

Governing the Abstract Object of Data: Towards a Distributive Integrity Framework

Anita Gurumurthy and Nandini Chami
IT for Change¹

¹This paper builds on the emerging body of work at IT for Change in thinking about data governance from the standpoint of the international political economy of development in the digital context.

Table of Contents

1 Introduction	3
2 Data's subsumption into capital's circuit of control	5
3. Current approaches to data governance – why they fall short	8
3.1 The individualist approach to data governance	8
3.2 Collectivist approaches to data governance	12
3.3 The 'community data' approach	14
3.4 Inequality and commodification: the core challenges in the resource governance of data	15
4 Governing data for distributive integrity	16
4.1 Going beyond the CPR and open access commons regimes	16
4.2 A semi-commons governance regime for data	18
4.3 Some considerations for institutionalising the semi-commons model	23
5 Post Script	26

Governing the Abstract Object of Data: Towards a Distributive Integrity Framework

Abstract

This paper presents the building blocks of an economic governance regime for data that can contribute to the growth of the digital economy by unlocking the potential of data. The de facto enclosure of data by platform monopolies not only creates a skewed market place with exclusionary effects, it also presents a huge opportunity cost in terms of the unrealised public and social value of data. The paper unpacks the limits of emerging approaches to data governance – both, individualist and collective – arguing why such frameworks do not go far enough. It offers a semi-commons approach to ordering the digital economy as a means to achieve ‘distributive integrity’.

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, excerpt from *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 productive debates on a wide-ranging plane. 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, presenting an urgency for a paradigm shift in the economic sphere.

As abstracted 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 the most appropriate one for evolving a new global digital economic paradigm centred on human flourishing.

Section 2 reflects on the *modus operandi* of data capitalism. Although data is born social, the *de facto* ownership of data and ensuing monopolisation 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, arguing that the question of economic rights in data has been disregarded by this approach. Subsection 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. Subsection 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 democratised. Subsection 3.4 underscores the urgency of a governance framework for data, calling attention to the rampant mis-recognition 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 that can draw a normative baseline in ‘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 normative 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 reorganising the social relations of production. The techno-social possibilities for commons-based peer production (Benkler 2006) seemed highly plausible, signalling 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 organising 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, with 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 the democratic, pluralist and decentralised web, the new Internet was increasingly assuming a hierarchical, centralised character. The frenetic increase in computing power and exponential growth of digital data generated either directly in the virtual space or indirectly in the physical space has seen the emergence of twenty first world order, understood widely as ‘platform capitalism’. From the innocent logic of social networks to the attention economy’s feverish aggregation of users into profiles, Uberisation (capitalist subversion of cooperative forms of peer production and exchange in the on-demand economy), Amazonification (parasitic capture of network economies) and finally, the universalisation of the platform model in the tangible economy (new processes of appropriation of the ownership of the means of production), the short history of platform capitalism stands tall as a spectacular myth.

The rapid consolidation of market power in the hands of winner-take-all platform behemoths – the “capitalist platform firm” (Srnicek 2017) – marks the platformisation 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 userbase. They also build an intelligence advantage through incessant algorithmic mining of data-based insights in the networks they control, by hyperoptimising 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 2017, Gurumurthy, et al 2019).

Data-based insight generation under digital capitalism² 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 (Sadowsky 2019). The data paradigm³ can be viewed as a breakdown of the

²The terms ‘digital capitalism’ and ‘data capitalism’ are used interchangeably in the paper.

³Suggesting the paradigm shift towards big data and associated technologies.

distinction between the level of individual component and that of aggregated structure, whereby every “whole is always smaller than 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 economy⁴ revolve around the accumulation of data for its potential, rather than, immediate, use value (Coyle et al 2020, UNCTAD 2019, Stanford Graduate School of Business 2017). 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, and could potentially become valuable, if new questions not yet thought of can be answered in future (Coyle et al 2020). The *raison d’être* of digital capitalism hence is to increase the option value of data sets, aggregate and recombine data in a myriad different ways to increase the possibility of generating analytical insights for a breadth of applications. Correspondingly, with developments in mobile, 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 represents 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 ensuing monopolisation 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; Taylor and Purtova 2019) is 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. The emergence of data as the object of capitalism in the past two decades has in fact been co-linear with the concentration of economic power in the hands of a few, a falling share in the returns for labour as compared to capital, and limited progress in reducing income inequality within and among countries

⁴The terms ‘data economy’ and ‘digital economy’ are used interchangeably in this paper.

