Economic Governance of Data

Balancing individualist-property approaches with a community rights framework

(Draft for discussion at Quarterly Roundtable of Data Governance Network[1])
January, 2021

Parminder Jeet Singh and Anita Gurumurthy,
IT for Change

Table of Contents

1 With apologies for the missing citations.

Introduction

Governance of data has mostly been from a harm prevention perspective, involving privacy and security. But data is also considered the central resource of a digital society. As digitalisation covers all sectors, data is set to become the most valuable resource for the overall economy. Value chains will be led by those who hold and control the most valuable data in any sector. Seven out of the top eight companies globally by market capitalisation today are digital companies.[2] The real value of these companies is increasingly in the data they hold and process even more than the monopoly platforms that most of them own. Talking about his priorities, the European Commissioner for Internal Market, Thierry Breton, recently observed that while 5G, artificial intelligence, industry, services, audiovisual, tourism, space, defence, etc., were all important, "their foundation, their common denominator, what runs through all their activities from end to end, is data".[3]

There however currently exists almost no economic governance of data. Economic governance of data is about apportioning the value of data, and related issues. Many reasons are put forward for this. It is claimed that data is a nebulous thing, both epistemologically and ontologically. This means that is difficult to know what to consider data and what not, and which is which kind of data. And, even if conceptually we could know it well enough, it is not easy – such is the nature of the networked entity of continually renewing big data – to physically pinpoint and circumscribe the precise object to which economic governance or rights may be applied. The nature of value creation and of value transactions are evolving by the day, and even those involved with data businesses do not yet fully understand them. As the conclusions of numerous papers on the subject testify, a common view is that it may be too early to set up frameworks for economic governance of data.[4] Instead, to the extent possible, existing instruments of economic regulation like in competition law and contract law should be employed for achieving various policy objectives related to economics of data.[5]

This genuine problem about the very nature of data, and the lack of maturity in understanding it, having been stated; the lack of alacrity in moving ahead with economic governance of data has at least as much to do with whose perspectives and interests dominate the data governance narrative. Keeping economic governance of data related discussions away from the policy table most suits those who currently partake of all the value of data collected from across the world, i.e. a few global digital corporations. It also serves the governments of those countries where almost all such corporations are based, the US, and now increasingly some in China. Non-governance of economics of data disadvantages most the countries and groups that are fast losing out in the emerging global digital economy equations, value chains and hierarchies, which are Southern countries, and its people. Unfortunately, these actors also possess the least capacity to develop new data-related understanding and the required frameworks for economic governance for data.

This has been most visible at the WTO, where in July 2016, the government of USA introduced a new template for global digital and data governance, in form of the proposed e-commerce rules.[6] The centrepiece of this proposal is global free flow of data, which basically means that whoever collects data globally retains it, and its entire value. These rules are meant as a pre-emptive measure against effective national regulation

2 That is if we do not consider Tesla a digital automobile company. If we do, all eight are digital corporations. https://en.wikipedia.org/wiki/List_of_public_corporations_by_market_capitalization#2020 Accessed January, 2020.

3 https://www.linkedin.com/pulse/europe-has-everything-takes-lead-technology-race-thierry-breton/

4 Cite

5 Cite

6 This came from the Trans Pacific Partnership trade agreement that the US was negotiation, but the Trump administration later refused to sign, for reasons unconnected to the digital content of this agreement. Indeed US has adopted the same template of global digital governance in other bi-lateral and multi-lateral trade treaties that it has entered into, or is negotiating.

of digital economy and data in the future. As developing countries like India and South Africa stoutly resist such efforts,[7] their task has been rendered difficult by absence of clarity at the domestic level on how to regulate a digital economy and data.

