provides for a way for individuals to mandate a non-profit entity with the capacity to file a complaint on their behalf, this is often not easy. This is because mandated non-profit entities may only lodge complaints against harm resulting from breaches of the GDPR and not other harms. In the EU, determining the mechanisms that enable such a complaint is left to the member states. Accordingly, a plethora of possibilities and an accompanying fragmentation are to be expected (McMahon et al 2020: 164).

HMBs ought to fulfil two primary functions. First, they should be a de facto monitoring body due to the harms that people report as occurring to them. Data controllers and public agencies may then access that information to improve the operation of systems of data governance. Second, in specific cases, where people have suffered financial harm and been unable to receive support elsewhere, they should also be able to pay money to harmed parties (McMahon et al 2020: 167). Unlike in formal legal mechanisms, there should be no requirement to prove wrongdoing or direct legal causation of the harms suffered.

HMBs should be established at the national level and exercise oversight over data uses connected to individuals residing in, and data using corporations established within, the national jurisdiction. To cover the operating costs of the HMB, as well as to provide funds for financial assistance to individuals harmed by a data use, all data controllers established in a given territory should contribute a percentage of their profit, or in the case of non-profits a percentage of their funding, to the HMB.

A practical template for an HMB can be found in Vienna's health care sector (and

which has equivalents in Austria's other eight states). The Vienna Care and Patient Advocacy (Wiener Pflege-, Patientinnen- und Patientenanwaltschaft – WPPA) offers patients in public clinics as well as residents of care homes free advice and support in all matters relating to health care in Vienna. In line with the envisaged functions of HMBs for data uses, the WPPA has two main functions. First, it supports patients and residents as well as their relatives when they suspect errors in the health care they received and in asserting claims for damages out of court. Second, the WPPA fulfils a monitoring function. It receives complaints regarding the Viennese health and care sector, investigates alleged problems, addresses grievances, and offers suggestions for solutions.

Nearly all public actors involved in the health care system in Vienna are required by law to cooperate with the WPPA. For example, if a patient suspects a treatment error, they can turn to the WPPA. To establish a factual account of what happened, the WPPA will request the necessary files from the hospital or surgery in question, which are then reviewed by its in-house lawyers. If the examination shows that a treatment error may have occurred, the WPPA will try to negotiate an out-of-court settlement for the complainant. If a settlement is reached, this is where the competence and the function of the WPPA end. Complainants can, however, always initiate legal proceedings. In cases where no settlement is reached or a medical malpractice cannot be sufficiently proven, the WPPA can pay out up to €100,000 to the injured party. Exactly as is envisaged for an HMB, this financial support is not restitution or compensation and need not be in a direct relation to the harm caused.

Since the WPPA commenced its work, important lessons have emerged for the design of HMBs related to digital data. First, it has been particularly successful in identifying 'bad actors'. For example, if several patients contact it within a relatively short period with complaints about a particular physician, the WPPA is aware that there may be a problem and can act to prevent future harm. Second, much of the WPPA's success depends on the cooperation of actors in the field. Most are legally compelled to cooperate. There is a broad exemption for private practices, but most of these cooperate with it on a voluntary basis, partly to prevent reputational damage. Third, the WPPA is particularly important to individuals who do not have the financial means or access to legal recourse. It thus helps to reduce the power imbalance between patients and medical institutions.

The problems that the WPPA has encountered are also instructive for HMBs. For example, the WPPA has had difficulties getting some powerful actors, such as the Viennese Medical Chamber, to comply with its requests. For example, access to relevant files is regularly denied. Ensuring that the legal basis for an HMB stipulates that all relevant actors are obliged to cooperate with the HMB seems crucial. Ensuring consistency in the treatment of harm is also important. Each of Austria's nine states have their own patient advocacy, and there is a considerable variation in how harms are addressed and in the amounts of financial support that can be provided (for instance, in Vienna, a patient can receive up to €100,000 while the maximum is limited to €20,000 in other states). Such unjustified variation can undermine the trustworthiness of harm mitigation instruments.

*A template for harm mitigation? The complaints procedure of the Information Regulator of the Republic of South Africa*

An example of a body that currently fulfils some of the functions of an HMB is the Information Regulator of the Republic of South Africa with its complaints procedure. One of the Regulator's duties is to receive and investigate complaints about alleged violations of the Protection of Personal Information Act (POPIA) and to attempt to resolve these complaints by means of dispute resolution mechanisms such as mediation and conciliation. Not only persons who suspect that their personal information has been interfered with that can lodge complaints with the Information Regulator; so can persons acting on their behalf or even with sufficient public interest in the subject matter (Section 73). Complaints must be made in writing and can be filed in a variety of ways (including in person, online or by post) (4.2/4.3 Rules of Procedure).

Two aspects of the complaint procedure of the Information Regulator are particularly noteworthy. First, the Information Regulator can choose different ways of addressing the complaint, which makes it flexible enough to adjust to the specifics to different types of complaints and harms. In some cases, mediation may be the ideal solution for affected parties (Section 80). In other cases, an official hearing with the involvement of legal assistance and with significant public attention is suited better to solve the conflict; for example, when there is one 'bad actor' that should be singled out for its actions (Section 81). Second, the Information Regulator's competences are carefully designed to complement existing legal remedies. For example, it cannot accept complaints that are still pending before other regulatory bodies or tribunals (4.5.3 Rules of Procedure). The decisions of the Information Regulator, however, can be appealed at the High Court, which renders its complaint mechanism de facto the first level of jurisdiction (Section 97).

### 5.3 Pillar III: Sharing the profits of corporate data use with the public

The third pillar of the proposed solidarity-based data governance framework seeks to ensure that some of the profits that corporate data users make come back into the public domain. This applies particularly to data uses that do not create significant public value (Table 1, Type B of data use).

Institutionalised benefit sharing mechanisms are significantly different from corporate social responsibility (CSR) and corporate philanthropy. CSR 'occurs when a business firm consciously and deliberately acts to enhance the social well-being of those whose lives are affected by the firm's economic operations' (Frederick 2018: 3). It is usually used as a voluntary strategy to remain competitive and ultimately increase commercial profits. Corporate philanthropy comprises charitable acts by corporations, such as donations or sponsoring (Gauthier and Pache 2015). In CSR and corporate philanthropy, the corporation retains control over its activities and is not accountable to the public (beyond strategic self-reporting). Institutionalised benefit sharing with the public, on the other hand, is legally binding, entails accountability to the public and transfers control away from the corporation to the community affected by its activities.

*Taxing data and data uses*

As digital economies grow, new ways of taxation are needed to ensure that digital practices do not increase social and economic inequities within and across societies. For example, there is currently a de facto exemption from taxation for services that people buy with data instead of money. When people get access to 'free' services in exchange for their personal data, these transactions are not taxed. If people paid for these same services with money, the same transactions would be taxed (Thimmesch 2016). Another important reason for the need for fairer ways to tax corporations in the digital era is that large parts of the profits that are earned in the digital economy do not currently come back into the public domain. Finally, the COVID-19 pandemic has brought some economic sectors to a near standstill while many digital businesses have boomed. This means that tax revenue from corporate taxation shrank while profits of these digital businesses skyrocketed. There is also a global justice dimension to this issue as, for example, many businesses in the digital economy 'have a significant economic presence in African countries and benefit from economic activity but have little obligations to pay tax because they do not have a physical presence in these countries' (African Union 2020: 2).

Several solutions have been proposed. For example, a globally binding set of rules for a corporate minimum tax that could counter tax avoidance, especially regarding digital platform businesses, is being negotiated on the international level (OECD 2021b). The OECD calls international companies' practice of shifting their income to low-tax countries 'base erosion and profit shifting'. This is particularly easy for data-driven businesses, whose commodities (data) can be transferred and processed across the globe almost without constraints, irrespective of the physical location of the company. The introduction of a global minimum tax for corporations could ensure that, among other transnational businesses, digital companies and their profits generated from digital personal data will eventually be taxed to an acceptable extent – even if the amount of tax they pay is not correlated to the intensity and kind of data use. Even where data or data use is not subject to taxation, some of the value derived from it can thus be returned to the public.