(Gurumurthy and Chami 2020). Just seven companies from two countries – the United States and China – account for two-thirds of the total market value of the world's top 70 digital platforms, cornering 75 per cent of all patents related to blockchain technologies, 50 per cent of global spending on the Internet of Things, and more than 75 per cent of the world market for public cloud computing (UNCTAD 2019).

b). The de facto control of data by lead digital firms creates a skewed political economy of digital development in which developing countries cannot build their competitive advantage. Many of these countries are often coerced into data-for-good initiatives that render their populations vulnerable to the predatory designs of Big Tech (Taylor 2016), and create dependencies that sidestep questions of domestic data infrastructure capabilities .

The rhetorical capture of policy debates on data by powerful Silicon Valley companies, with the backing of the United States government and its trade allies, is another challenge for developing countries. Any attempt by developing countries to introduce policy interventions in the data economy is cast as a threat to the global democratic Internet and free trans-border information flows. As Singh and Vipra (2019) highlight, discursive ploys to promote 'data flows with trust' claim to take care of the concerns of developing countries through privacy and intellectual property protections in global data flow frameworks. This obscures 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 prospects for the flourishing of smaller businesses and local communities, locking up data value in the silos of rentier capitalism, and prevents positive externalities in the form of data's public and social value from being realised. It affects indigenous people from the global South disproportionately. The shift towards datafication in traditional sectors like biodiversity through genetic sequencing techniques without a corresponding economic governance regime for data creates unprecedented challenges that destabilise existing economic rights frameworks and rules in these sectors.⁵

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 governance question about data and its unique properties as an economic resource.

⁵Indigenous communities enjoy rights and privileges under the Nagoya Protocol of the Convention on Biodiversity. Digital sequencing of genetic resources allows Big Pharma and Northern researchers to take away genetic sequence information in direct violation of these rights.

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 that lacks both a robust data privacy legislation and a competition law framework that can effectively challenge market dominance. 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 centred 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 European Union 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 approach⁶, 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 market⁷ specified by personal data protection legislation (Viljeon 2019). Personal data that is anonymised and machine-observed data that do 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 maximised for the development of the data market (European Union Regulation on the Free Flow of Non-Personal Data 2021). In this approach, except in the case of willful/inadvertent deanonymisation 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* anonymised personal data or *their* data footprints in machine-observed data is completely sidestepped.

This approach suffers from the following shortcomings:

⁶As in the case of the EU’s General Data Protection Regulation

⁷Including obligations of data processors

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 oftentimes become visibilised 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-behavioural 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 anonymisation, thus posing regulatory challenges to prevent collective harm. As Fitbit's privacy policy indicates, de-identified personal data is used to "inform the health community about trends; for marketing and promotional use; or for sale to interested audiences" (Ciani 2018).

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 get 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.

As Yu (2018) highlights, 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 2018). 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 European Union 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 debate has been around the use of competition law as an instrument for effectively breaking 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'⁸ 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"⁹ (Cramer et al 2019). However, these efforts may not go very 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. It is not because access might be pro-competitive that a claim to access is justified under competition law, but because its dismissal by the holder of the data serves or results in a restriction of competition that is not otherwise justifiable (Ullrich 2019).

Another discussion has been around addressing the disadvantage that traditional firms face when negotiating with digital companies in smartifying value chains in the real economy, by introducing a new data producer's right for non-personal, anonymised machine-generated data (Yu 2018). As elucidated in the European Commission'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.

⁸The Essential Facilities Doctrine is a structured rule originally developed to decide under which conditions the denial of access to infrastructures should be considered anti-competitive. Under the test, a company with a dominant position in the provision of a facility, product or service which is indispensable to compete in a downstream market abuses its dominant position where, without objective justification, it refuses to grant access to this facility, product or service.

⁹Markets that are part of the broader ecosystem served by the firm that is the data controller.

The European Strategy for Data (2020) makes no mention of this proposal. But in any case, this proposal suffers from two critical drawbacks. Firstly, even though the right has been conceptualised as a de facto possession right rather than an ownership right, it allows rights-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 bringing injunctions against unauthorised uses (Stepanov 2020). The collective benefits of a social resource like data may still not materialise simply with a complex mesh of private property claims.

In summary, 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 in preventing the moral inappropriateness of converting inalienable personal information into an alienable economic object (Prainsack 2019), and leaves untouched the exclusive control that data collectors have over the social data they hoard.