EU's role has been most interesting in this regard. Bob van Dijk, the head of South Africa's biggest digital company Naspers, thinks that the EU is beginning to be a US's digital colony.[8] EU leaders are alarmed about their dismal digital position vis a vis the US and China. This has inter alia caused fervent activity about data governance, including attempts to address the economic dimensions of data. At global forums like the WTO, however, EU sticks to its traditional global political economy camaraderie with the US and faithfully toes latter's 'free flow of data' line. This is despite a planned pluri-lateral, Trade in Services Agreement, having failed because the US and the EU could not agree on issues around data flows.[9] At the domestic level, even as EU's analyses of data-related issues and problems are generally very sound, its recent remarkably intensive efforts at economic governance of data still remain weak in their concrete implications, as will be discussed later.[10]

The main reason for this is that while its analyses have invariably demonstrated the need for greater domestic access to data, which requires more data sharing, the EU's approach remains contained within individualistic and property based frameworks. Data's greatest economic value (other than for personalisation of services which requires individual or personal data) is in the patterns found in aggregated big data. It is such abstractions from the individual or the personal that produce the all important general insights and digital intelligence.[11] The EU's draft Data Governance Act (2020) focuses on establishing "domain-specific common European data spaces as the concrete arrangements in which data sharing and data pooling can happen".[12] The nature of data's value, and the well-recognised need for data sharing, both have an evident collectivist character to them. No wonder then that an exclusive reliance on individualist and market-property centred thinking is unable to address the 'data problem', as discussed throughout this paper.

Exclusive hoarding of world's, and any country's, data by a few digital corporations, with the consequent lack of access to such data for the domestic digital industry, is the key 'data problem'. The corresponding 'solution' is to ensure widespread data access and sharing. This calls for some kind of a collective-commons approach to economic governance of data. One such framework, anchored around the concept of 'community data', has begun to take shape in the policy discourse of India.[13]

This paper first examines the emerging efforts at economic governance of data, presenting their useful points as well as limitations. The latter arise largely from the reluctance of current efforts to go beyond exclusively individualist, market and private property frameworks. In its second part, the paper argues a community and commons based approach to economic governance of data, within which various kinds of private claims and rights to data can be framed.

7 Cite

8 "India should never be a digital colony like Europe: Naspers CEOC" Economic Times, https://economictimes.indiatimes.com/small-biz/startups/newsbuzz/india-should-never-be-a-digital-colony-like-europe-naspers-ceo-bob-van-dijk/articleshow/63074511.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

9 TiSA discussions hit privacy, data protection roadblock, https://iapp.org/news/a/tisa-discussions-hit-privacy-data-protection-roadblock/

10 Detailed note here on what is happening, including new acts

11 cite

12 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

13 This has been seen in various meeting of Southern delegates on data related issues held at the South Centre, Geneva. South Africa has shown special and sustained interest in this regard.

Part I: Existing data governance – Individualist and private property based

Economic governance of personal data

EU's General Data Protection Regulation, or GDPR, enables a data subject to withdraw her data and/or port it to other service providers. The data collector is under obligation to facilitate this process. Similar data portability provisions figure in most data protection laws or proposals, including the Personal Data Protection Bill currently with the Indian Parliament. This appears to be only data related economic right that privacy regulations give to the data subject, going beyond the negative right of protection to a positive right of obtaining the best economic value from one's data.

Such a right is considered by some observers to be akin to endowing data ownership upon data subjects. Others have argued that since all residual rights about how the collected data may be used, beyond the explicit provisions of the GDPR, remain with data collectors, it is the data collector who gets data ownership.[14] The reality may lie somewhere in-between. Right of dispossession is an important ownership right. However, the numerous residual rights untouched by the GDPR are equally important. For instance, the data collector unilaterally captures all the value arising from deanonymisation and aggregation of data. It is under no obligation to share any part of this enormous value back with the original contributors of data, individually or collectively.

Data portability and other obligations on data collectors ostensibly exist for individual data subject's empowerment.[15] But the GDPR also had a hidden industrial policy objective. European companies were expected to turn privacy protection into a competitive edge, thus challenging the US digital giants.[16] This did not happen. If anything, the GDPR may have helped further consolidate the power of US's digital corporations in Europe at the expense of smaller companies.[17]

A second phase of promoting data subjects' economic control over their personal data now comes with the recognition that individuals may not be able to manage their data and its controlled availability on their own. The solution is specialised intermediary data institutions that would assist individuals to achieve such control. Entirely private projects like MyData exist in this respect.[18] There are also regulation based initiatives like open banking in the EU[19] and Accounts Aggregators in India[20], both concerning financial data. Information Banks of Japan extend such an institutional framework for controlling personal data to all kinds of data, enabling data subjects to even monetise their data.[21] It is a lightly regulated self-governance initiative involving certification by an industry body that increasing trustworthiness. This line of thinking on personal

