Data Governance Network
Anchored by IDFC Institute

January 2022

Working Paper 23

Governing the Resource of Data: To What End and for Whom?

Conceptual Building Blocks of a Semi-Commons Approach

Anita Gurumurthy and Nandini Chami

IT for Change

Data Governance Network

The Data Governance Network is developing a multi-disciplinary community of researchers tackling India's next policy frontiers: data-enabled policymaking and the digital economy. At DGN, we work to cultivate and communicate research stemming from diverse viewpoints on market regulation, information privacy and digital rights. Our hope is to generate balanced and networked perspectives on data governance — thereby helping governments make smart policy choices which advance the empowerment and protection of individuals in today's data-rich environment.

About Us

IT for Change aims for a society in which digital technologies contribute to human rights, social justice and equity. Our work in the areas of education, gender, governance, community informatics and internet/digital policies push the boundaries of existing vocabulary and practice, exploring new development and social change frameworks.

Disclaimer and Terms of Use

The views and opinions expressed in this paper are those of the authors and do not necessarily represent those of the organisation.

Design

Cactus Communications

Suggested Citation:

Gurumurthy, A., Chami, N. (2022). Governing the Resource of Data: To what end and for whom?, Data Governance Network Working Paper 23.

Supported by a grant from Omidyar Network

Abstract

This paper presents the building blocks of an economic governance regime for the social commons of data that can contribute to decentralizing the benefits of the digital economy. The enclosure of data by platform monopolies not only creates a skewed, exclusionary marketplace, but also represents a huge opportunity cost in terms of the unrealized public and social value of data. Emerging approaches to data governance – both, individualist and collective – do not go far enough in addressing this challenge. This paper argues that only a semi-commons approach can effectively reorder the digital economy to achieve a much needed distributive integrity.

Table of Contents

1. Introduction

> "Nature builds no machines, no locomotives, railways, electric telegraphs, self-acting mules etc. These are products of human industry...the power of knowledge, objectified. The development of fixed capital indicates to what degree general social knowledge has become a direct force of production, and to what degree, hence, **the conditions of the process of social life itself have come under the control of the general intellect and been transformed in accordance with it.**" (emphasis ours).
>
> - Karl Marx, The Fragment on Machines segment of Grundrisse (1857-61)

More than 160 years after Marx's conception of the "control of the general intellect" as capital's modus operandi, the subsumption of social life into capital stands at a crossroads. With the inexorable march of technology, data emerges as the economic object par excellence for capital to refine its pursuit of social control.

The advent of the data revolution has triggered wide-ranging debates. If and how society's brand new general intellect in the form of data's intelligence will transform the question of value creation and distribution, and how social and public value can be contemplated afresh, is at the core of these discussions. Data's unique character has moved these debates into the arenas of policy and law, reflecting the urgent need for a paradigm shift in the governance of the economic sphere.

As abstract social knowledge from data becomes a force of production, choices about its resource governance will shape our collective social future. From this starting point, our paper argues how a sui generis semi-commons approach to data governance, grounded in the norm of distributive integrity, is necessary for evolving a new global digital economic paradigm centered on human flourishing.

To begin with, Section 2 of the paper reflects on the modus operandi of data capitalism. Although data is born social, the de facto ownership of data and ensuing monopolization of its intelligence advantage by lead platform firms suggests a lawlessness in the digital economy, exacerbating inequality on a planetary scale. Section 3 highlights how the lack of effective resource governance approaches to data has contributed to this crisis. Sub-section 3.1 traces the roots of this problem to the dominant, individualist approach pioneered by the European Union (EU) that has disregarded the question of economic rights in data. Sub-section 3.2 shows how alternative data stewardship approaches – anchored as they may be in collective control – end up recasting social data into a 'pro-capitalist' commons. Sub-section 3.3 examines the emerging community data approach in India, suggesting that its vision of community control over material resources needs to engage further with questions of how data value can be democratized. Subsection 3.4 underscores the urgency of a governance framework for data, calling attention to the rampant misrecognition and maldistribution in the current economic paradigm.

Having established the details of the crisis at hand, Section 4 proceeds to articulate how to effectively govern data for distributive integrity. Subsections 4.1 and 4.2 lay out the building blocks of a semi-commons governance approach with a normative baseline in the 'freedom of open use' in data resources, proposing a rights-based resource ownership regime with varying degrees of differentiated access rights and associated conditionalities for economic actors across the spectrum. Subsection 4.3 reflects on some considerations that should inform the institutional design of the data semi-commons model to embed it in a whole-of-economy approach for distributive integrity. The paper concludes with brief reflections on the need for further conceptual work.

2. Data's subsumption into capital's circuit of control

At the turn of the millennium, techno-optimist readings of the internet presented new visions for reorganizing the social relations of production. The techno-social possibilities for commons-based peer production (Benkler 2006) seemed highly plausible, signaling the prospect of new pathways to economic justice. The new communitarian technologies of Silicon Valley were also seen as blending market and non-market forms of production in the best possible way, rendering the classical polarities of capitalism and socialism irrelevant in organizing the economy (Kelly 2009, Shirky 2009). Techno-optimists believed that the internet, as a technology of commonsification, would usher in the end of post-Fordist capitalism and its enclosure of knowledge rents, and culminate in an abundance of information and knowledge (Peters 2019).

By the late 2000s, it was impossible to ignore the writing on the wall. If the original internet was basically a democratic, pluralist, and decentralized web, the new internet was increasingly assuming a hierarchical, centralized character. The frenetic increase in computing power and exponential growth of digital data generated – either directly in virtual space or indirectly in physical space – has since seen 'platform capitalism' emerge as the grand narrative of our times. The rapid consolidation of market power in the hands of winner-take-all platform behemoths – the "capitalist platform firm" (Srnicek 2016) – marks the platformization of production and market exchange based on capitalist control of "network-data architectures" (Gurumurthy et al. 2019).

The wealth accumulation strategy of platforms is based on the extraction of network monopoly rents and algorithmic rents (Mazzucato, Entsminger, and Kattel 2020). Platforms do not just leverage network effects in growing their user base. They also build an intelligence advantage through incessant algorithmic mining of data-based insights in the networks they control, by hyper-optimizing network interconnections. Data accumulation plays a central role in the monopoly power of platforms – to expand their reach and create new markets where none existed previously (Srnicek 2016, Gurumurthy et al 2019).

Data-based insight generation under digital capitalism[1] is vastly different from the pre-digital knowledge paradigm. Data as a form of discrete pieces of digital information is easily agglomerated and highly mobile (Sadowski 2019). The data paradigm[2] can be viewed as a breakdown of the distinction between the level of individual component and that of aggregated structure, whereby every "whole is always smaller that the sum of its (mobile) parts" (Latour et al, 2012). The intelligence advantage in capitalist platform ecosystems, therefore, depends on two premises – seamless data hoarding and continuous data extraction. To put it differently, value propositions in the data-led digital economy revolve[3] around the accumulation of data for its potential, rather than immediate, use value (Coyle et al 2020, UNCTAD 2019). Also, the immediate use of a data point or any single collection is less important than the unceasing flow of data creation (Sadowski 2019).

Individual sources of data often have considerable option value,[4] and could potentially become valuable, if new questions not yet thought of can be answered in the future (Coyle et al 2020). Hence, the *raison d'etre* of digital capitalism is to increase the option value of data sets, aggregate, and recombine data in a

[1] The terms 'digital capitalism' and 'data capitalism' are used interchangeably in the paper.

[2] Suggesting the paradigm shift towards Big Data and associated technologies.

[3] The terms 'data economy' and 'digital economy' are used interchangeably in this paper.