Another way in which institutionalised benefit sharing could take place is in the form of data taxes. One version of this would be to impose a tax on the data itself, following the argument that data has become a new asset class 'by making user data measurable, legible, and monetizable, such as through subscriptions or selling access' (Birch, Cochrane, and Ward 2021: 13. See also World Economics Forum 2011: 5). Another approach would be to impose a tax on the data use by corporate entities that, although their data use does not pose unacceptable risks, it is not likely to create significant public value (see Table 1, Type B data use). In this way, corporations that benefit financially from such data use would pay some return on the investment that has been made by public infrastructures and other elements that have facilitated the curation, storage or sharing of data, and that also reward citizens for the risks they take by having data about themselves used by others. Revenue collected from such a data use tax could be paid into HMBs at regional and national levels, for example, so that it benefits citizens who were harmed by

data use. Alternatively, it could also be earmarked for initiatives to close digital divides or other equity-related goals.

Regarding the scenarios in which taxes would be imposed on data or data use, let us consider each of these two variants in more detail. First, for a tax on data to be effective, public authorities need to specify where and when data is to be taxed, and by whom. Data as a digital asset class is different from other products and services in that it is regularly very difficult to be located and billed, and also because data can be in several places simultaneously (Section 3). Digital businesses can host their services on servers that are located in different countries from where they provide their services. The data that these businesses collect and commodify can easily be transferred across borders without constraints.

The dissociation of the service provider and the place of consumption (De Filippi 2013: 3) on the one hand and the uncertainty about the classification of commodified data as a taxable asset on the other hand pose crucial challenges to fair data taxation. At present, it is common to consider the criteria of source and residence to determine where international businesses pay their taxes (de Man 2021: 78). This means that a company is usually obligated to pay taxes in the country where it resides; that is, where its main commercial operations are performed (de Man 2021: 94). That these criteria tend to blur in the case of data driven businesses poses yet another obstacle for data to become subject to taxation (De Filippi 2013: 4). A company could provide online services in one country but collect and process data in another. This allows data-driven businesses to avoid direct (for example, corporate income tax) and indirect taxes (for example, value added taxes) in a relatively easy way (Olbert and Spengel 2019).

As noted, a second option would be to impose taxes not on data but on data use. Here too, one of the challenges consists of making data use visible and countable. Moreover, only data use that does not create significant public value should ideally be subjected to a data tax. (Activities such as research on diseases, to decrease carbon emissions or to improve social services would typically be exempt). Individuals and small and medium-size enterprises should be exempt from such taxes to avoid that they burden small data users while powerful large ones with expensive lawyers and advisers find loopholes and other workarounds (similar to the way in which they negotiate and find ways out of taxation in other contexts at present.)

One possible way to ensure that data use by individuals as well as small and medium-size enterprises is exempt from taxation would be to impose such taxes only on data-driven businesses that handle a relatively large amount of data and create substantial network effects, and not on other corporate entities or individuals. As to determining what constitutes a data-driven business, the OECD, for example, considers the degree of mobility, the reliance on data or network effects as possible criteria (OECD 2015: 11). In another example, in 2013, France's government proposed the taxation of data at the place of interaction of a data-driven business (for example, a social network) with the persons who use their digital services, 'which is where the process of value creation actually takes place' (De Filippi 2013: 3 f). Taxation of highly diverse data streams would require regulation and technical

instruments that allow relevant public institutions to regionally manage and categorise personal data collected by data-driven businesses and enable regional taxation. This is a particular challenge given the high mobility of data and the nature and frequency of data transfers performed by data-driven businesses (see, for example, European Commission 2018 on 'Significant Digital Presence'). Although France introduced a form of digital tax in 2019, targeting advertising revenue from international digital companies that has accrued in France, it has not yet been implemented, in the hope of finding an international solution through OECD negotiations. Preceding an internationally binding solution for data taxation, a 'transitory internet tax' (De Filippi 2013: 4) could be imposed at national levels. It could be based on the collection and processing of user data that was obtained in a particular state and is being stored for a longer period of time, because, 'just like goods, data can be stored, aggregated and reused at a later time, so as to provide new value in the long term' (De Filippi 2013: 4).

As noted, data or data use taxes, as well as tax incentives given by states to multinational corporations, would probably need to be detached from the exact volume of data, data use or revenue. Debates in other policy areas such as environmental protection are currently considering charges (in the wide sense of the word) on businesses, depending on their energy intensity and the measures they take to reduce it. The aim is to create an incentive for companies to switch to more climate-friendly modes of production (see, for example, German Federal Ministry for Economic Affairs and Climate Action 2022 on the promotion of sustainable energy sources and reduction of carbon dioxide emissions). Governments can also impose specific compensations or levies to get companies 'to behave in a way that is compliant with the general interests of society' (De Filippi 2013: 5). For data-driven businesses, this could mean tax incentives for companies that comply with relevant data protection ambitions and respect their users' privacy (De Filippi 2013: 5). This could motivate data-driven businesses to support higher privacy and security standards for their users, with companies that do not offer such control functions having to pay levies.

*Benefit-sharing Agreements (BSAs)*

Similar to data taxes, benefit-sharing agreements (BSAs) can ensure that profits or benefits generated by data controllers are shared with those who made those profits possible in the first place. BSAs originate from extractive industries and are commonly used in mining, oil and gas extraction, the expansion of renewable energy sources and the use of genetic resources (Musall and Kuik 2011; Siebenmorgen and Bradshaw 2011; Tulaeva and Tysiachniouk 2017; Tysiachniouk et al. 2018). In the context of natural resource extraction, the goal of BSAs is to 'share the economic value created by extractive activities' between 'extraction companies and stakeholder communities' (Adebayo and Werker 2021: 1). Such benefits may consist of direct financial compensation – for example, through royalties – but they may also include 'preferential access to contracting opportunities for local firms' (Adebayo and Werker 2021: 1), the provision of infrastructure (for example, roads and schools) or guarantees of direct employment.

BSAs are also common in the governance of genetic resources. Article 5 of the Nagoya Protocol – the supplement to the 1992 Convention on Biological Diversity (CBD) – prescribes that 'benefits arising from the utilization of genetic resources as well as subsequent applications and commercialization shall be shared in a fair and equitable way with the Party providing such resources [...]. Such sharing shall be upon mutually agreed terms.' In fact, 'the concept of benefit-sharing originated from' the CBD (Tulaeva and Tysiachniouk 2017).

BSAs could serve as a model for policy instruments with the objective of sharing benefits derived from the use of digital data for two reasons. First, they respond to the fact that power is unevenly distributed between entities performing the (data) extraction activities and the communities whose (data) resources are extracted. This uneven distribution of power does not lead to extraction activities being prohibited, but instead a mechanism is devised that mitigates it. Second, BSAs reflect an acknowledgement that the communities from whom (data) resources are extracted play an indispensable role in the process of profit generation and as such ought to be compensated through financial or other means. These two elements can be seamlessly applied to BSAs for entities generating profits from the use of digital data.

BSAs can cover any sort of benefit (within the legal constraints) and are not limited to financial compensation as outlined above. While for the mining sector guaranteed jobs and contracting opportunities may be a suitable benefit, benefits in BSAs relating to digital data would be embedded in a different political economy and could change depending on the activities of the entity profiting from the data. For example, an entity analysing large amounts of data from which information about individuals can be extracted may be required as part of a BSA to provide the data bundled and prepared in a desired manner to an applicant free of charge. The same entity may also be required to share findings from its data analysis with the people and communities whom the data come from. For example, it could be required to tell people which of their online activities negatively influence their credit score. This is by no means an exhaustive list of benefits that could be included in a BSA and merely serves as inspiration.

There is one crucial difference between BSAs in natural resource mining and data mining that requires further consideration. In the former, the community 'affected' by the practices of extraction is typically limited; for example, a particular community next to a mining site or the community owning the land. While such narrow views of who is affected may attract justified criticisms on their own (for example, a community residing upstream may still be disadvantaged from mining activities and not be included in the BSA), defining such a community for BSAs in the digital sphere could be a difficult task that deserves dedicated analysis.