---

14 p-17, "The economics of ownership, access and trade in digital data", European Commission JRC Digital Economy Working Paper 2017-01, https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf

15 Recognising their economic rights aspect, these portability rights are sought to be expanded through new laws that are largely economic governance minded, like using these rights for creating 'personal data spaces'. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=EN#footnoteref54

16 Placeholder https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2019/gdpr-compliance-as-a-competitive-advantage

17 GDPR has been a boon for Facebook and Google, WSJ, https://www.wsj.com/articles/gdpr-has-been-a-boon-for-google-and-facebook-11560789219

18 https://mydata.org/mydata-101/#:~:text=MyData%20is%20a%20human%20centred,us%20in%20our%20everyday%20actions

19 cite

20 cite

21 https://asia.nikkei.com/Business/Business-trends/Japan-s-information-banks-to-let-users-cash-in-on-personal-data

data monetisation has led to calls for recognising personal property rights in ones data.[22]

Monetising personal data mostly brings measly returns.[23] Personal data that is sold is either used as such to target the concerned data subject, or it is anonymised for developing general intelligence. If employed for personalisation, to the extent such data sharing is a fair win-win, there is no point in one side remunerating the other. Even if it was to happen, it will become a part of the cost structure of the overall product or service transaction, thus cancelling any benefit to the remunerated party. It is like sharing information with the salesperson at a shop, say, about what kind of washing machine one needs which helps him connect one to the right model, thus fairly benefiting both sides.[24] It is not meaningful even to charge someone for surfacing needs that one may otherwise not have consciously felt, as advertisements often do, if it ends in a net positive service. On the other hand, data transfer mostly being a transaction between very unequal parties – an individual and possibly a large and digitally-powerful corporation, and every piece of data providing one-sided cumulative intelligence about the subject, it is highly likely that the digital corporation takes a much greater and unfair share of the 'data dividend'. A data-based relationship may even be of net negative value to the data subject compared to obtaining a similar service without providing her data, or as involving possible alternatives involving fair sharing of data. Selling data does not address this problem; it only exacerbates it because data subjects may willingly sell much more data than they otherwise would have shared, causing larger overall net loss to them.

If data is sold for anonymised sharing, or otherwise for building larger knowledge or intelligence for the data buyer that may not be specific to the data seller,[25] there is another kind of problem. Firstly, the marginal value of any one set of personal data to an anonymised big data pool being extremely low, much remuneration cannot be expected. More importantly, there is a significant negative externality problem here. Any individual selling data feels little or nothing lost in doing so, while the long-term collective and cumulative loss to the 'group of people like her' or 'community' – about which new intelligence gets generated – can be enormous in the form of data-based exploitation. This shows the limitation of focussing exclusively on expanding personal choice in the face of problems that may be of a collective nature.

The recent draft EU Data Governance Act, 2020, goes furthest in providing legal and institutional clarity about data intermediary services, or data sharing services. Regulation is aimed at ensuring greater trustworthiness of such services. The draft Act seeks to regulate only pure data intermediaries services, leaving untouched the mainstream integrated business model of giant digital corporations who collect personal data and freely move it around in various ways (Cambridge Analytica was just one more egregious instance of it)[26].

In leaving out mainstream practices of data collection, use and re-use, it is not clear whether the new regulation will encourage independent data intermediary services or discourage and impede them; because they now bear new regulatory burden not carried by integrated data value chain based business models that

22 In 2019, in California, Governor Gavin Newsome proposed the idea of a "data dividend" that will allow California residents to profit from their online data. Shortly after, Senator John Kennedy (R-LA) introduced the "Own Your Own Data Act of 2019," which provides that "each individual owns and has an exclusive property right in the data that individual generates on the internet" and requires social media companies to obtain licenses for the use of data. https://www.linkedin.com/pulse/you-dont-need-property-rights-monetize-your-personal-data-turrecha/

23 https://www.wired.com/story/i-sold-my-data-for-crypto/

24 An online parallel will be someone making it known in an online profile that she is considering, say Spain or India for holidays this year, which enables tourism companies dealing with these destinations to get in touch with her.