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 (ODI 2019). 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 modelled 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 (Mozilla Insights 2020). The EU's draft Data Governance Act 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 government-to-business data sharing through data stewardship mechanisms (to be added).

The notion of stewardship has been received positively across the ideological spectrum. However, even as they seek to target corporate data extractivism, under the de facto economic regime for data, data stewardship models will most likely end up as a device for large data monopolies to externalise their regulatory burden, reducing administrative costs and reputational risks in the process of data collection and processing (Stuart Mills 2020). Stewardship may also serve as a perverse incentive in the trading of aggregated personal data (Viljeon 2019). The EU is contemplating the possibility of data stewardship to enhance trust-building in business-to-government data sharing, compensating businesses for the data shared with public authorities through such arrangements (EU B2G data sharing document). Such a model only legitimises data enclosures further 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 governance of the project's data. But as the then Information and Privacy Commissioner Brian Beamish pointedly submitted in a 2019 letter to the Waterfront Board, the urban data trust would end up becoming a privatised, unaccountable mechanism governing data collection and processing under the project 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 endeavour that seeks to co-design, pilot and scale flexible data governance mechanisms. The aim is to unlock data from existing siloes and create opportunities for both the public good and commercial benefit (to be added). 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 (WEF 2021). As an effort at data stewardship, the initiative is an attempt at transparent rule setting for trade in data and to clarify private property claims (from the data originator to the initial data collector to 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 catalyse data-based value creation and remedy the suboptimal use of data resources. While transparency-based exchange may be a welcome departure from the dark markets in which personal data is widely traded today for modest sums, the effort's ability to unlock 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 poor quality data infrastructure, and hence, starting off with a handicap. Transparency in data exchange does not also automatically incentivise smaller players in the data field, 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, political philosopher George Caffentzis, terms, “a pro-capitalist commons”. Data market propositions that convert the social relational resource that is data for capital accumulation leave untouched the unequal ownership of data as a means of production.

Self-organised data communities may at best represent enclaves of powerful, alternate visions¹⁰, 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” (Appendix 7 to the Report of the Committee on Non-Personal Data, GOI, December 2020). Community resource governance frameworks for the natural resource commons, such as the Nagoya Protocol¹¹ 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-behavioural datasets and intelligence about communities by private corporations, and unlocking them for the common good (ibid).

The Indian approach moves the needle in the data governance debate by acknowledging data’s social moorings, and hence, anchoring data rights in its associated communities, and also identifying data-enabled value creation as a vital national public policy issue. Its proposal for actualising 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 resources¹²

¹⁰Take for instance, ‘The Distributed Cooperative Organisation’ framework proposed by progressive organisations 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.

¹¹on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation.

¹²In specific, Article 39(b) of the Directive Principles of State Policy. See <https://www.datagovernance.org/files/research/1604381845.pdf>

(Mittal 2019) – 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 modelled 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 operationalised for democratising 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 characterised 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 (Chatzigiannakis et al 2018) 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.

The commodification of the social data commons fortifies and multiplies an alarming inequality in the global economic order, naturalising data as a free-for-all commodity through rhetorical capture. A hegemonic, ‘data-must-flow’ illogic legitimises its exclusive possession by a few elite corporate firms. As discussed in Section 1, in the world economic paradigm reorganised on the basis of data-enabled intelligence, countries in the global South find themselves at the margins, lacking in the capacity for

data-age innovation and production. Equally, the context bespeaks a huge opportunity cost in terms of data's untapped public and social value.

The de facto regime of data governance undermines human rights, offering little protection to people against the harms that stem from profiling. The proliferation of smart connected objects for aggregation of socio-behavioural 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; Hendriks, 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 externalise privacy risks to populations from developing countries without PDP frameworks. The steady datafication of community knowledge on community-managed biodiversity resources by the biotech industry exacerbates the plundering of such resources for profit, jeopardising 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 2019).

4 Governing data for distributive integrity

4.1 Going beyond the CPR and open access commons regimes

Considering that the value proposition of data rests in its potential for the discernment of 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 2, fledgling collectivist approaches do not challenge the de facto ownership of data collectors.