# 6. Other ways to enable and improve data solidarity

Besides the three pillars of the proposed solidarity-based data governance framework, there are other policies and strategies that can help to realise data solidarity. Two important aspects are highlighted below.

**Data solidarity and collective data ownership**

Data solidarity implies a specific take on data ownership. Ownership is not the same as property. While 'property' has clearly definable legal meaning, ownership can also refer to a moral claim on something. People who say that they own their personal data do not typically mean this to be a legal assessment. Instead, they often express that they should have a say in who uses their data, what they do with it and who benefits from it, including to ensure that their data will not be used against them. This understanding also ought to prevent free-riding. Data solidarity does not understand the availability of data as a free-for-all and grab-as-you-go. Instead, communities and other collectives should have control over the conditions of access and use; for example, a for-profit entity may be required to pay for access to data while a public research institute may not. In this sense, the underlying goods that ownership is expected to protect are privacy and dignity, both of which are not merely individual needs but also have a collective dimension.

When ownership is used to refer to property rights, then the underlying needs can still be privacy and dignity, but also the creation of financial profits. Property rights are understood as a bundle of entitlements that grant the rights holder control over the thing that they hold property rights to. The right to do whatever one wants with the object and to hinder everyone else to do the same are the two entitlements that set property rights apart from other kinds of ownership (for example, right to use). Debates on property rights to data, which have been particularly lively in recent years (see, for example, Boerding et al. 2018; Purtova 2015; Stepanov 2020), typically assume that these rights are, or should be, held by individual citizens or individual organisations. Sometimes individual property rights are claimed to offer the strongest form of individual control over data, offering better protection to individuals than privacy rights do (Hummel et al. 2021a). At other times they are explicitly justified by the desire to enable individual-level monetisation. Only if individual data subjects are seen to hold property rights to their personal data, so the argument goes, can they sell it or charge fees for its use.

From a data solidarity perspective, individual-level monetisation is a highly problematic idea (Prainsack and Forgó 2022). At a superficial level, it may seem emancipatory as it gives people the possibility to make money from their personal data (see also Hummel et al. 2021a: 557-8). Especially in societies where there are no public services to meet people's basic needs, but where needs such as health care, housing, transport, and education are to be met via the private purchase of goods and services, individual-level monetisation is likely to lead to a situation where the rich pay with money while the poor pay with data. Besides increasing social and economic inequalities, this would also mean that privacy becomes a service that

only the wealthy can afford (Elvy 2017).

Instead of endorsing individual-level property rights data, a solidarity-based perspective supports the treatment of data as collective property (see also Montgomery 2017). This means that the advantages of property rights – as the strongest level of control, including the right to sell and to destroy data – are preserved, but at the level of collectives such nations, communities, or commons (see Section 3). While individual data subjects should have meaningful control over their individual-level data wherever this helps to protect the goods of privacy and dignity (for example, individual consent to data use in the medical, insurance or consumer rights contexts), the stronger entitlements of property rights should be a collective right. Commons, communities, and nations should decide together what parts of people's lives and bodies should be datafied, and how data, once it exists, should be used and for whose benefits. This collective right to data ownership also acknowledges the large investments that societies make in enabling data work by providing education and training as well as funding the material and immaterial infrastructures that data curation and use often relies on. Although data is often considered the new oil, unlike oil it is not just 'out there' to be appropriated by companies that process and refine it. It is created by people and societies who need to have the biggest say in how data is used.

**Genuine data commons and data trusts as an alternative way of organising data practices**

Under the term 'data trust' or 'data steward', a variety of concepts are discussed that in some form or another aim at promoting data practices that are sustainable, compliant and balance the interests of diverse stakeholders (Ulmenstein 2020: 28; see also Delacroix et al. 2019 and Ruhaak 2019). A data trust is typically an entity that is supposed to facilitate wider data access and sharing, while at the same time protecting the rights of the involved parties. It thus lies at the intersection of data protection and the data economy. So far, however, no unified framework for data trusts has been established and real-world examples of possible models are still scarce (Blankertz 2020). As pointed out by Blankertz, the concept of data trust is subject to different expectations and requirements depending on the specific sector and use case (Blankertz 2020). In the health context, for example, where sensitive data is processed, an important function of data trusts is to protect the privacy of individuals when pooling and analysing patient data for medical research (Kolain et al. 2019). Data trusts carry out the pseudonymisation or anonymisation process, control the identity and access management, and facilitate the provision of consent by the data subject (Kolain et al. 2019). Data trusts in the circular economy, in contrast, typically aim to overcome the logistic and managerial challenges of sharing product data between companies (Blankertz 2020). Rath and colleagues (2016), for example, discuss the potential of data trusts to store personal data of European data subjects on behalf of US companies. By restricting the access of the US government, the data trust could provide a solution to the legal dilemma since the 2020 Schrems II judgment of the Court of Justice of the European Union, which struck down EU-US Privacy Shield.

Another important motivation for the development of the data trust model has been to strengthen the position of data subjects to counter power asymmetries and allow for participatory data governance (Ada Lovelace Institute and UK AI Council 2021). Data trusts can play a role in linking individual interests to collective governance and public value. They can also help people to communicate their privacy preferences and to enforce them against public authorities or commercial entities.

A decisive element in determining how powerful data trusts are in strengthening collective control and oversight (instead of merely functioning as a private data vault for individuals) is the extent to which individual data subjects can assign their individual rights to control their data to the data trust. Personal data or consent management systems, for instance, provide the technical infrastructure for data exchange between different parties (Mahoney, Fahs and Marti 2021) but do not represent data subjects vis-à-vis data users in any way. Rather than jointly owning and governing data, such entities are making the communication and enforcement of individual preferences more effective. Genuine data trusts, however, go beyond merely facilitating transactions between data subjects and data users. Not dissimilar to unions, who bundle the power of individual workers and obtain weight by speaking for many of them, data trusts should ideally negotiate the conditions for data processing on behalf of their members (see also Ada Lovelace Institute and UK AI Council 2021). The concept of the 'consent to be governed' introduced by the bioethicist Barbara Koenig (2014) can be helpful in this context.

**Whistle-blowers protection to prevent future harm**

As indicated earlier, harm can occur without any laws being broken (McMahon et al. 2020) or without anyone being able to pinpoint what caused the harm. But harm can also go unseen for extended periods of time. For instance, a specific group of people could be discriminated against based on their socio-economic status if a ride-sharing algorithm applied higher rates to any rides in their neighbourhood due to a higher crime rate there. Although the company's ride-sharing algorithm would not use metrics of socio-economic status directly as a category to set prices, this would de facto amount to discriminating against poor people in a neighbourhood with a higher crime rate. The customers are unlikely to discover discriminatory treatment without access to data that is severely restricted due to trade secrecy and personal data protection. In this context, whistle-blowers can reveal misconduct and prevent or mitigate harm.

Whistle-blowers are persons who report unethical or unlawful conduct obtained within a work-related context, either within the organisation concerned or outside it (for example, public authority or society at large). Whistle-blowing is considered an essential mechanism in striving for integrity and public interest (Wolfe et al. 2014), and it has brought to light some of the most controversial cases of unethical data-related practices in recent years. Examples include the US presidential election and the Brexit referendum in the United Kingdom in 2016. The stronger the protection mechanisms are, the better the collective control over the (mis)use of data. Thus, the framework for protecting whistle-blowers remains an essential

element in balancing private and public interests.

The framework for the protection of whistle-blowers shifts the burden of proof from a violation of the law to a violation of ethical principles. In doing so, whistle-blower protection contributes to ensuring that data-driven science and innovation can take place while putting adequate safeguards in place. At the same time, ensuring the effectiveness of whistle-blowers protection requires that potential whistle-blowers have access to legal counselling, as well as financial and psychological support. There is also a need for greater awareness of the political, social, economic, and ethical aspects of digital practices and the political economies that they are embedded in. This, in turn, requires covering these topics in education at all levels, and in particular within curricula and courses related to data sciences.