25 Some traditional cases of data sharing of this kind are various kinds of surveys and clinical trials.

26 cite

they are to compete with. The expectation of the regulation apparently is that with independent data intermediaries rendered more trustworthy because of certification under it, data subjects will exercise rational choice in their favour over handing their data directly to mainstream digital platforms.

This puts the onus back on an individual's initiative and personal responsibility to best use her data, now that the law makes available regulated, trustworthy data intermediaries 'as an option'. This is missing the lessons of GDPR's implementation, that individuals do not seem to respond to such incentives, caught that they are in very alluring, intense, and sticky interactions with digital giants. It is not the lack of trusted data intermediary services – useful as they are – that is keeping them from more closely controlling the collection and use of their data. Even if data intermediaries services were to really take off and become a credible alternative to existing business models, it is not evident what prevents dominant digital companies from employing – what is there their old trick – the carrot of some additional conveniences and perhaps other sundry benefits, and the stick of many kinds of hidden switching-costs, to retain almost their data providing consumers. This clearly is the most likely scenario, from what is known in the past.

It is a naive and unconvincing expectation that the new regulation will in any significant manner change the current situation, where the few existing data intermediary services get scarcely used, just by providing assurance of their neutrality and quality to potential users.

It would be different if regulation mandates that firms providing data based services can deal with data subjects only through such regulated independent data intermediaries. The regulation states that to bring trust and more control for data subjects it "require(s) structural separation between the data sharing service and any other services provided" . But without mandating such structural separation, it exist only conceptually and in the minds of the regulators. Any such structural separation needs to be enforced by law, something which the draft Data Governance Act does very little about.[27] It stops data intermediary services from undertaking data value add, but does not act in the reverse direction, of stopping value add services providers from being data intermediaries – whereby a Facebook or Google would be prohibited from sharing any user data with anyone, with or without conditions, even with data subject's permission.[28] In fact, real structural separation would exist if these corporations are also banned from collecting data directly from data subjects and are able to only go through data intermediaries. That would render independent data intermediaries a viable business model.

The draft Act recognises possibilities of collective bargaining and action through 'data cooperatives'. The strength of collective control of their data by data subjects, and collective bargaining based on it, could in principle certainly challenge mainstream models. But again, it remains unclear what incentives data subjects have to get into such a cooperative mode as against the current individual mode of consuming digital services and providing data as a part of the same process.

It would have been different if data cooperatives had some kinds of collective rights, for instance to their non-personal data – a category of data which definitionally cannot be individually controlled. The draft Act does speak of "data pertain(ing) to several data subjects within … (a) group' and the role of a data cooperative in "potentially solving disputes between members of … (the) group on how data can be used".

27 https://datagovernance.org/report/breaking-up-big-tech-separation-of-its-data-cloud-and-intelligence-layers

28 Such an enforcement does bring another problem: these corporations will internalise more and more of data business possibilities rather than doing things in partnerships, as Facebook did for many kinds of targeted advertisement after the Cambrige Analytica scandal.

However, what is required is a collective right of the group over such common data, that can be exercised through a 'data cooperative', say, with respect to with a digital platform. Such a collective right to aggregate data is implicit in the language of the Draft Act when discussing 'data cooperatives'. [29] That would provide a great incentive for data subjects to get into various kinds of data cooperatives.[30] For instance, business users of large digital platforms can very usefully employ such a data cooperative exercising a collective right, something discussed later in this paper. Merely 'allowing' data cooperatives of natural or legal persons to exist, without laying down any special basis or rights for them, remains largely a statement of good intention rather than bear any chance of real change.

Such efforts of EU substantiate how even though its policy makers get their problem analysis mostly right, actual regulations remain hamstrung by a lack of will to get into the needed hard law and new data rights frameworks. In this case, these could respectively have been in the from of mandated structural separation and new collective economic rights around data. The EU still lingers at the stage of stating its intent – however clearly and powerfully, and, to some extent, of tentative explorations of what may be needed, and hopefully can fructify in the future. The time though may be running out for it, and others, such is the pace of entrenchment of global digital dominations.