A CPR regime in data through which community rights over it can be realised 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 in the case of communities with traditional rights in the material natural resource commons with pre-established criteria for membership in the collective, data communities are extremely fluid in that 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 categorise the social data commons for a layered separation of collective economic claims among discrete data communities, there is another critical 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 democratising 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.¹³

This brings us to whether data as an abstract object 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.¹⁴ “Openness of use” is vital for the continuous enlargement of such commons. An ‘*open access*’ route – rapidly gaining currency as the favoured governance modality for public data sets – is not the same. Open access reduced to free-for-all does not carry the “duty 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,

¹³Heller’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 disincentivising optimal use.

¹⁴Intangible objects such as information and knowledge.

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 democratise 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'¹⁵ or the intelligence obtained through algorithmic analysis or manipulation of mixed data sets (personal data sets combined with non-personal data).

Personal data comprise any information relating to an identified or identifiable natural person ("data subject")

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.

Source: European Commission (2019).

<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 *tour de force* 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.

¹⁵According 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."

We propose a semi-commons governance regime for data as a suitable way forward; encoding data as a medium of [economic] democracy” (Salome 2016) and addressing its unique propensities. The semi-commons framework creates and demarcates the boundaries between common property and private property in data resources ownership. Ownership, in this 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, possess enjoy or transfer” a particular asset (Johnson 2012). Rights in this framework are opposed by a negative right, or the absence of the entitlement that was embodied in the positive right. Needless to add, the normative framework that is used to determine the boundaries of these negative and positive rights and the obligations of rights-holders to others is at the heart of the question of economic fairness (ibid).

With a normative baseline in ‘freedom of open use’ in data resources, the semi-commons approach proposes a rights-based resource ownership regime, with varying degrees of differentiated access rights and associated conditionalities for economic actors across the spectrum.

The key elements of the ‘bundle of rights’ in the data semi-commons approach are 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 that:

- In the base layer of data that data-holders have acquired either through direct collection in compliance with personal data protection legislation, the law of contracts and other laws of the land, they have the ‘right to use’ – that is the processing of such data for the generation of inferred data; and they have 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 bound by the obligation to respect data subjects’ right to privacy. What this means is that 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 organisations 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 (excepting 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 anonymisation. (Adapted from Committee of Experts on Non-Personal Data (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 conceptualised 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 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.¹⁶
- 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 organisations. 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 organisations) 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 organisations 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 data 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 could 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 organisations. They may be allowed to gain such access through agreements negotiated under the law of contracts. This would ensure that datasets controlled by altruistic organisations are not easily co-opted by private firms, without appropriate compensation or benefit sharing mechanisms.

¹⁶Discussions 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 (2020) 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.”

Altruistic/not-for-profit organisations 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.

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-commons¹⁷

Data holder →	Private for profit entity		Altruistic data organisation		Public agency	
	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 organisations	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 institutionalising 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 under-use 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

¹⁷Adapted from Tommaso Fia

effective mechanisms to create market certainty. A traceability obligation on data businesses and altruistic data organisations should be mandated by the law. Entities collecting data will need to disclose their sources of data collection to an appropriate authority.¹⁸ 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.

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 galvanise 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 (xxxx). With support from the EU’s DECODE consortium, the city has set up a publicly funded data infrastructure using the following mechanisms (xxxx):

- 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, coops, social sector organisations that create public value through data-based innovations.

Francesca Bria, a key architect of the initiative,¹⁹ has called attention to how this model could potentially evolve into a trans-European networked cloud and data architecture – providing support to public agencies, 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 (xxxx).

¹⁸The 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.

¹⁹Francesca Bria served as the Chief Technology Officer of the city of Barcelona from 2016 to 2020.

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”²⁰ (A4AI, 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 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.

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 xxxx). 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 anonymisation and risks for individuals and groups in Section 2. The rapid rise of ambient intelligence environments is bound to only complicate an already fraught discourse, with an urgency for eliminating/minimising harms arising from decontextualised information processing. For instance, 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 are key issues that the law needs to grapple with. The debates

²⁰In 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>

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.²¹

c. Regulating the digital economy

Operationalising the semi-commons approach involves the articulation of 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).

5 Post Script

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 whilst 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 organised with the objective of dismantling rentier capitalism and socialising 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. It has also argued how the question of property rights in the abstract object of data is about institutionalised social relations. 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 so that a million flowers can bloom, and economic activities can contribute to human flourishing remains central to scholarly

²¹See 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.

Draft version. Not to be cited.

endeavour. Deeper work is needed to explore how data economies based on cooperativist production can be orchestrated. In a globalised, 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.