Whistle-blower protection remains a favourable mechanism from the perspective of trade secrecy as it assumes internal and external reporting channels. Most of the regulations implemented worldwide prioritise internal reporting mechanisms that aim to ensure confidentiality and protection from retaliation. Only if such internal mechanisms do not exist, or if they are not effective, can a potential whistle-blower report the misconduct externally – for instance, to law-enforcement authorities – or disclose the information to the general public. For example, in France, the reporting system foresees three levels. In the first instance, the problematic conduct needs to be reported within the organisation. If the issue is not adequately addressed at that level, a whistle-blower is allowed to inform public authorities or a professional association. As a last resort, if the organisation in charge has not responded to a report within three months, the report may be made public (OECD 2017). The evidence suggests that most whistle-blowers do indeed report wrongdoings within their institution or organisation before going public (OECD 2017).

International and regional standards for protecting whistle-blowers have so far been developed mainly by the OECD and the European Union. In 2019, the EU Directive 2019/1937 on the protection of whistle-blowers entered into force, aiming to harmonise fragmented regulation across the union. In many countries, there is no protection of whistle-blowers, or it is limited to certain sectors, such as banking (for example, in Australia) or public service (for example, in Chile) (Ambler 2015).

**A need for a new declaration of human rights in the digital era?**

Because ever more aspects of human lives are being captured in data, and due to the increasing power asymmetries that characterise digital societies, initiatives are currently underway to codify new rights of people in the digital era (see, for example, Cities Coalition for Digital Rights 2022; Frontiers of Tech Governance Initiative 2022; Mind Bank AI 2022; Paradigm Initiative 2022). Policymakers should consider which of these proposals should and could meaningfully be implemented via national, supranational, or international legislation. In particular, we propose the consideration of data non-discrimination legislation as analogous to genetic non-discrimination legislation issued in the last two decades (see, for example, Joly et al. 2020; Prainsack and Van Hoyweghen 2020).

# 7. Summary and recommendations

As the Governing Health Futures Report points out, it no longer makes sense to speak of 'digital health', as if there was any part of human health that is entirely independent of digital practices. Digital practices are woven into every aspect of personal, social and political life. The report's notion of digital determinants of health helpfully conveys that access to digital tools, data and information also shapes how people can protect, learn about and act upon their own health. Asymmetries of digital power and resources within and across societies are always also relevant for people's health, whether directly or indirectly (Kickbusch et al. 2021: 1732-3).

As noted in this White Paper, the increasing concentration of data and power in the hands of only a few corporate actors can lead to multiple societal harms. Individual-level control over data alone cannot improve this situation. Data solidarity seeks to complement individual-level control over data by strengthening instruments of collective oversight, control, and ownership. It seeks to balance the risks and benefits of digital practices and to support those practices that bring significant added value to societies as a whole. Next to making more effective use of the law to prevent harm, it promotes policy instruments that support people who have experienced harm as a result of data use and that ensure that commercial profits obtained on the basis of data are shared with the public.

As noted, the proposed solidarity-based data governance framework has three main pillars (Table 1 and Illustration 1). Pillar I seeks to facilitate data use that is likely to create significant value and unlikely to pose unacceptable risks of harm. While societies need to do their best to prevent harm from data use, affected people need to receive adequate support when it occurs. The instruments for providing this support need to be easy and fast to access, and they must not depend on any laws being broken, or on the harmed party being able to prove whose act or omission caused the harm (Pillar II). Pillar III aims to strengthen mechanisms of benefit sharing to ensure that some of the profits that emerge from commercial data use come back to the public domain, which has enabled the data use via the data work of individuals and public infrastructures.

Specifically, to realise data solidarity, we recommend the following:

1. Regulation should shift the emphasis away from different data types that are assumed to imply specific risks or benefits. Instead, regulation being tailored to different types of data use, based on the benefits and harms that such data use is likely to yield. (The authors of this White Paper are currently developing a tool that enables a structured assessment of the public value of data uses).

2. Data use that poses unacceptable risks to individuals or communities should be outlawed, with fines high enough to deter even large corporations and with effective enforcement mechanisms. Reliance on self-regulation by the corporate sector, as well as the assumption that 'ethics' alone can solve the problem without the help of the law, are toothless approaches

in the current political economy.

3. Data use that is likely to create significant public value without posing unacceptable risks should receive more public support than is currently the case. This can take the form of laws that set out to make easier those types of data use that create significant public value or of public services that help organisations find, access, or use data for public benefit (see Section 5.1).

4. In the digital age, harm can occur without any law being broken or without anybody being legally liable. It is often impossible for the harmed party to pinpoint who or what exactly caused the harm they experienced. For this reason, and because there can be economic or social barriers to legal remedies, support for people who experienced harm in the context of data use must be improved. They should be able to access support independent of their social or economic status, and independent of whether or not they qualify for legal remedies. As one possibility to achieve this goal, we support the introduction of Harm Mitigation Bodies (see Section 5.2).

5. Another way to prevent and to mitigate harm is through disclosures of problematic and harmful practices to the public. Whistle-blower protection in the context of data use should be strengthened.

6. In instances where data use is unlikely to create significant public value but yields commercial profits, some of these profits need to be returned to the public domain. Large data-using businesses should be required to share profits with the public; for example, through taxation or other benefit sharing requirements (see Section 5.3). Benefit sharing should help to improve global justice.

7. Individual people should never be paid for their data. This would deepen social and economic inequalities by creating societies in which the poor pay with their data for services and goods they cannot afford. Rather than treating it as (quasi-)individual property, data should be considered collective property that is governed collectively – at the level of nations or at the level of regional or even local commons. Collective ownership also enables certain types of users (for example, quasi-monopolist commercial companies or those that have broken legal or ethical rules) to be excluded from data use or to have restrictions imposed on their use.

8. Policymakers should look at current initiatives that seek to codify human rights in the digital era and deliberate which of their proposals should and could meaningfully be implemented via national, supranational, and international legislation. Specifically, we propose the consideration of data non-discrimination legislation in analogy to the genetic non-discrimination legislation issued in the last two decades (see, for example, Joly et al. 2020; Prainsack and Van Hoyweghen 2020).

As the Lancet and Financial Times Commission's Report notes, digital transforma-

tions affect all areas of life (Kickbusch et al. 2021). They reconfigure dividing lines and connections between people and societies. We believe that this work on data solidarity will help to strengthen the important agenda for research and policy set by the European Commission.

We fully endorse the principles of Indigenous data sovereignty, which share many ontological and value commitments with data solidarity. Our recommendations also correspond closely to calls for a United Nations' effort to use the data revolution for achieving the Sustainable Development Goals. The UN has recognised the unique opportunities that digital practices, and data-driven analytics specifically, bring. At the same time, it emphasises the need to adequately protect privacy and ensure the highest ethical standards of data processing (UN Development Group 2017).

# 8. Future research agenda

In terms of a future research agenda, we believe that the following activities are particularly important in the next years:

1. The policy instruments proposed in this White Paper should be implemented and evaluated, with particular attention to their effects on equity and justice, and also with regard to possible unintended consequences.

2. A systematic review of legal systems and policy approaches in all world regions should be carried out to determine (a) what laws and other policy instruments help to realise data solidarity, and with what effects so far, and (b) what laws, other policy instruments and other arrangements are the biggest impediments to justice and equity in the digital era

3. A systematic review of initiatives to codify a new set of rights in the digital era should be undertaken to see what national and international norms and agreements are under way to avoid harmful digital practice, and to hold those accountable that engage in them.

# References and appendices


Ada Lovelace Institute and UK AI Council. 2021. 'Exploring Legal Mechanisms for Data Stewardship - Data for the Public Good. Working Group - Final Report'. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf (accessed Sept 30, 2022).

Adebayo, Eric and Eric Werker. 2021. 'How Much Are Benefit-Sharing Agreements Worth to Communities Affected by Mining?' Resources Policy 71: 101970. https://doi.org/10.1016/j.resourpol.2020.101970.