Indian Personal Data Protection Bill (2019) has provisions of consent managers that will be registered, and subject to various "technical, operational, financial and other conditions as may be specified by regulations".[31] India's central bank has notified such consent managers for the financial sector called Account Aggregators. The Indian government has floated a discussion paper on 'Digital Empowerment Protection Architecture' for providing data and consent intermediation services in all sectors.[32] The paper admits that while this can be more easily ensured in regulated sectors like finance, health or telecom, for unregulated sectors (for instance social media, e-commerce, education & jobs, etc.) incentives structures will need to be devised. How this will be achieved is not at all obvious. Even with Accounts Aggregators under regulation of the central bank, very limited kinds of informations get mediated through third party agents whereas the kind of personal data held by any firm on its customers today is multifarious and huge – a large part of it unknown to the data subject. Importantly, there exists no way for data subjects to protect, have access to, and/or partake in the value of, non personal data developed from their personal data, which definitionally requires collective approaches.

Economic governance of data of platform dependent business actors

Personal data protection laws do provide a good starting point for economic governance of data pertaining to

---

29 The relevant portion of the Draft Data Governance Act says; "In this context it is important to acknowledge that the rights under Regulation (EU) 2016/679 (GDPR) can only be exercised by each individual and cannot be conferred or delegated to a data cooperative." It is appropriate that individuals rights cannot be handed over to a collective entity. However, what is needed is a collective right over the top of, and not at the expense of, these individual rights. It is fine to rely on a collection of GDPR given individual rights if a data cooperative is dealing with a collection of personal data sets. However, an individual has no right to data that is anonymised across a group, which kind of common data is spoken of in the Draft Act as 'data pertain(ing) to several data subjects within ...(a) group'. What leverage does a group have vis a vis such common data, say, with a data collecting platform? It can only be through a collective right of the group concerned over such data. Without such a collective right, data cooperatives cannot function appropriately. To the extent that rights over such common can only be exercised by the concerned group coming together in a data cooperative – and cannot be exercised individually – an implicit collective right already exist in these formulations of the draft Data Governance Act. The disinclination to make any such collective right explicit is the main problem with current data governance regimes.

30 The key question is; can platforms refuse to deal with data cooperatives. If there is no rights based framing, they perhaps can. The draft Data Governance Act only enables persons or businesses to form a data cooperative, but does not enjoin a platform to have to deal with them. This is where an explicit right framing comes in, whereby they cannot refuse to deal with them. If a group has a collective right to do something, no action specifically meant to disable such a right will be legal.

31 cite

32 https://niti.gov.in/sites/default/files/2020-09/DEPA-Book_0.pdf

individual natural persons, by defining the primary rights holder for personal data. But for legal persons, like small business units, no such basic framing exists. With all sectors getting platformised, monopolistic dominant platforms hold enormous data pertaining to various platform-dependent 'small economic actors'.[33] Such data may be directly contributed by these small businesses, or created by their activities, or of their consumers, over the respective platforms. Much less such data pertaining to dependent small business actors provide them economic benefit, it often gets employed to their detriment; as platforms use it to develop competing goods or services of their own, or to otherwise rig the bargaining equation between platforms and them.

Uber drivers have tried to leverage data access rights under the GDPR for economic purposes. But the regulation being basically about privacy protection, the results have not been very encouraging.[34] As individual actors, Uber drivers can at least attempt to extend GDPR's use for obtaining their economic rights. But what about small traders on Amazon – in a similar situation of data-based exploitation as Uber drivers – who may have legal and not natural personhood in relation to Amazon? They cannot use the GDPR, and have no other recourse. This is a discordant situation of different rights and protections for similarly placed economic actors.

In a study undertaken by the Competition Commission of India, restaurant owners complained how food delivery companies did not share with them data pertaining to consumers of their products, thus disadvantaging them in major ways.[35] Food delivery platforms employ such data to set up their own production facilities, often called cloud kitchens, in direct competition to the restaurants whose data they initially rely on. EU's competition regulators[36] as well as US's Federal Trade Commission[37] are investigating whether and how Amazon uses data pertaining to merchants who trade goods on its platform for developing competing products.