[4] The new potential uses and services and the associated future income streams that are not possible to anticipate at the point of collection of data.

myriad ways to increase the possibility of generating analytical insights for a breadth of applications. Correspondingly, with developments in mobile, Internet of Things(IoT), and related technologies, data accumulation in the digital economy extends beyond the web of advertising into a range of environments – physical and social, and the human interactions therein – through smart connected objects. Increasingly, such accumulation comprises the basis of production and market exchange activities at a whole-of-economy level, no longer confined to the internet.

The unremitting flows of data into ever-expanding platform enclosures represent a crisis for value creation and distribution in the data economy. First-mover platform firms have unbridled, unilateral power to exclude all other parties from accessing and using the data resources they control. Given that data is a social resource, born of social relationalities (Viljoen 2020; Taylor and Purtova 2019), the de facto ownership of such data (Fia 2020), and the ensuing monopolization of its intelligence advantage by lead platform firms reveals a lawlessness in the digital economy.

A digital wild west arising in the absence of a resource governance regime for data (Purtova 2017) is exploitative, and hence, unsustainable, for several reasons:

a. It propels a global economic order that is a far cry from the original promise of an egalitarian internet economy contained in the techno-materiality of network-data resources.

b. The control of data by lead digital firms creates a skewed political economy of digital development in which developing countries cannot build their competitive advantage, thus obscuring the right of people and communities in these countries to access, use, and benefit from their own data and intelligence resources.

c. The digital economy as we know it, forestalls the economic prospects of smaller businesses and relevant data communities, locking up data value in the silos of rentier capitalism, and preventing positive externalities in the form of data's public and social value from being realized.

Recent policy developments calling for regulation of Big Tech and data sovereignty (mostly at the EU level), do mark a fork in the road. However, as argued in the next section, they fall short of addressing the excesses of digital capitalism, failing to grapple with the economic governance of data in relation to its unique properties.

3. Current approaches to data governance – why they fall short

A quick scan of the current data governance landscape reveals that there are three prominent strands in the debate. The 'breaking up Big Tech' approach has gained traction primarily in the United States, which lacks both a robust data privacy legislation and a competition law framework that can effectively challenge market dominance of lead firms. The criterion of business size in the proposition to break up Big Tech, however, is unlikely to make a dent on the enclosure of the 'social data commons' (Mazzucato 2019). Such measures may just end up replacing a platform monopoly with a duopoly/oligopoly (ibid).

The other two strands – a dominant, individualist approach pioneered by the EU and a more nascent, alternative approach that is centered on evolving a set of collective controls over data – aim to fix the unchecked expropriation of social relationalities for capitalist accumulation. We examine both these approaches below.

3.1 The individualist approach to data governance

The EU may be seen as a pioneer of the individualist approach to data governance that has caught on in many other parts of the globe. In this approach,[5] individual subjects have quasi-ownership rights in their personal data. This includes the right to determine if, and on what terms, their personal data enters the data market through a notice-and-consent regime, within the boundaries for the data market[6] specified by personal data protection legislation (Viljoen 2020). Personal data that is anonymized and machine-observed data that does not have personal identifiers at the point of collection are treated as alienable 'non-personal data', whose free and unrestricted flows as an economic object must be maximized for the development of the data market (European Union 2018). In this approach, except in the case of willful/inadvertent deanonymization that reveals personal identifiers, there are no claims that citizens can make on data processors with respect to non-personal data processing. More importantly, non-personal data is treated implicitly and automatically as the private property of data processors. The question of the economic claims of citizens in the data value generated from their anonymized personal data or *their* data footprints in machine-observed data is completely sidestepped.

This approach suffers from the following shortcomings:

a. Inattention to privacy risks arising from processing of non-personal data

The assumption that all privacy risks pertain only to personal data processing is deeply flawed. In reality, risks to privacy as informational self-determination are often visibilized only in higher order processing in the data value chain, and not at the lower levels of data collection. For instance, at first glance, it may appear that in the observed data points collected by smart energy systems, temperature, light, and motion sensors, there is nothing relevant to privacy risks as they otherwise arise in the collection of personal data. However, as these machine-observed data points move up the data value chain, they hold the latent potential for smart home manufacturers to infer quite a lot of socio-behavioral insight that can profile individual households when clubbed with other data sets (Nissenbaum 2019).

Importantly, data-based profiling may occur with the mixing of machine-observed data with personal information, even without lapses in anonymization, thus posing regulatory challenges to prevent collective harm.

b. Failure to address market fairness in the data economy

The lack of a clear framework outlining economic claims in non-personal data is accentuated because of the increasing smartification of real economy value chains. As more and more production chains in agriculture and manufacturing are transformed because of the deployment of smart connected objects, questions for data value sharing between traditional and new age data firms in the economy arise. As discussed, these are over and above older questions about the capture of value from data mining in social interactions from advertising platforms and the consumer Internet of Things.

[5] As in the case of the EU's General Data Protection Regulation.

[6] Including obligations of data processors.

When it comes to the datafied real economy value chains in which traditional and new age firms participate, the lion's share of value steadily accrues to the firms with data power – that is, firms producing the "sensors, processors, embedded software, data storage systems, and automated services" (Yu 2019). There is also a concern that even among digital firms, dominant players will consolidate their market advantage by leveraging the monopoly power of data enclosures, resulting in a market loaded against new entrants.

Within the confines of the overall individualist approach, solutions have been proposed/deliberated upon by policymakers in the EU to restore fairness in the data market by opening up private data enclosures. But these measures have not really succeeded, as discussed below.

The main policy debate has focused on using competition law as an instrument to effectively challenge the anti-competitive advantage of firms enclosing data on a massive scale. Competition authorities in the EU have been exploring the extension of Article 102 of the Treaty on the Functioning of the EU, a legal provision intended to regulate monopolies. The proposal seeks to update the Treaty's 'essential facilities doctrine'[7] and open up access to the aggregate non-personal data held by dominant firms to downstream activities in the data value chain with its numerous "aftermarkets"[8] (Cremer et al. 2019). However, these efforts may not go far in overturning the *de facto* property claims of platform companies, and establishing non-exclusive access to data as the new norm, because of the limits of competition law in its granting of defensive and circumstantial, rather than categorical or abstract, access claims, to address market abuse. Competition law does not justify a claim to data access on the basis that such a claim promotes market competition. Rather, it permits access where the dismissal of such a claim by the data holder is seen to result in a restriction of competition that is not otherwise justifiable (Ullrich 2019).

Another related discussion has been around addressing the disadvantage that traditional firms face when negotiating with digital companies in 'smart' value chains in the real economy, by introducing a new data producer's right for non-personal, anonymized machine-generated data (Yu 2019). As elucidated in the EU's Communication on Building a European Data Economy (2017):

> A right to use and authorise the use of non-personal data could be granted to the "data producer", i.e. the owner or long-term user (i.e. the lessee) of the device. This approach would aim at clarifying the legal situation and giving more choice to the data producer, by opening up the possibility for users to utilise their data and thereby contribute to unlocking machine-generated data. However, the relevant exceptions would need to be clearly specified, in particular the provision of non-exclusive access to the data by the manufacturer or by public authorities, for example for traffic management or environmental reasons. Where personal data are concerned, the individual will retain his right to withdraw his consent at any time after authorising the use.