African Union. 2020. 'Taxing the Digital Economy: COVID-19 Heightens Need to Expand Resource Mobilization Base'. https://au.int/sites/default/files/pressreleases/39159-pr-4th_high-level_policy_dialogue_pre-event_pr.pdf (Accessed 14 June 2022).

Ambler, Leah. 2015. 'Towards Comprehensive Public & Private Sector: Whistleblower Protection. OECD Anti-Corruption Division, St Petersburg, Conference of States Parties to the UN Convention against Corruption. Nov 2-6, 2015' https://www.oecd.org/corruption/ethics/whistleblower-protection/ (accessed Sept 30, 2022).

Autili, Marco et al. 2019. 'A Software Exoskeleton to Protect and Support Citizen's Ethics and Privacy in the Digital World'. IEEE Access 7: 62011–21.

Azaria, Danae. 2020. 'The Legal Significance of Expert Treaty Bodies Pronouncements for the Purpose of the Interpretation of Treaties', International Community Law Review, 22(1), 33-60.

Bero, Lisa. 2019. 'When Big Companies Fund Academic Research, the Truth Often Comes Last'. The Conversation. October 2, 2019. https://theconversation.com/when-big-companies-fund-academic-research-the-truth-often-comes-last-119164 (accessed January 17, 2020).

Bezuidenhout, Louise M., Sabina Leonelli, Ann H. Kelly and Brian Rappert. 2017. 'Beyond the Digital Divide: Towards a Situated Approach to Open Data'. Science and Public Policy 44(4): 464–75.

Blankertz, Aline. 2020. 'Designing Data Trusts - Why We Need to Test Consumer Data Trusts Now'. Stiftung Neue Verantwortung. https://www.stiftung-nv.de/sites/default/files/designing_data_trusts_e.pdf (accessed Sept 30, 2022).

Birch, Kean, D.T. Cochrane and Callum Ward. 2021. 'Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech'. Big Data & Society 8(1): 205395172110173.

Boerding, Andreas et al. 2018. 'Data Ownership - A Property Rights Approach from a European Perspective'. Journal of Civil Law Studies 11(2): 323–69.

Bollier, David and Silke Helfrich. 2019. Free, fair, and alive: the insurgent power of the commons. Gabriola Island, BC: New Society Publishers.

Braun, Matthias and Patrik Hummel. 2022. 'Data justice and data solidarity'. Patterns, 3(3),

p. 100427. https://doi.org/10.1016/j.patter.2021.100427.

Chapman, Audrey R. 2009. 'Towards an Understanding of the Right to Enjoy the Benefits of Scientific Progress and Its Applications'. Journal of Human Rights, 8:1, 1-36.

Carrier, James. 1991. 'Gifts, Commodities, and Social Relations: A Maussian View of Exchange'. Sociological Forum 6(1): 119–36.

Choi, Hanbyul, Jonghwa Park and Yoonhyuk Jung. 2018. 'The Role of Privacy Fatigue in Online Privacy Behavior'. Computers in Human Behavior 81: 42–51. https://doi.org/10.1016/j.chb.2017.12.001.

Cities Coalition for Digital Rights. 2022. 'Digital Rights Governance Project'. https://citiesfordigitalrights.org/digitalrightsgovernance (accessed Oct 2, 2022).

De Angelis, Massimo. 2017. Omnia Sunt Communia: On the Commons and the Transformation to Postcapitalism. London: Zed Books.

De Filippi, Primavera. 2013. 'Taxing the Cloud: Introducing a New Taxation System on Data Collection?' Internet Policy Review 2(2). https://doi.org/10.14763/2013.2.124.

Delacroix, Sylvie and Neil D. Lawrence. 2019. 'Bottom-up Data Trusts: Disturbing the "One Size Fits All" Approach to Data Governance'. International Data Privacy Law 9(4): 236-52. https://doi.org/10.1093/idpl/ipz014.

de Man, Fernando Souza. 2021. 'From BEPS Action 1 to Unified Approach: Where is the OECD Heading?' In: Kraft, Wolfgang W. and Andreas Striegel (eds). WCLF Tax und IP Gesprächsband 2019. Wiesbaden: Springer Fachmedien, 77–95. https://link.springer.com/10.1007/978-3-658-32073-7_7.

Dulong de Rosnay, Mélanie and Felix Stalder. 2020. 'Digital Commons'. Internet Policy Review 9(4). https://doi.org/10.14763/2020.4.1530.

Dixon, Pam and Robert Gellman. 2014. 'The scoring of America. How Secret Consumer Scores Threaten Your Privacy and Your Future'. World Privacy Forum. April 2, 2014. https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/ (accesses Oct 2, 2022).

El-Sayed, Seliem and Barbara Prainsack. 2022. 'The European Health Data Space fails to bridge digital divides'. BMJ 378. https://doi.org/10.1136/bmj-2022-071913.
Elungu, Alphonse. 2009. 'Article 13: Solidarity and Cooperation'. In: The UNESCO Universal Declaration on Bioethics and Human Rights: Background, principles and application, eds. Henk A. M. J. ten Have and Michèle S. Jean. Paris: UNESCO Publishing, 211–17.

Elvy, Stacy-Ann. 2017. 'Paying for privacy and the personal data economy'. Columbia Law Review 117: 1369.

European Commission. 2018. '. Proposal for a Council Directive laying down rules relating to the corporate taxation of a significant digital presence, COM(2018) 147 final'. https://ec.europa.eu/taxation_customs/system/files/2018-03/proposal_significant_digital_presence_21032018_en.pdf (accessed Oct 2, 2022).

European Commission. 2020. 'Proposal for a Regulation of the European Parliament and

of the Council on European data governance (Data Governance Act), COM(2020) 767 final'. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 (accessed Oct 2, 2022).

European Commission. 2022a. 'What Is Personal Data?' https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en (accessed Oct 2, 2022).

European Commission. 2022b. 'Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM(2022) 197 final'. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 (accessed Oct 2, 2022).

Farzanehfar, Ali, Florimond Houssiau and Yves-Alexandre de Montjoye. 2021. 'The Risk of Re-Identification Remains High Even in Country-Scale Location Datasets'. Patterns 2(3): 100204. https://doi.org/10.1016/j.patter.2021.100204.

Figueira, Rui et al. 2022. 'Guidance for Private Companies to Become Data Publishers through GBIF'. https://docs.gbif-uat.org/private-sector-data-publishing/2.0/en/guidance-for-private-companies-to-become-data-publishers-through-gbif.en.pdf (accessed Oct 2, 2022).

Findata. 2021. 'A Year in Review – Summary of Findata's Operations in 2020'. https://findata.fi/en/uutiset/a-year-in-review-summary-of-findatas-operations-in-2020/ (May 19, 2022).

First Nations Centre. 2007. OCAP: Ownership, Control, Access and Possession. Sanctioned by the First Nations Information Governance Committee. Assembly of First Nations. Ottawa: National Aboriginal Health Organization. https://icwrn.uvic.ca/wp-content/uploads/2013/08/FNC-OCAP.pdf (accessed Oct 2, 2022).

First Nations Information Governance Centre. 2014. Ownership, Control, Access and Possession (OCAP™): The Path to First Nations Information Governance. Ottawa: The First Nations Information Governance Centre. https://achh.ca/wp-content/uploads/2018/07/OCAP_FNIGC.pdf (accessed Oct 2, 2022).

Frederick, Willam C. 2018. 'Corporate Social Responsibility: From Founders to Millennials'. In: Weber, James and David M. Wasieleski (eds). Corporate Social Responsibility. First edition. Bingley, UK: Emerald Publishing Limited, 3-38.

Frontiers of Tech Governance Initiative. 2022. 'Initiate Digital Rights in Society'. https://digitalrights.ai/process-timeline/ (accessed Oct 2, 2022).

Fry, Hannah. 2019. Hello World: How to Be Human in the Age of the Machine. London: Black Swan.