Competition law based *ex post* remedies applied to data-misuse by platforms tend not to be very effective, as discussed in the next section. Further, platforms not only use data pertaining individually to a 'business user' to adversely affect latter's interests, they also employ data aggregated across many similar actors in an collectively adverse manner. This may be in the form of developing competing products, or otherwise disadvantaging business users in their bargaining power *vis a vis* the platform, as the latter develops vast and deep intelligence based on such aggregated data. Therefore, not just an *ex ante* approach is required but also a collective one *vis a vis* all similarly placed small economic actors, and the aggregated data pertaining to them.

The clearest and strongest approach to the issue of data about business actors or users has been taken by the recent draft Digital Markets Act of the EU. It does three things. It extends data portability rights to 'business users' of platforms, which may be legal persons and not natural persons. It prohibits data pertaining to business users – both individual and aggregated kinds – from being employed in ways or activities that compete with them. And, third, it gives business users access to both individual and aggregated data pertaining to them that is held by the platform.

---

33 A formulation we employ to represent all platform dependent economic actors, especially small ones with little or no bargaining power vis a vis the might of giant digital platforms who closely govern and control their economic activities..

34 https://www.personneltoday.com/hr/uber-drivers-personal-data-gdpr/

35 cite

36 cite

37 cite

Digital platforms are prohibited from employing data contributed by, or generated from the activities of, their business users, or of the consumers of the latter's services, from being employed in competition to such business users. Importantly, such a prohibition includes aggregated as well inferred data arising from the stated kinds of data. This far-reaching regulation can put a spanner in the wheels of the dominant platform business model. This is especially so because aggregate and inferred data are covered in the prohibition, which should include all intelligence derived from the described data classes – i.e."data which is generated through activities by those business users, including by the end users of these business users". Does it mean that, for instance, Uber cannot employ intelligence built from operating its fleet of drivers run taxis to develop autonomous taxi services, which apparently is where it plans to head?[38] Similar kinds of automation are intended by most platforms.

Unfortunately, no prohibition or check has been imposed on platforms using data pertaining to platform dependent 'small economic actors' (a term we prefer to 'business users') to harm their bargaining power, other than employing it for competing purposes. Dominant platforms do so all the time in various ways, as they accumulate huge intelligence built from data "generated through activities by ... business users".[39] Such data may be data about an individual businesses as well as aggregate data across a number of similarly placed dependent small business actors. Developing and employing such digital intelligence is central to the business model of platforms and their ever growing economic power.

The draft Digital Markets Act also creates a new kind of data access right for dependent business actors. Business users must be provided "free of charge, with 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 the products or services provided by those business users". This is a far-reaching new economic data right for small dependent business actors. Individual businesses can employ such data, especially that about consumers of their products and services, to better develop their products/services. This new right can also improve their bargaining power and reduce dependency *vis a vis* respective platforms. This is because they not only now have direct access to useful commercial data pertaining to them, they can even switch platforms taking such data along.

What is significant here is that platform-dependent actors can also access and use their aggregated data. It is unclear how this could operate as an individual business's right, because aggregated data presumably includes such data that may be shorn of its identification with individual businesses. Accessing such aggregated data by definition means also accessing data of other similarly placed businesses whose data together constitutes such aggregates. This access right therefore appears to be a kind of collective data access right of similarly placed 'small economic actors'. We discussed how such a collective right to aggregated data is also implicit in the concept of 'data cooperatives' in EU's draft Data Governance Act. Data cooperatives are supposed to help persons as well business users of platforms exercise collective control over their data. A combination of the enabling form of 'data cooperatives', and the collective right to aggregate data implied in both draft Digital Markets Act and draft Data Governance Act, can go a long way in addressing the current unsustainable power imbalance between platforms and the dependent 'small economic

38 https://www.cnbc.com/2020/01/28/ubers-self-driving-cars-are-a-key-to-its-path-to-profitability.html

39 Even when providing products or services in competition to their business suers is expressly banned, for instance for foreign-funded e-commerce companies in India, that does not seem to have any great effect on platform's economic power. This shows that they intelligence based ever growing bargaining and exclusion power is more important, which is not addressed by the draft Digital Markets Act.

actors'.