The European Strategy for Data (2020) makes no mention of this proposal. Even so, this proposal suffers from two critical drawbacks. First, even though the right has been conceptualized as a possession right rather than an ownership right, it allows right-holders to assert a private claim to excludability (the legal right to unilaterally exclude others) from the data they hold by controlling downstream uses and

[7] The Essential Facilities Doctrine is a rule originally developed to decide under which conditions the denial of access to infrastructures should be considered anti-competitive. Under the test proposed by the Doctrine, a company with a dominant position in the provision of a facility, product, or service, which is indispensable to compete in a downstream market is determined as abusing its dominant position if, without objective justification, it refuses to grant access to this facility, product, or service.

[8] Markets that are part of the broader ecosystem served by the firm that is the data controller.

bringing injunctions against unauthorized uses (Stepanov 2019). Second, the collective benefits of a social resource like data may not materialize merely with a complex mesh of private property claims. In order to ensure that an economic regime does not perpetuate inequality, it is critical to evaluate whose property rights are secured through legal frameworks (Lawson-Remer 2012). The idea of a data producer right fails this test as it strengthens the claims of upstream data producers against that of newer entrants.

To summarize, an individualist governance regime for data is predicated on the separation of data governance considerations into two water-tight compartments; one that preserves the inalienability of personal information as an extension of the self, and the other that enables the existence of a data market that increasingly relies on mixed data sets of aggregate personal and non-personal data. This 'solution' treads on uneasy ground on two counts: it fails to prevent the moral inappropriateness of converting inalienable personal information into an alienable economic object (Prainsack 2019), and leaves the exclusive control that data collectors have over the social data they hoard untouched.

3.2 Collectivist approaches to data governance

A fledgling narrative on unlocking data value for society is emerging in policy and civil society circles. A closer examination reveals that there are two distinct threads in this alternative approach: a data stewardship narrative that has caught the imagination of policymakers and thinktanks in the Global North, and a community data conception emerging out of Indian policy developments.

Data stewardship, simply put, refers to any institutional arrangement where a group of people come together to pool their data and put in place a collective governance process for determining who has access to this data, under what conditions, and to whose benefit (Hardinges 2020). The institutional arrangement may take a range of specific forms: a data cooperative where pooled data is co-owned and democratically controlled by its members using decision-making processes modeled after traditional cooperatives; a data trust where a trusteeship mechanism is put in place for the fiduciary management of a group's data; or a data collaborative where there is a public private partnership for the pooling of private sector data to aid governance decision-making (van Geuns and Branducescu 2020). The EU's draft Data Governance Act (2020) seeks to facilitate the establishment of "data altruism organisations" to enable the pooling of non-personal data for non-profit, "general interest" purposes. The hope is that such registered special purpose entities can play the role of data stewards to shepherd data-based innovations for social good. Similarly, the EU is also exploring the means to encourage business-to-government data sharing through data stewardship mechanisms (EU High-Level Expert Group, 2020).

The notion of stewardship has been received positively across the ideological spectrum. However, even as stewardship models seek to target corporate data extractivism, they will, most likely, end up as a device for large data monopolies to externalize their regulatory burden, reducing administrative costs and reputational risks in the process of data collection and processing (Mills 2020). They also legitimize data enclosures by implicitly endorsing private companies' *de facto* ownership rights in the data they have collected and aggregated.

Data stewardship is also reduced to an ethics-washing strategy in the hands of digital corporations and their lobbies. For instance, in the case of the now-scrapped Toronto Waterfront Smart City project, Google-owned Sidewalk Labs had proposed an urban data trust mechanism based on data stewardship principles for governing the project's data. But as the then Information and Privacy Commissioner Brian Beamish (2019) pointedly submitted in a 2019 letter to the Waterfront Board, the urban data trust would end up becoming a privatized mechanism with absolutely no accountability either to privacy regulators or city authorities.

More recently, in December 2020, the World Economic Forum (WEF) announced the Data for Common Purpose initiative – a global multistakeholder endeavor that seeks to co-design, pilot, and scale flexible data governance mechanisms. It aims to unlock data from existing silos and create opportunities for both the public good and commercial benefit. The initiative will create government-backed ethical data marketplaces that enable the equitable sharing of benefits in data through non-personal data exchange, with appropriate valuation based on the level of processing (World Economic Forum 2021). As an effort at data stewardship, the initiative is an attempt at transparent rule setting for trade in data and clarifying private property claims (from the data originator to the initial data collector and firms who have acquired the data from collectors and done some level of processing). The assumption here is that enabling a data market bounded by a privacy ethics framework can catalyze data-based value creation and remedy the sub-optimal use of data resources. While a transparency-based exchange may be a welcome departure from the dark markets in which personal data is widely traded today for modest sums, the viability of the effort in unlocking data's positive externalities remain to be seen. Public value creation through data depends on a host of factors, including the ability of state agencies to benefit from such collaborations. Most public agencies in the Global South are unfortunately placed at a relative disadvantage in this regard with their sub-par data infrastructure. Transparency in data exchange also does not automatically incentivize smaller players in the data field. In fact, it often ends up giving firms who already squat over large data sets an advantage in such collective experiments.

In the final analysis, across different empirical settings, the idea of data stewardship primarily corresponds to the creation of what the political philosopher George Caffentzis (2010) terms "a pro-capitalist commons". Data market propositions that convert the social relational resource of data for capital accumulation do not address the unequal ownership of data as a means of production.

Self-organized data communities may, at best, represent enclaves of powerful, alternate visions,[9] more idealistic than pragmatic. At worst, they hijack the virtues of cooperativism for underwriting capitalist appropriation of value.

3.3 The 'community data' approach

We turn to another collectivist approach emerging in Indian policy circles: the 'community data' approach. In this view, data resources are seen as akin to natural resources in that both are part of "a nation's or community's collective resources as arising from their natural and/or social spaces, and should be governed as such" (Committee of Experts on Non-Personal Data Governance Framework, 2020). Permanent sovereignty over natural resources has been acknowledged as a "basic constituent of the right to self-determination" in UN General Assembly Resolution 1803 (Alam and Faruque 2019). Community resource governance frameworks for the natural resource commons, such as the Nagoya Protocol[10] of the Convention on Biological Diversity, are seen as the foundations for evolving a new data governance regime that prevents the enclosure of valuable socio-behavioral datasets and intelligence about communities by private corporations, and unlocking them for the common good.

[9] Take, for instance, 'The Distributed Cooperative Organisation' framework proposed by progressive organizations in the EU that seeks to leverage innovations in the commons, P2P, open co-operativism, and other digital alternatives for evolving new enterprise models that produce value in socially sustainable ways. The sustainability of such frameworks hinges on appropriate public support in the form of tax incentives, community spaces, commons-public partnerships and co-operative development funds that may be lacking in developing countries.

[10] on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization.

The Indian approach moves the needle in the data governance debate by acknowledging data's social moorings, thereby anchoring data rights in its associated communities, and also identifying data-enabled value creation as a vital national public policy issue. It provides a fresh starting point to solving the dilemma of how best to account for collective claims in shared data resources, an area that EU policymakers attempting to move out of the individualist approaches to data governance are yet to grapple with fully. Take the case of the EU's proposed Digital Markets Act (2020) that attempts to conceptualize a unique data access right for platform-dependent business users. Under this Act, business users have "effective, high-quality, continuous and real-time access and use of aggregated or non-aggregated data, that is provided for or generated in the context of the use of the relevant core platform services by those business users and the end users engaging with products or services provided by those business users". However, the legislation is unable to clarify the nature of how such data access claims are to be operationalized – especially in relation to business users' collective rights in aggregate data (Singh and Gurumurthy 2021).