Fuster Morell, Mayo. (2011). 'An Introductory Historical Contextualization of Online Creation Communities for the Building of Digital Commons: The Emergence of a Free Culture Movement'. In: Hellmann, Sebastian, Philipp Frischmuth, Sören Auer and Daniel Dietrich (eds). OKCon 2011. Open Knowledge Conference Proceedings of the 6th Open Knowledge Conference Berlin, Germany, June 30 and July 1, 2011, Volume 739 of CEUR Workshop Proceedings. Universität Leipzig, Germany and Technical University Berlin, Germany.

Gao, Henry S. 2021. 'Data Sovereignty and Trade Agreements: Three Digital Kingdoms'. SSRN Electronic Journal. https://www.ssrn.com/abstract=3940508 (October 17, 2022).

Gauthier, Arthur and Anne-Claire Pache. 2015. 'Research on Corporate Philanthropy: A Review and Assessment'. Journal of Business Ethics 126: 343-369.

German Federal Ministry for Economic Affairs and Climate Action. 2022. 'Sustainability in the Economy'. https://www.bmwk.de/Redaktion/DE/Dossier/nachhaltigkeit.html (accessed October 17, 2022).

Gray, Colin M. et al. 2021. 'Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective'. Proceedings of the 2021 CHI Conference on Human Factors in Computing System. 172: 1–18. https://dl.acm.org/doi/10.1145/3411764.3445779.

Heeks, Richard and Jaco Renken. 2018. 'Data Justice for Development: What Would It Mean?' Information Development 34(1): 90–102.

Hummel, Patrik, Matthias Braun. 2020. 'Just data? Solidarity and justice in data-driven medicine'. Life Sciences, Society and Policy 16(1): 1-18. 8 https://doi.org/10.1186/s40504-020-00101-7.

Hummel, Patrik, Matthias Braun and Peter Dabrock. 2021a. 'Own data? Ethical reflections on data ownership'. Philosophy & Technology 34(3): 545-72.

Hummel, Patrik Matthias Braun, Max Tretter and Peter Dabrock. 2021b. 'Data sovereignty: A review'. Big Data & Society 8(1): 2053951720982012.

Information Regulator (South Africa). 2021. 'Rules of procedure relating to the manner in which a complaint must be submitted and handled by the Information Regulator'. https://inforegulator.org.za/wp-content/uploads/2020/07/20211012-InfoReg-RulesOfProcedure-HandlingPOPIAcomplaints.pdf (accessed Oct 2, 2022).

Johnson, Jeffrey Alan 2016. 'The Value – and Limits – of Distributive Justice in a Justice-Centered Approach to Information Privacy'. Paper presented at Western Political Science Association 2016 Annual Meeting, San Diego, CA, 23-26 March 2016.

Joly, Yann et al. 2020. 'Looking Beyond GINA: Policy Approaches to Address Genetic Discrimination'. Annual Review of Genomics and Human Genetics 21(1): 491–507. https://doi.org/10.1146/annurev-genom-111119-011436.

Kickbusch, Ilona et al. 2021. 'The Lancet and Financial Times Commission on Governing Health Futures 2030: Growing up in a Digital World'. The Lancet 398(10312): 1727–76.

Kitchin, Rob. 2014. ‚The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences'. London: SAGE Publications Ltd. https://methods.sagepub.com/book/the-data-revolution (accessed Oct 17, 2022).

Koenig, Barbara A. 2014. 'Have we asked too much of consent?' Hastings Center Report 44(4): 33-34.

Kolain, Michael, Ramak Molavi and Philipp Otto. 2019. 'Zukunft Gesundheitsdaten—Wegweiser zu einer forschungskompatiblen elektronischen Patientenakte'. Berlin: Bundesdruckerei & iRights.Lab. https://www.bundesdruckerei-gmbh.de/files/dokumente/pdf/studie_zukunft-gesundheitsdaten.pdf (accessed Oct 18, 2022).

Kruesz, Corina and Felix Zopf. 2021. 'European Union — The Concept of Data Altruism of the Draft DGA and the GDPR: Inconsistencies and Why a Regulatory Sandbox Model May Facilitate Data Sharing in the EU'. European Data Protection Law Review 7(4): 569–79.

Krutzinna, Jenny and Luciano Floridi. 2019. 'Ethical medical data donation: a pressing issue'. In: Krutzinna, Jenny and Luciano Floridi (eds). The Ethics of Medical Data Donation. Cham, CH: Springer, 1-6.

Kukutai, Tahu and John Taylor (eds). 2016. 'Indigenous Data Sovereignty. Toward an Agenda'. Canberra: Australian National University Press.

Levin, Nadine et al. 2016. 'How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice'. Bulletin of Science, Technology & Society 36(2): 128–41. https://doi.org/10.1177/027046761666876.

Lok Sabha Secretariat. 2021. 'Report of The Joint Committee on The Personal Data Protection Bill, 2019'. http://164.100.47.193/lsscommittee/Joint%20Committee%20on%20the%20Personal%20Data%20Protection%20Bill,%202019/17_Joint_Committee_on_the_Personal_Data_Protection_Bill_2019_1.pdf (accessed Oct 2, 2022).

Luchilo, Lucas, María Guillermina D'Onofrio and María Victoria Tignino. 2018. 'Case study: The Argentine science and technology information portal'. https://community.oecd.org/servlet/JiveServlet/downloadBody/141311-102-1-248448/Argentina.pdf (accessed Oct 2, 2022).

Mahoney, Maureen, Ginny Fahs and Don Marti. 2021. 'The State of Authorized Agent Opt Outs Under the California Consumer Pirvacy Act'. https://advocacy.consumerreports.org/wp-content/uploads/2021/02/CR_AuthorizedAgentCCPA_022021_VF_.pdf (accessed Oct 2, 2021).

McDonald, Sean. 2022. 'A Digital Sovereign, by any other name'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4035822 (accessed Oct 2, 2022).

McGoey, Linsey. 2015. 'No Such Thing as a Free Gift. The Gates Foundation and the Price of Philanthropy'. London: Verso.

McMahon, Aisling, Alena Buyx and Barbara Prainsack. 2020. 'Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond'. Medical Law Review 28(1): 155–82. https://doi.org/10.1093/medlaw/fwz016.

Mind Bank AI. 2022. 'Petition: A New UN Human Rights Article That Protects Your Personal Digital Twin'. https://www.change.org/p/united-nations-human-rights-council-a-new-un-human-rights-article-that-protects-your-personal-digital-twin (accessed Oct 2, 2022).

Ministry of Electronics and Informational Technology. 2022a. 'Draft — India Data Accessibility & Use Policy 2022'. https://www.meity.gov.in/content/draft-india-data-accessibility-use-policy-2022 (accessed Oct 2, 2022).

Ministry of Electronics and Informational Technology. 2022b. 'Draft — National Data Governance Framework Policy'. https://www.meity.gov.in/content/draft-national-data-governance-framework-policy (accessed Oct 2, 2022).

Ministry of Health and Family Welfare and National Health Authority. 2022. 'Ayushman Bharat Digital Mission - Draft Health Data Management Policy, April 2022, Version 02'. https://abdm.gov.in:8081/uploads/Draft_HDM_Policy_April2022_e38c82eee5.pdf (accessed Oct 2, 2022).

Ministry of Social Affairs and Health. 2022. 'Secondary Use of Health and Social Data'. https://stm.fi/en/secondary-use-of-health-and-social-data (accessed May 16, 2022).

Mittelstadt, Brent Daniel et al. 2016. 'The Ethics of Algorithms: Mapping the Debate'. Big Data & Society 3(2): 205395171667967. https://doi.org/10.1177/2053951716679679.

Monahan, John and Jennifer L. Skeem. 2016. 'Risk Assessment in Criminal Sentencing'. Annual Review of Clinical Psychology 12(1): 489–513.

Musall, Fabian David and Onno Kuik. 2011. 'Local Acceptance of Renewable Energy—A Case Study from Southeast Germany'. Energy Policy 39(6): 3252–60. https://doi.org/10.1016/j.enpol.2011.03.017.

Montgomery, Jonathan. 2017. 'Data sharing and the idea of ownership'. The New Bioethics 23(1): 81–86. https://doi.org/10.1080/20502877.2017.1314893.