In this context, the Indian policy proposal for actualizing data's potential for the economy through principles of community access, use, and benefit sharing – enshrined in constitutional principles for the fair distribution of a community's material resources[11] (Mittal 2020) – comprises a useful normative compass for devising any economic governance regime for data. However, the institutional design of a governance regime for data as a societal commons cannot be modeled exactly after the natural resource commons. This is because of the differences in the nature of the two resources, as elaborated below.

Unlike natural resources, data resources do not have clear boundaries – "digital data is multiple in that it can be in several places at the same time" (Prainsack 2019). The same piece of location data, for instance, can be in multiple data sets. What this means is that it is in the creation of a specific data pool that a data community takes shape, making it difficult to establish data communities a priori. To make things more complicated, even those from whom data is not collected may be impacted by the use of data-based digital intelligence in various settings. In other words, not only are we all contributors to multiple data communities at the same time, we may also find ourselves in the target community of data-based businesses even if we are not part of the source community (the group from whom data was initially compiled). This raises complex issues for the rules of exclusion-inclusion and the evolution of representative decision-making mechanisms in the institutional mechanisms of data governance. Mechanisms for excludability of 'outsiders' that communities are able to implement in the instance of Common Property Resources (CPR), such as forests, are not easily amenable in the case of data resources. The community data approach, thus, needs to grapple further with how claims will be managed and operationalized for democratizing data's value.

3.4 Inequality and commodification: The core challenges in the resource governance of data

Value creation in the data economy under the current governance impasse is characterized by the commodification of data – essentially a social commons – and its appropriation by a few first-mover firms. What we note from the above analysis is that prevailing governance approaches – whether individualist or collectivist – are unable to effectively negotiate the 'finders, keepers' scenario that gives exclusive possession rights in perpetuity to the data that these behemoths collect. The result is a crisis at the whole-of-economy level, with colossal corporate dominions that control data enclosures locking out material access to data for value creation by other economic actors.

[11] In specific, Article 39(b) of the Directive Principles of State Policy. See https://www.datagovernance.org/files/research/1604381845.pdf

The commodification of the social data commons fortifies and multiplies an alarming inequality in the global economic order, normalizing data capture. A hegemonic, 'data-must-flow' illogic legitimizes its exclusive possession by a few elite corporate firms (James 2020). The current regime of data governance also undermines human rights, offering little protection to people against the harms that stem from profiling. The proliferation of smart connected objects that aggregate the socio-behavioral footprints of people's interactions with their physical and social environment has opened up a range of concerns. Data mining from smart homes to smart cars – part of the fintech industry's strategies for psychological risk profiling (Marafie et al 2018; Hendricks, accessed 2021) – is one such egregious practice. Trade agreements have become a route to gain access to training data sets from the Global South, allowing European firms to externalize privacy risks to populations from developing countries without personal data protection frameworks. The steady datafication of community knowledge on community-managed biodiversity resources by the biotech industry exacerbates the plundering of such resources for profit, jeopardizing community rights to cultural self-determination (ETC Group 2020). The paralysis of policy has seen the twin-ills of misrecognition (identity-based exclusion and harm) and maldistribution (unfair distribution of access and benefits) in the data economy (Hummel et al. 2020).

4 Governing data for distributive integrity

4.1 Going beyond the CPR and open access commons regimes

Since the value proposition of data rests in its potential for discerning societal relations (among nature-things-people) as the basis for creating products and services, it seems quite self-evident that the starting point for any resource governance regime in data must be collectivist. However, as argued in Section 3.2 and 3.3, fledgling collectivist approaches do not challenge the market dominance of data collectors.

A CPR regime in data through which community rights over it can be realized seems like an attractive proposition at first glance. But the distribution of data across time and space (Prainsack 2019) and the long tail of its downstream uses in aftermarkets makes identifying the boundaries of data as a resource extremely difficult. The fact that data communities are nested/overlapping and not discrete/fixed renders clear identification of community claims to value a near-impossible task. To explain this further: unlike communities with traditional rights in the material natural resource commons with pre-established criteria for membership in the collective, data communities are extremely fluid. Their membership is made and re-made over numerous instances of collection, processing, use, and re-use in the unfolding segments of data value chains. This also means that data communities are generated through specific acts of data processing and individuals may not know the potentially innumerable communities they are part of; or even be in a position to identify other individuals in these communities.

But even if we were to arrive at some acceptable, albeit imperfect, yardstick to categorize the social data commons for a layered separation of collective economic claims among discrete data communities, there is another problem. The CPR regime focuses on how a clearly defined group of people/community with legally/customarily granted economic rights in a resource can exclude non-members from the use of the resource, to avoid its depletion, overuse, and co-option by a few powerful interests. But in the case of the social data commons, we face the opposite dilemma – how can exclusive ownership of data in the hands of a few firms be wrested away for democratizing its use and benefit? The problem in data is one of assembling ownership to a social optimum to address what Heller (2013) terms "wasteful underuse". Multiple and fragmented parcels of ownership in this situation may lead to a tragedy of the

anticommons.[12]

This brings us to whether data as a resource lends itself to open access commons regimes applied in the case of immaterial information and knowledge resources. Open access commons, as Benkler (2016) highlights, are not ungoverned. They are a family of institutional arrangements that focus on "guaranteeing symmetric use privileges to an open general class of users, rather than assigning an asymmetric exclusion right to an individual or known class of individuals".

Making a case for the inappropriability of knowledge as a precondition for social innovation, Drahos (2016) asserts how, more rather than less abstract objects should remain open to use in the intellectual commons.[13] "Openness of use" is vital for the continuous enlargement of such commons. An open access route – rapidly gaining currency as the favored governance modality for public data sets – is not the same. Open access reduced to free-for-all does not carry the "duties of nurture" that must underpin the commons of abstract objects. Unconditional open access also ignores the fact that "accessibility" (ibid) – the capacity and competence to access – is deeply linked to market power.

An appropriate resource regime for data must preserve the 'openness of use', also promoting 'accessibility', that is, the freedom for all economic actors to meaningfully leverage data for unlocking its value. However, any such regime must start with an acknowledgment of the unique nature of data. While similar to information, data is not information. Data is constituted by three distinct layers – a) the semantic/content layer, which encapsulates the information being encoded; b) the syntactic layer, which is the representation of the information collected as machine-readable datasets; and c) the physical layer, the networked infrastructure through which data is extracted (Stepanov 2019).

What we are dealing with in the governance challenge to democratize data value is the ability of the regime to prevent the possessor of the physical-syntactic layers (the network-data architecture) from claiming exclusive rights over the semantic layer in all possible re-uses of the data generated. Any collectivist or commons-based governance framework for data, therefore, needs to deftly manage the delicate balance between openness as non-exclusive accessibility of data's syntactic content and openness as duty to nurture use of data's semantic propositions.

4.2 A semi-commons governance regime for data

In the data economy, the value proposition is in 'inferred data'[14] or the intelligence obtained through algorithmic analysis or manipulation of mixed data sets (personal data sets[15] combined with non-personal data[16]).

[12] Heller's (2013) "tragedy of the anticommons" describes a situation of wasteful underuse of a resource arising from fragmented ownership rights and regulatory controls. The right of exclusion of multiple data communities carries the risk of disincentivizing optimal use.

[13] Intangible objects such as information and knowledge.

[14] According to the Committee of Experts on Non-Personal Data Governance, Government of India (2020), inferred data refers to "an inferred/derived view of data where insights are developed by combining different data points typically involving trade secrets, algorithms, computational techniques, advanced analytics etc."

[15] Personal data comprise any information relating to an identified or identifiable natural person ("data subject"). See https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-250-F1-EN-MAIN-PART-1.PDF

[16] Non-personal data include, firstly, data which originally did not relate to an identified or identifiable natural person. Secondly, data which were initially personal data, but were later made anonymous. Data aggregated to the extent that individual events (such as a person's individual trips abroad or travel patterns which could constitute personal data) are no longer identifiable, can be qualified as anonymous data. See https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-250-F1-EN-MAIN-PART-1.PDF

Private actors, today, rely on trade secrets protection to preserve their commercial interests in specific use-cases of 'inferred data', locking up the base layer of data. Because data is an 'infrastructural good' – a resource whose value proposition stems from its deployment into a wide range of economic activities – high exclusion curbs its innovation potential (Frischmann 2012; Benkler 2016).

To reclaim data as a building block for a fair and equitable socio-economic paradigm, the economic governance regime for data needs to be grounded in the principle of distributive integrity – ordering data value creation and distribution on a set of norms and rules that promote a multiplicity of sustainably productive economic communities.

We propose a semi-commons governance regime for data as a suitable way forward; encoding data as "a medium of [economic] democracy" (Viljoen 2020) and addressing its unique propensities. The semi-commons framework creates and demarcates the boundaries between common property and private property in data resource governance.

Property ownership in the semi-commons framework is understood not as the simple and non-social relationship between a person and a thing, but a complex set of legal relations in which individuals are interdependent and which determines the limits of an individual's or group's freedoms to "use, posses enjoy or transfer" a particular asset (Johnson 2007). These legal relationships are "sets of claims and entitlements in tension with each other, held by people against one another" (ibid), giving rise to a maze of rights and obligations.

The semi-commons framework affords a way forward to optimally balance private and public claims in data. However, if the goal of distributive integrity in the data economy has to be met, any demarcation of private property and common property boundaries must upend the exclusive claims of first movers over all possible future uses of the data collected, as argued above. Common use exemptions in private claims accorded by the mainstream IP tradition cannot measure up to this task (Ciani 2018). Distributive integrity depends upon a radically new normative baseline that affords non-exclusive accessibility of data's syntactic content for all. This means the boundary between private and public claims in data must start from the legal recognition of data as inappropriable social commons with commensurate freedom of open use for all, balanced by limited privileges for data producers.

Such a rights-based resource ownership regime under the data semi-commons framework will confer varying degrees of differentiated access rights and associated conditionalities for economic actors across the spectrum as outlined below.

a. Right to non-exclusive access in the base layer of data

Since data is a shared societal resource, data-holders – the private for-profit/not-for-profit legal entity or public agency that determines the purpose and means of data processing – only have the right to non-exclusive access over the base layer of data they have collected, without exclusive possession rights. The base layer of data includes raw, non-personal data, personal data and aggregate data sets combining personal and non-personal data.

Non-exclusive access in data resources that are self-collected implies the following:

In the data they have directly collected (in compliance with personal data protection legislation and the law of contracts/other relevant laws of the land), data-holders have a 'right to use'. This right includes the right to processing of such data for the generation of inferred data; and the right to obtain profits from inferences/intelligence subject to legally laid out limits for market fairness.

In processing the data collected and in the routes they pursue to generate value from 'inferred data', data-holders are perennially obligated to respect data subjects' right to privacy. So, in addition to safeguards in personal data protection, data-holders must ensure that even with respect to the non-personal data that they have collected, any act of processing does not lead to profiling.

Data-holders have a mandatory duty to share data as required by their obligation to respect the rights of data-seekers, as explained below.

b. Right to seek data

The corollary of the right to non-exclusive access in the data semi-commons is the right to seek data in the datasets collected, aggregated, and controlled by for-profit legal entities, altruistic organizations and public agencies through an entitlement of accessibility. The operation of the differential rights to seek data is explained below and depicted in Table 1.

Data seekers can be individual data subjects, public agencies, or private legal persons. They may access raw non-personal data and/or aggregate non-personal data (except for individuals who only have a right to their own data) defined as follows:

Raw non-personal data: Any non-personal data that is defined by its representative characters and has not undergone any automated or human activity of analysis, reuse, or other manipulation aimed at extracting meaningful information from it (Fia 2020).

Aggregate non-personal data: An aggregated view of the data (like mean, median, mode of the data sets), across several personal and machine-observed data points, without revealing the specific base-level data points and with due safeguards for irreversible anonymization. (Adapted from Committee of Experts on Non-Personal Data Governance Framework, 2020).

Data seekers do not necessarily have an unconditional right to access. Access may vary from one scenario to another, and boundaries will be differentially determined through appropriate institutional mechanisms. Conditional access depends (at minimum) on the following:

The type of data seeker and the type of data holder from whom access is sought.

The specific parts of raw, non-personal data and aggregate non-personal data in which access is sought.

The purpose (contextual applicability) for which data is sought.

The right to seek data is conceptualized as an entitlement granted through law. It does not, however, preclude rights that arise out of private contracts in the data economy.

The establishment of limits of operations of private contracts in the data economy and the downstream rights they produce have an implication for economic fairness. The answers for this cannot be found within the semi-commons itself, and this needs a whole-of-economy approach to governing the data economy. The subsequent section raises some questions on this, but addressing these concerns in theoretically sound ways, requires deeper investigation.

The right to seek, as a conditional right, is exercised by different data seekers in the following ways:

> Individual data subjects have the rights to data access and portability in their personal data and non-personal data they generate through use of smart connected objects.[17]
>
> Public agencies have a right to 'authority access' in the raw non-personal data and aggregate non-personal data held by other private for-profit entities and altruistic organizations. Authority access refers to entitlements of public agencies to access data on the grounds of fulfilling legitimate public policy functions, backed by specific legislation. Data access between public agencies would be governed as per rules and protocols in public policies.
>
> Private legal persons (for-profit entities and altruistic organizations) have a right to conditional access in the raw non-personal data and aggregate non-personal data held by other private for-profit entities, altruistic organizations, and public agencies.

c. Conditionalities in the right to seek

The ends of distributive integrity in a semi-commons regime for data will depend on how permutations and combinations deployed by public policy calibrate the right to seek across competing interests in the data economy. Through institutional mechanisms that ascribe rights, obligations, and privileges to economic actors variously, public policies and legislative frameworks underpinning a semi-commons framework will need to manage the fragile balance between data extractivism and underuse. The conditional access that data seekers have, therefore, depends on broader economic and social policy choices of a particular country or jurisdiction.

For instance, where the seeker is a for-profit entity, only smaller and upcoming private firms would be allowed access to raw non-personal data and aggregate non-personal data of dominant players. This right may not function the other way around. In the datasets controlled by public agencies, for-profit entities can have a right to access on compliance with conditionalities to prevent free-riding (including criteria set by the state for accessing a public data pool or licensing conditionalities with respect to inventions created from the pool and so on).

Similarly, for-profit entities may be disallowed from exercising a right to seek data in datasets held or controlled by altruistic organizations. They may be allowed to gain such access through agreements negotiated under the law of contracts. This would ensure that datasets controlled by altruistic organizations are not easily co-opted by private firms, without appropriate compensation or benefit sharing mechanisms.

Altruistic/not-for-profit organizations may be permitted to access raw non-personal data and aggregate non-personal data held by private-for-profit players when they are setting out to initiate socially relevant data-supported projects, subject to a specific institutional mechanism for public scrutiny. Purpose limitation of potential uses and capital market regulation safeguards would be important to put in place so that dominant private players do not put up a false front in order to gain access to other private players' aggregate datasets through the backdoor.

[17] Discussions in the EU context point to technical challenges in extending access and portability rules to data generated through smart connected objects. As Turner et al (2021) highlight, "IoT devices are not only diverse, but competing vendors collect, store, and process data differently. In particular, technical complexities, such as missing IoT and interoperability standards, the scale and extent of collected data, as well as data subjects' lack of awareness of the nature of data processing can hinder the transmission of data across different systems."