Neidhardt, Julia, Hannes Werthner and Stefan Woltran. 2022. 'It Is Simple, It Is Complicated'. In: Hannes Werthner, Erich Prem, Edward A. Lee and Carlo Ghezzi (eds). Perspectives on Digital Humanism. Cham: Springer International Publishing, 335–41. https://link.springer.com/10.1007/978-3-030-86144-5.

Nerlich, Brigitte, Sarah Hartley, Sujatha Raman and Alexander Smith (eds). 2018. 'Science and the Politics of Openness'. Manchester: Manchester University Press. http://www.manchesterhive.com/view/9781526106476/9781526106476.xml (accessed Oct 17, 2022).

Nida-Rümelin, Julian. 2022. 'Digital Humanism and the Limits of Artificial Intelligence'. In: Hannes Werthner, Erich Prem, Edward A. Lee and Carlo Ghezzi (eds). Perspectives on Digital Humanism. Cham: Springer International Publishing, 71–75. https://link.springer.com/10.1007/978-3-030-86144-5.

Nowotny, Helga. 2022. 'Digital Humanism: Navigating the Tensions Ahead'. In: Hannes Werthner, Erich Prem, Edward A. Lee and Carlo Ghezzi (eds). Perspectives on Digital Humanism. Cham: Springer International Publishing, 317–22. https://link.springer.com/10.1007/978-3-030-86144-5.

Obar, Jonathan A. and Anne Oeldorf-Hirsch. 2020. 'The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services'. Information, Communication & Society 23(1): 128–47.

OECD. 2015. 'Addressing the Tax Challenges of the Digital Economy, Action 1 - 2015 Final Report'. Paris: OECD Publishing. https://www.oecd.org/tax/addressing-the-tax-challenges-of-the-digital-economy-action-1-2015-final-report-9789264241046-en.htm (accessed Oct 17, 2022).

OECD. 2017. 'The Detection of Foreign Bribery, Chapter 2. The Role of Whistleblowers and Whistleblower Protection'. https://www.oecd.org/corruption/anti-bribery/OECD-The-Role-of-Whistleblowers-in-the-Detection-of-Foreign-Bribery.pdf (accessed Oct 2, 2022).

OECD. 2021a. 'Recommendation of the OECD Council Concerning Access to Research Data from Public Funding'. https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm (accessed Oct 2, 2022).

OECD. 2021b. 'OECD/G20 Inclusive Framework on BEPS Progress Report July 2020 - September 2021'. https://www.oecd.org/tax/beps/oecd-g20-inclusive-framework-on-beps-progress-report-july-2020-september-2021.pdf (accessed Oct 2, 2021).

Olbert, Marcel and Christoph Spengel. 2019. 'Taxation in the Digital Economy – Recent Policy Developments and the Question of Value Creation'. ZEW Discussion Paper No. 19-010. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368092 (accessed Oct 2, 2022).

Paradigm Initiative. 2022. 'Connecting African Youth with Digital Opportunities and Ensuring Digital Rights for All'. https://paradigmhq.org/ (accessed Oct 2, 2022).

PERVADE. 2017. 'Pervasive Data Ethics. We're Awake - But We're Not At the Wheel'. https://medium.com/pervade-team/were-awake-but-we-re-not-at-the-wheel-7f0a7193e9d5 (accessed Oct 2, 2022).

Phan, Thao, Jake Goldenfein, Monique Mann and Declan Kuch. 2022. 'Economies of virtue: the circulation of 'ethics' in Big Tech'. Science as culture 31(1): 121-135.

Popkin, Gabriel. 2019. 'How Scientists can Team up with big Tech'. Nature 565: 665–67. doi: https://doi.org/10.1038/d41586-019-00290-y.

Powles, Julia and Hal Hodson. 2017. 'Google DeepMind and healthcare in an age of algorithms'. Health and technology 7: 351-67.

Prainsack, Barbara. 2019. 'Logged out: Ownership, exclusion and public value in the digital data and information commons'. Big Data & Society 6(1): 2053951719829773.

Prainsack, Barbara. 2020. 'The political economy of digital data: Introduction to the special issue'. Policy Studies 41(5): 439-46.

Prainsack, Barbara and Alena Buyx. 2011. 'Solidarity: reflections on an emerging concept in bioethics'. London: Nuffield Council on Bioethics.

Prainsack, Barbara and Alena Buyx. 2016. 'Thinking ethical and regulatory frameworks in medicine from the perspective of solidarity on both sides of the Atlantic'. Theoretical Medicine and Bioethics 37(6): 489–501. https://doi.org/10.1007/s11017-016-9390-8.

Prainsack, Barbara and Alena Buyx. 2017. 'Solidarity in biomedicine and beyond'. Cambridge: Cambridge University Press.

Prainsack, Barbara and Nikolaus Forgó. 2022. 'Why paying individual people for their health data is a bad idea'. Nature Medicine 28: 1989-91. https://doi.org/10.1038/s41591-022-01955-4.

Prainsack, Barbara and Ine Van Hoyweghen. 2020. 'Shifting Solidarities: Personalisation in Insurance and Medicine'. In: Ine Van Hoyweghen, Valeria Pulignano and Gert Meyers (eds). Shifting Solidarities. Cham: Springer International Publishing, 127–51. http://link.springer.com/10.1007/978-3-030-44062-6_7.

Prime Minister's Office. 2022. 'Leveraging the Power of JAM: Jan Dhan, Aadhar and Mobile'. https://www.pmindia.gov.in/en/government_tr_rec/leveraging-the-power-of-jam-jan-dhan-aadhar-and-mobile/ (accessed Oct 2, 2022).

Purtova, Nadezhda. 2015. 'The Illusion of Personal Data as No One's Property'. Law, Innovation and Technology 7(1): 83–111.

Ragnedda, Massimo and Anna Gladkova (eds). 2020. 'Digital inequalities in the Global South'. London: Palgrave Macmillan.

Raj, Minakshi et al. 2020. 'Do People Have an Ethical Obligation to Share Their Health Information? Comparing Narratives of Altruism and Health Information Sharing in a Nationally Representative Sample'. PLoS ONE 15(12): e0244767. https://doi.org/10.1371/journal.pone.0244767.

Rath, Michael, Christian Kuß and Christoph Maiworm. 2016. 'Die neue Microsoft Cloud in Deutschland mit Datentreuhand als Schutzschild gegen NSA & Co.?'. Computer Und Recht 32(2). https://doi.org/10.9785/cr-2016-0206.

Roessler, Beate and Dorota Mokrosinska. 2013. 'Privacy and social interaction'. Philosophy & Social Criticism 39(8): 771-91. https://doi.org/10.1177/0191453713494968.

Ruhaak, Anouk. 2019. 'Data Trusts: Why, What and How'. https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34 (accessed Oct 2, 2022).

Saxinger, Gertrude and First Nation of Na-Cho Nyak Dun. 2018. 'Community Based Participatory Research as a Long-Term Process: Reflections on Becoming Partners in Understanding Social Dimensions of Mining in the Yukon'. The Northern Review 47: 187–207. https://doi.org/10.22584/nr47.2018.009.

Sharon, Tamar. 2021. 'Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers'. Ethics and Information Technology 23(1), 45-57. https://doi.org/10.1007/s10676-020-09547-x.

Siebenmorgen, Peter and Benjamin Bradshaw. 2011. 'Re-Conceiving Impact and Benefit Agreements as Instruments of Aboriginal Community Development in Northern Ontario, Canada'. OGEL 9(4): 1–23. https://www.ogel.org/article.asp?key=3142 (accessed Oct 2, 2022).

Sina.com. 2021. , ' '. https://finance.sina.com.cn/money/bank/bank_hydt/2021-10-24/doc-iktzqtyu3238774.shtml (accessed Oct 15, 2022).

Snell, Karoliina, Heta Tarkkala, and Aaro Tupasela. 2021. A solidarity paradox – welfare state data in global health data economy. Health [online first; doi:10.1177/13634593211069320]

State Council of the People's Republic of China. 2022. ' 2022 14 _ '. http://www.gov.cn/zhengce/content/2022-06/23/content_5697299.htm (accessed Oct 15 2022).