In the case of datasets controlled by public agencies, it is important to ensure that licensing conditionalities are put in place to prevent free riders from capturing and enclosing the value of such data.

Table 1: Differential rights to seek data in the data semi-commons[18]

Data Holder	**Private for-profit entity**		**Altruistic data organization**		**Public agency**	
Access seeker	**Raw non-personal data**	**Aggregate non-personal data**	**Raw non-personal data**	**Aggregate non-personal data**	**Raw non-personal data**	**Aggregate non-personal data**
Individual data subject	Right to data access and portability	No	Right to data access and portability	No	Right to data access and portability	No
Private for-profit entity	Right to conditional access	Right to conditional access	No	No	Right to conditional access	Right to conditional access
Altruistic data organizations	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access
Public Agency	Authority Access	Authority Access	Authority Access	Authority Access	State Policy	State Policy

4.3 Some considerations for institutionalizing the semi-commons model

The institutional design of the data semi-commons is not merely an economic question about a seemingly autonomous zone of the market, but a political one of norm-setting to reorder a data society gone wrong. Delineating the specifics of the institutional design is beyond the scope of this paper, but we are able to reflect on some key considerations and questions that will play a critical role in the norm-setting at the heart of this model.

a. Building a fair and equitable data market and enabling cooperativist production

A tragedy of the anticommons (high fragmentation and underuse) often results from the fact that underuse of a resource is a hidden problem. For example, several patent owners may block a promising line of drug research and it would not be known what lifesaving cures were abandoned (Heller 2013). An institutional framework encouraging open use and data pooling would depend on effective mechanisms to create market certainty. A traceability obligation on data businesses and altruistic data organizations should be mandated by the law. Entities collecting data will need to disclose their sources of data collection to an appropriate authority.[19] Such disclosure norms will automatically limit the over-broad application of trade secrets that is at the heart of the data hoarding problem in the current paradigm. Additionally, regulatory arrangements that extend to standards creation for IoT will be necessary for ease of data portability, robust dispute settlement, and effective collaboration mechanisms in the data economy.

[18] Adapted from Tommaso Fia. (2020).

[19] The metadata register created from the disclosure of data by significant data businesses, that India's Committee of Experts on Non-Personal Data (2020) has recommended, is a pertinent suggestion in this regard.

The potential for innovation in downstream markets, however, hinges on a strong role for public provisioning of data and related infrastructure. The productive capacity of individual firms and the capability of the economy as a whole in leveraging the data economy – as the case of Barcelona demonstrates – requires public agencies at national and sub-national levels to galavanize a new production culture that values distributive integrity. The municipality of Barcelona has implemented its own approach to a 'smart city' where the focus is on creating data infrastructure as a new meta-utility (similar to other urban infrastructure such as water, roads, and electricity), financed through public funding and managed democratically using the principle of citizen data sovereignty (Bria 2019). With support from the EU's DECODE consortium, the city has set up a publicly funded data infrastructure using the following mechanisms (Bass & Old 2020):

> introduction of 'data sovereignty' clauses in all public service contracts that impose a mandatory obligation on any supplier to the Barcelona municipality to share associated data in machine-readable format and using open APIs and open standards to guarantee interoperability across the data pool;
>
> smart contracts and cryptographic tools that enable citizens themselves to directly contribute data to the city data commons in privacy-compliant ways and with full autonomy over the terms and conditions of data sharing; and
>
> opening up the data commons to local companies, cooperatives, social sector organizations that create public value through data-based innovations.

Francesca Bria, a key architect of the initiative,[20] has called attention to how this model could potentially evolve into a trans-European networked cloud and data architecture – providing support to public agencies, Micro, Small and Medium Enterprises (MSMEs), and non-profits for generating public value from collective data through open and shared services – if backed by appropriate licensing conditionalities and public funding.

The initial years of the digital revolution demonstrated that without public funding and policy intervention for universal access, communities who were not seen as investment-worthy by telecommunications providers would be locked into a permanent "access trap"[21] (Alliance for Affordable Internet 2013). Similarly, it is becoming evident that in the data revolution, economic and social needs or service propositions not easily translatable into lucrative aftermarkets may be completely neglected by dominant firms. A 2018 report by the McKinsey Global Institute found that among 19 sectors evaluated in India, the potential value of AI for agriculture was in the bottom tercile. NITI Aayog[22] has also pointed out that a push for AI in agriculture may ultimately be only a lukewarm prospect for the private sector given that, "efforts from private sector may neither be financially optimal nor efficient on a standalone basis" (NITI Aayog, cited in Gurumurthy and Bharthur [2018]). Critical sectors like agriculture that provide livelihood options for millions in developing countries urgently need data-based services for a much-needed productivity boost, but the private sector cannot be relied on to fill this gap.

[20] Francesca Bria served as the Chief Technology Officer of the city of Barcelona from 2016 to 2020. See https://ec.europa.eu/research-and-innovation/en/projects/success-stories/all/enabling-citizens-take-control-their-own-data

[21] In several large middle income countries, although high-end broadband customers in urban areas were well-served in the early 2010s by network operators, poorer communities in urban and rural areas were left behind on account of limited disposable income, low levels of digital literacy and low availability of relevant content. With limited competition, network operators had no incentive to invest in new markets leading to an 'access trap'; a deadlock in which limited demand and lack of incentives for new market entrants left these populations without access to connectivity. http://a4ai.org/wp-content/uploads/2013/12/Affordability-Report-2013-FINAL.pdf

[22] A public policy think tank of the Government of India.

b. Grappling with data protection considerations in mixed data sets

A semi-commons regime for data would need to be bounded by privacy as contextual integrity. Contextual integrity is a theory of information privacy which argues that social relations are guided by norms in information flows and that privacy is violated when these norms are violated (Joshi 2020). For instance, contextual integrity requires that we interrogate what information about a person can be collected in a particular context; and to what extent information once collected can be deployed/shared further in a different context. A burgeoning array of networked, sensor-enabled devices (IoT), and data-ravenous machine learning systems present new challenges to privacy; the crucial question is whether privacy norms governing lower-order data are sufficient for the inferred higher-order data (Nissenbaum 2019). A social relational approach to privacy, hence, calls for agility about all current and potential future uses of data. Distributive integrity of data and the contextual integrity of privacy are two sides of value integrity in the digital economy.

We flagged the lapses in anonymization and risks for individuals and groups in Section 3.1. The rapid rise of ambient intelligence environments is bound to only complicate an already fraught discourse, with an urgency for eliminating/minimizing harms arising from decontextualized information processing. The law needs to grapple with key issues such as protection against reidentification in the sharing of aggregate data that includes mixed datasets, and safeguards against profiling risks that arise in downstream re-combinations of raw non-personal data. The debates in the EU about fulfilling data sharing obligations arising out of competition law with due attention to privacy safeguards are a useful precedent in this regard.[23]

c. Regulating the digital economy

Operationalizing the semi-commons approach involves articulating an institutional regime of norms and principles as well as rules and protocols that mediate data interests at a whole-of-economy level. The local economy and its regenerative potential is core to the distributive integrity of data. The immense power of digital corporations needs reining in through boundary setting of cross-border data flows – controversial, no doubt – but a necessary measure to tackle laissez faire data colonialism and to protect strategic interests, especially in the context of developing countries.

The updation of the 'essential facilities doctrine' (discussed in Section 2) to curb monopolies and enable competition in the downstream markets of data value chains, and new legal protections to prevent business actors from being harmed in multi-sided digital markets, are important directions for a fair digital economy, as the EU experience demonstrates. But in order to fully address the natural monopoly effect of platform capitalism, we may also need interventions for effective structural separation in all layers of data value chains: data layer, cloud computing layer, intelligence layer, and consumer facing intelligent services layer (Singh 2020).

[23] See the decision of the French Competition Authority in its decision of 9 September 2014 in the Direct Energie/GDF Suez case where a fine balance was established between data access under competition law and personal data protection with respect to a data sharing scenario that involved mixed data sets, with inextricably linked personal data.

5. Conclusion

The quest for a suitable economic governance regime for data is not limited to seeking a share of the data pie for individuals and communities while keeping production arrangements in the data economy unruffled. On the contrary, it is about ensuring that every point in the data value chain – from collection, processing, and generation of digital intelligence and the re-use of data and intelligence in various aftermarkets – is organized with the objective of dismantling rentier capitalism and socializing data value.

This paper has demonstrated how the data governance status quo needs to change. The obfuscation of the economic rights question in data governance needs to be urgently remedied through a sui generis semi-commons approach grounded in the norm of distributive integrity. As abstracted social knowledge from data becomes the most valuable means of production, political choices about its resource governance will shape our collective social destiny.

The task of ensuring that data-enabled social knowledge is future-proofed, and economic activities can contribute to human flourishing remains central to scholarly endeavor. Deeper work is needed to explore how data economies based on cooperativist production can be orchestrated. In a globalized, datafied world, theoretical work is also needed to chart a path for a new global constitutionalism that can roll back the tide of digital colonialism in the multilateral order.

References

Alliance for Affordable Internet. (2013). *The Affordability Report 2013*. http://a4ai.org/wp-content/uploads/2013/12/Affordability-Report-2013-FINAL.pdf.

Alam, S., & Faruque, A. (2019). *From Sovereignty to Self-Determination: Emergence of collective rights of indigenous peoples in natural resources management.* The Georgetown Environmental Law Review, 32. https://www.law.georgetown.edu/environmental-law-review/wp-content/uploads/sites/18/2020/01/GT-GELR190045.pdf.

Bass, T., & Old, R. (2020, January). *Common Knowledge: Citizen-led data governance for better cities.* Nesta. https://media.nesta.org.uk/documents/DECODE_Common_Knowledge_Citizen_led_data_governance_for_better_cities_Jan_2020.pdf.

Beamish, B. (2019, September 24). Personal Communication, *Sidewalk Labs' Proposal.* Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/2019/09/2019-09-24-ltr-stephen-diamond-waterfront_toronto-residewalk-proposal.pdf.

Benkler, Y. (2006). *The Wealth of Networks: How social production transforms markets and freedom.* Yale University Press. http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf.

Benkler, Y. (2016). *Open Access and Information Commons.* Oxford Handbook of Law and Economics: Private and Commercial Law.
http://www.benkler.org/Open%20Access%20Commons%20Oxford%20Handbook%20Prepub.pdf.

Bria, F. (2019, January). *The Right to the (digital) City.* Barcelona Metropolis.
https://www.barcelona.cat/metropolis/en/contents/right-digital-city.

Caffentzis, G. (2010). The Future of the 'Commons': Neoliberalism's 'Plan B' or the Original Disaccumulation of Capital?
https://www.researchgate.net/publication/233541411_The_Future_of_%27The_Commons%27_Neoliberalism%27s_%27Plan_B%27_or_the_Original_Disaccumulation_of_Capital

Ciani, J. (2018). *Governing Data Trade in Intelligent Environments: A taxonomy of possible regulatory regimes between property and access rights.* Intelligent Environments.
https://www.academia.edu/37751449/Governing_Data_Trade_in_Intelligent_Environments_a_Taxonomy_of_Possible_Regulatory_Regimes_between_Property_and_Access_Rights.

Committee of Experts on Non-Personal Data Governance Framework. (2020, December 16). *Report by the Committee of Experts on Non-Personal Data Governance Framework.* Ministry of Electronics and Information Technology, Government of India. https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf.

Coyle, D., Diepeveen, S., Wdowin, J., Tennison, J., & Kay, L. (2020, February). *The Value of Data – Policy Implications.* Bennett Institute for Public Policy, University of Cambridge.
https://www.bennettinstitute.cam.ac.uk/publications/value-data-policy-implications/.

Cremer, J., Montjoye, Y., & Schweitzer, H. (2019). *Competition Policy for the Digital Era.* European Union.
https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

Drahos, P. (2016, October 31). *A Philosophy of Intellectual Property.* Routledge.

Drexl, J. (2020, December 29). Chapter 16 – *The (Lack of) Coherence of Data Ownership with the Intellectual Property System.* Transition and Coherence in Intellectual Property Law, Cambridge University Press. https://www.cambridge.org/core/books/transition-and-coherence-in-intellectual-property-law/ED5D2D6D4B0ADB7DF2F97DA26AAE1EF7.

ETC Group/Action Group on Erosion, Technology and Concentration. (2020). *Food for All or Feeding the Data Colossus? The future of food in a digital world.* IT for Change. https://itforchange.net/digital-new-deal/2020/10/30/food-for-all-or-feeding-the-data-colossus-the-future-of-food-in-a-digital-world/.

EU High-Level Expert Group. (2020). *Towards a European Strategy on Business-to-Government Data Sharing for the Public Interest.* European Union. https://www.euractiv.com/wp-content/uploads/sites/2/2020/02/B2GDataSharingExpertGroupReport-1.pdf.

European Union. (2018, November 28). *Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.* Official Journal of the European Union, 303-59. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN.

European Commission. (2017, January 10). Communication on building a European Data Economy. https://wayback.archive-it.org/12090/20171013230250/https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy. Accessed May 2021.

European Commission. (2020, February 19). A European strategy for data. Communication From The Commission to the European Parliament, The Council, the European Economic And Social Committee and the Committee of the Regions. https://eur-lex.europa.eu/legal-content/en/txt/html/?uri=CELEX:52020DC0066&from=EN

European Commission. (2020, November 25). Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN

Fia, T. (2020, September 23). *An Alternative to Data Ownership: Managing access to non-personal data through the commons.* Global Jurist, 21-1. https://www.degruyter.com/document/doi/10.1515/gj-2020-0034/html.

Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources.* Oxford Scholarship.

Gurumurthy, A. & Bharthur, D. (2018). Taking Stock of Artificial Intelligence in Indian Agriculture. http://library.fes.de/pdf-files/bueros/indien/15808.pdf

Gurumurthy, A., Bharthur, D., & Chami, N. (2019). *Platform Planet: Development in the intelligence economy.* IT for Change. https://itforchange.net/platformpolitics/wp-content/uploads/2019/06/Platform-Planet-Development-in-the-Intelligence-Economy_ItfC_2019.pdf.

Hardinges, J. (2020, March 17). *Data Trusts in 2020.* Open Data Institute. https://theodi.org/article/data-trusts-in-2020/.

Heller, M. (2013, January 02), *The Tragedy of the Anticommons: A concise introduction and lexicon.* Modern Law Review, 76-1, 6-25. https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2230.12000.

Hendricks, D. *6 Ways Big Data in Fintech is Creating a Better Customer Experience*. IEEE Computer Society. https://www.computer.org/publications/tech-news/trends/6-ways-big-data-in-fintech-is-creating-a-better-customer-experience.

Hummel, P., Braun, M., & Dabrock, P. (2020, June 15). *Own Data? Ethical reflections on data ownership.* Philosophy & Technology, 34, 545-572. https://link.springer.com/article/10.1007/s13347-020-00404-9.