Stepanov, Ivan. 2020. 'Introducing a Property Right over Data in the EU: The Data Producer's Right – an Evaluation'. International Review of Law, Computers & Technology 34(1): 65–86.

Szoszkiewicz, Lukasz. 2021. 'Open Data: Toward Achieving and Measuring the Sustainable Development Goals'. In: Walter Leal Filho et al. (eds). Industry, Innovation and Infrastructure, Encyclopedia of the UN Sustainable Development Goals. Cham: Springer International Publishing, 1–11. https://link.springer.com/10.1007/978-3-319-71059-4_129-1.

Talbot, Colin. 2006. 'Paradoxes and prospects of 'Public Value''. Paper presented at Tenth International Research Symposium on Public Management, Glasgow.

Taylor, Linnet. 2017. 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally'. Big Data & Society 4(2): 205395171773633.

Taylor, Mark. 2012. 'Genetic Data and the Law: A Critical Perspective on Privacy Protection'. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511910128.

TEHDAS. 2021. 'Presentation of a First Set of Data Altruism Definitions, Use Cases and Findings'. https://tehdas.eu/app/uploads/2021/09/tehdas-presentation-of-a-first-set-of-data-altruism-definitions-use-cases-and-findings.pdf (accessed Oct 2, 2022).

Thimmesch, Adam B. 2016. 'Transacting in Data: Tax, Privacy, and the New Economy'. Denver Law Review 94(1): 145.

Tulaeva, Svetlana and Maria Tysiachniouk. 2017. 'Benefit-sharing arrangements between oil companies and indigenous people in Russian Northern regions'. Sustainability 9(8): 1326.

Tysiachniouk, Maria, Laura A. Henry, Machiel Lamers and Jan P.M. van Tatenhove. 2018. 'Oil and Indigenous People in Sub-Arctic Russia: Rethinking Equity and Governance in Benefit Sharing Agreements'. Energy Research & Social Science 37: 140–52. https://doi.org/10.1016/j.erss.2017.09.004.

Ulmenstein, Ulrich von. 2020. 'Datensouveränität durch repräsentative Rechtswahrnehmung. Begriffliche Prägung und normative Gestaltung sogenannter „Datentreuhänder"' (‚Data souvereignty through representative rights management. Conceptual determination and normative design of so-called data fiduciaries'). Datenschutz und Datensicherheit 44: 528–34 (in German).

UNESCO. 2021a. 'UNESCO Recommendation on Open Science'. Paris: United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en (accessed Oct 2, 2022).

UNESCO. 2021b. 'Recommendation on the Ethics of Artificial Intelligence'. Paris: United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed Oct 17, 2022).

United Nations Conference on Trade and Development (UNCTAD). 2021. 'Data Protection and Privacy Legislation Worldwide'. https://unctad.org/page/data-protection-and-privacy-legislation-worldwide (accessed Oct 2, 2022).

United Nations Development Group. 2017. 'Data Privacy, Ethics and Protection. Guidance Note on Big Data for Achievement of the 2030 Agenda'. https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf (accessed May 30, 2022).

United Nations General Assembly. 2007. 'Resolution adopted by the General Assembly on 13 September 2007, Res. 61/295. United Nations Declaration on the Rights of Indigenous Peoples'. https://legal.un.org/avl/pdf/ha/ga_61-295/ga_61-295_ph_e.pdf (accessed Oct 17, 2022).

Van Dijck, José, Thomas Poell and Martijn De Waal. 2018. 'The Platform Society: Public Values in a Connective World'. Oxford: Oxford University Press.

Vienna Manifesto on Digital Humanism. 2019. https://dighum.ec.tuwien.ac.at/dighum-manifesto/ (accessed Oct 2, 2022).

Vohs, Kathleen D. et al. 2008. 'Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiative'. Journal of Personality and Social Psychology 94(5): 883–98.

Werthner, Hannes, Erich Prem, Edward A. Lee and Carlo Ghezzi (eds). 2022. 'Perspectives on Digital Humanism'. Cham: Springer International Publishing. https://link.springer.com/10.1007/978-3-030-86144-5.

Wolfe, Simon, Mark Worth, Suelette Dreyfus and A J Brown. 2014. 'Whistleblower protection laws in G20 countries : priorities for action'. Blueprint for Free Speech, The University of Melbourne, Griffith University, Transparency International Australia. https://catalogue.nla.gov.au/Record/6608111 (accessed Oct 18, 2022).

Wolkovich, Elizabeth. M., James Regetz and Mary I. O'Connor. 2012. 'Advances in global change research require open science by individual researchers'. Global Change Biology 18(7): 2102-2110. https://doi.org/10.1111/j.1365-2486.2012.02693.x.

World Economic Forum. 2011. 'Personal Data: The Emergence of a New Asset Class'. https://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf (accessed Oct 2, 2022).

World Health Organization (WHO), Regional Office for South-East Asia. 2017. 'Universal health coverage technical brief: data exclusivity and other 'TRIPS-plus' measures'. https://apps.who.int/iris/bitstream/handle/10665/272979/Data-exclusivity.pdf?sequence=1&isAllowed=y (accessed Oct 2, 2022).

Young, Iris Marion. 1990. 'Justice and the Politics of Difference'. Princeton, NJ: Princeton University Press.

## Glossary of Key Terms

**Datafication**
Datafication refers to the process by which different kinds of subjects, objects and practices are transformed into digital data.

**Digital and data commons**
Commons are social institutions for governing common-pool resources. Digital and data commons are approaches to collaboratively govern resources – such as data, information, culture and knowledge – that have been created by communities, according to the principles of fairness, equality, justice, and sustainability.

**Data harm**
Harm to data subjects can be of a financial, social or emotional nature. It can occur without any laws being broken, or without anyone knowing what or who exactly caused the harm. In the case of large amounts of personal data being analysed, harms may be systemic, affect more than one individual and pose a threat to societies as a whole.

**Data trusts**
A data trust is typically an entity that is supposed to facilitate wider data access and sharing, while at the same time protecting the rights of the involved parties. Data trusts can also strengthen the position of data subjects to counter power asymmetries and allow for participatory data governance. They can also play a role in linking individual interests to collective governance and public value.

**Digital economies**
The OECD defines the digital economy as an economic sector characterised by high mobility, reliance on data, network effects, a wide range of products and services on offer, a tendency towards monopoly or oligopoly and volatility (OECD 2015). The types of business models include e-commerce, app stores, online advertising, cloud computing, participative networked platforms, high-speed trading and online payment services.

**Personal and non-personal data**
The European Commission defines personal data as any information that relates to an identified or identifiable living individual (European Commission 2022a). Different pieces of information that collected together can lead to the identification of a particular person also constitute personal data. Furthermore, personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the GDPR.
Non-personal data does not contain any direct information about or relating to a person who is directly or indirectly identifiable. This can be geographical data, climate data or higher-level economic data.

**Acknowledgements**

We thank The Lancet and Financial Times Commission on 'Governing health futures 2030: Growing up in a digital world' for giving us the opportunity to write this paper, and for their inspiration and helpful feedback. We are grateful for their financial support for research assistance.

We also thank the following individuals who have contributed to individual sections, or supported this White Paper in other substantive ways: Matthias Braun, Henry Gao, Theresa Henne, Ivan Josipovic, Shi Li, Katja Mayer, Gert Meyers, and Ketan Modh.

We are grateful to Heidrun Åm, Tima Otu Anwana, Peter Biegelbauer, Paul Dunshirn, Ulrike Felt, Carrie Friese, Hanna Kienzler, Jenny Krutzinna, Migle Laukyte, Gertrude Saxinger, Lukas Schlögl, Karoliina Snell, Elisabeth Steindl, Sandra Wachter, and Alice Vadrot, who provided helpful comments on various versions of this White Paper.

The work on this paper was made possible by the Research Platform Governance of Digital Practices at the University of Vienna, Austria.