James, D. (2020). *Digital Trade Rules: A disastrous new constitution for the global economy, by and for Big Tech.* Center for Economic and Policy Research. https://cepr.net/wp-content/uploads/2020/07/digital-trade-2020-07.pdf.

Johnson, D. R. (2007). *Reflections on the Bundle of Rights.* Vermont Law Review, 32. https://lawreview.vermontlaw.edu/wp-content/uploads/2012/02/johnson2.pdf.

Joshi, D. (2020, September 17). *Privacy Theory 101: Privacy as contextual integrity.* Centre for Law & Policy Research. https://clpr.org.in/blog/privacy-theory-101-privacy-as-contextual-integrity/.

Kelly, K. (2009, June 23). *The New Socialism.* Wired UK. https://www.wired.co.uk/article/the-new-socialism.

Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). *'The whole is always smaller than its parts' – a digital test of Gabriel Tardes' monads.* The British Journal of Sociology, 63-4. http://www.bruno-latour.fr/sites/default/files/123-MONADS-BJSpdf.pdf.

Lawson-Remer, T. (2012). *Property Insecurity.* Brooklyn Journal of International Law, 38-1. https://core.ac.uk/download/pdf/228597125.pdf.

Marafie, Z., Lin, K., Zhai, Y., & Li, J. (2018, September 03). *ProActive Fintech: Using intelligent IoT to deliver positive InsurTech feedback.* IEEE Xplore. https://ieeexplore.ieee.org/document/8453934.

Marx, K. *Grundrisse – Marxists Internet Archive.* https://www.marxists.org/archive/marx/works/download/pdf/grundrisse.pdf.

Mazzucato, M. (2019, October 2). Preventing Digital Feudalism. https://www.project-syndicate.org/commentary/platform-economy-digital-feudalism-by-mariana-mazzucato-2019-10

Mazzucato, M., Entsminger, J., & Kattel, R. (2020, September 29). *Public Value and Platform Governance.* UCL Institute for Innovation and Public Purpose. https://www.ucl.ac.uk/bartlett/public-purpose/wp2020-11.

Mills, S. (2020, January). *Who Owns the Future? Data trusts, data commons, and the future of data ownership.* Future Economies Research and Policy Paper #7, Manchester Metropolitan University. https://www.mmu.ac.uk/media/mmuacuk/content/documents/business-school/future-economies/Mills-2020.pdf.

Mittal, A. (2020, November). *Exploring the Constitutional Tenability of Data Sharing Policies.* Working Paper 11, Data Governance Network. https://www.datagovernance.org/files/research/1604381845.pdf.

Nissenbaum, H., (2019). *Contextual Integrity Up and Down the Data Food Chain.* Theoretical Inquiries in Law, 20-1. https://www7.tau.ac.il/ojs/index.php/til/article/view/1614.

Peters, M. A. (2019, April 04). *Digital Socialism or Knowledge Capitalism?* Educational Philosophy and Theory, 52(1), 1-10. https://www.tandfonline.com/doi/full/10.1080/00131857.2019.1593033.

Prainsack, B. (2019, February 14). *Logged Out: Ownership, exclusion and public value in the digital data and information commons.* Big Data & Society. https://journals.sagepub.com/doi/full/10.1177/2053951719829773.

Purtova, N. (2017, November 13). *Do Property Rights in Personal Data Make Sense after the Big Data Turn?: Individual control and transparency.* Journal of Law and Economic Regulation, Tilburg Law School Research Paper No. 2017/21. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070228#maincontent.

Sadowski, J. (2019, January 07). *When Data is Capital: Datafication, accumulation and extraction.* Big Data & Society. https://journals.sagepub.com/doi/full/10.1177/2053951718820549.

Shirky, C. (2009, February 24). *Here Comes Everybody: The power of organizing without* organizations. Penguin Books.

Singh, P. J., & Gurumurthy, A. (2021, January). *Economic Governance of Data: Balancing individualist-property approaches with a community rights framework.* IT for Change. https://itforchange.net/sites/default/files/1880/Economic-governance-of-data.pdf.

Singh, P. J. (2020, June). *Breaking up Big Tech: Separation of its Data, Cloud and Intelligence Layers.* IT for Change. https://itforchange.net/sites/default/files/add/Regulating_data__cloud_and_intelligence_-_Paper_9-21.pdf.

Srnicek, N. (2016, November 25). *Platform Capitalism.* Polity Press.

Stepanov, I. (2019, July 01). *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. https://www.tandfonline.com/doi/full/10.1080/13600869.2019.1631621.

Taylor, L., & Purtova, N. (2019, July). *What is Responsible and Sustainable Data Science?* Big Data & Society. https://www.researchgate.net/publication/334533376_What_is_responsible_and_sustainable_data_science.

Turner, S., Quintero, J. G., Turner, Si., Lis, J., & Tanczer, L. M. (2021). *The Exercisability of the Right to Data Portability in the Emerging Internet of Things (IoT) Environment.* New Media & Society. https://journals.sagepub.com/doi/pdf/10.1177/1461444820934033.

Ullrich, H. (2019, September 27). *Technology Protection and Competition Policy for the Information Economy. From property rights for competition to competition without proper rights?* Max Planck Institute for Innovation & Competition Research Paper No. 19-12. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3437177.

UNCTAD. (2019). *Digital Economy Report 2019 – Value creation and capture: Implications for developing countries.* United Nations. https://unctad.org/system/files/official-document/der2019_en.pdf.

van Geuns, J. & Brandusescu, A. (2020). Shifting Power through Data Governance. Mozilla Insights. https://assets.mofoprod.net/network/documents/ShiftingPower.pdf

Viljoen, S. (2020, November 11). *A Relational Theory of Data Governance.* Yale Law Journal. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562.

World Economic Forum. (2021, April). *Data-driven Economies: Foundations for our common future.* https://www3.weforum.org/docs/WEF_WP_DCPI_2021.pdf#137520_WP_DCPI.indd%3A.44303%3A163.

Yu, P. K. (2019). *Data Producer's Right and the Protection of Machine-Generated Data.* Tulane Law Review, 93. https://scholarship.law.tamu.edu/cgi/viewcontent.cgi?article=2302&context=facscholar.

Acknowledgements

We thank Parminder Jeet Singh for the many conversations on the economic governance of data that inspired the core ideas of this paper. We also extend our gratitude to Arindrajit Basu of the Centre for Internet and Society for his peer review of, and all members of the Data Governance Network for their feedback on, the draft version of this paper. We thank our colleagues, Amay Korjan and Sneha Bhagwat, for their editorial inputs, and Sakhi Shah, for copyediting assistance. We are grateful for the collective knowledge processes at IT for Change that enrich our thinking and analysis.

About the Authors

Anita Gurumurthy is a founding member and executive director of IT for Change, where she leads research on the platform economy, data and AI governance, democracy in the digital age, and feminist frameworks on digital justice. Anita actively engages in national and international advocacy on digital rights and contributes regularly to academic and media spaces. She serves as advisor and expert on various bodies including the United Nations Secretary-General's 10-Member Group in support of the Technology Facilitation Mechanism, the Paris Peace Forum's working group on algorithmic governance, Save the Children's ICT4D Brain Trust, and Minderoo Tech & Policy Lab's Board.

Nandini Chami is Deputy Director and Fellow, IT for Change. Her work focuses on research and policy advocacy in the domains of digital rights and development, and feminist digital justice.

datagovernance.org dgn@idfcinstitute.org

@datagovnetwork /datagovnetwork /datagovnetwork