As to why data access and use rights – both individual and collective – are not explicitly stated, Data Governance Act does make it cleat that it aims "at facilitating data sharing including by reinforcing trust in data sharing intermediaries', but "does not aim to grant, amend or remove the substantial rights on access and use of data". This "type of measures is envisaged for a potential Data Act (2021)". Hopefully, the proposed Data Act (2021) will expressly allocate the required individual and collective economic rights to data both for natural persons and 'small economic actors' as legal persons.

Economic governance from a competition point of view

A worldwide clamour for regulating Big Tech has resulted in even the erstwhile adherents of self-regulation accept some role for competition authorities in this matter. Since the key data-related issue here is anti-competition potential of data hoarding, economic governance of data can be subsumed in competition regulation. Without going into details of such possibilities and the efforts being undertaken in this regard, this section briefly visits their limitations.

A fundamental conceptual problem is that competition law considers two or more entities operating in a given 'relevant market', one of which is considered dominant. Platforms and their dependent 'small economic actors' do not so much operate in a market, as platforms provide the market in which dependent 'small economic actors' operate.[40] It may be impossible to effectively apply 'industrial age' competition law to such a new digital economy institutional framework centred not on markets as much as on sector-spanning platforms.[41]

Competition law remedies come in *ex post* after sufficient social loss has been demonstrated. Even then such interventions may address specific implicated actors, rather than deliver more broadly applicable economy-wide corrections. Competition law must indeed be updated to give attention to the role, if not centrality, of platforms and data in new economic structures. Many jurisdictions have undertaken or are considering such changes.[42] But even with the best and most updated competition regime, *ex ante* regulation, which could be in the form of data rights, structural separation, and so on, will need to complement *ex post* competition law. The draft Digital Markets Act observes:

> Regulation and competition enforcement already coexist in other sectors, such as energy, telecoms or financial services. The Digital Markets Act addresses unfair practices by gatekeepers that either (i) fall outside the existing EU competition control rules, or, (ii) cannot always be effectively tackled by these rules **because of the systemic nature of some behaviours, as well as the ex-post and case-by-case nature of competition law.** The Digital Markets Act will thus **minimise the harmful structural effects of these unfair practices ex-ante**, without limiting the EU's ability to intervene ex-post via the enforcement of existing EU competition rules. (Emphasis added.)

The German competition law is most advanced in this area, with newly included data related understanding and provisions. Regarding these most advanced data related competition provisions, an observer notes:[43]

40 cite
41 Commonwealth paper, quote parts and p no.
42 Germany, India
43 https://www.econstor.eu/bitstream/10419/190945/1/104386220X.pdf

> On the one hand, “an undertaking which holds a dominant position has a special responsibility not to allow its conduct to impair genuine undistorted competition in the internal market” (Court of Justice of the European Union, 2011, para. 24). Accordingly, a dominant company can generally be forced to grant access to data or information. On the other hand, two preconditions must be fulfilled for forcing a dominant company to grant access (Bundeskartellamt, 2017, 7):
>
> ▪ the access to the requested data has to be important for economic success in a market and
>
> ▪ competing or not competing companies in the same or different markets are unable to buy or collect a set of data that is similar, or at least as useful (natural monopoly). “
>
> However, **both preconditions are generally not fulfilled at the same time** (Demary/Rusche, 2018, 52 ff.). In essence, the **provision of data by using antitrust regulation is not very likely** because for every single case, a company has to be deemed dominant because of data, it has to be proven that the data is necessary for success **and that there is no other source of suitable data.** Furthermore, t**he long duration of proceedings** also plays a role, especially in the dynamic digital economy. For example, the proceedings in the Microsoft case ran from 2004 to 2007 (Court of Justice of the European Union, 2007). (Emphasis added.)

The above lays out very well the limitations of relying exclusively on competition law for the necessary economic governance of data. As we discuss later in the section on data infrastructures, natural monopoly situation – the second condition above for mandatory data sharing – actually does apply for important sectoral data. However, first, data as infrastructure needs to be understood in its special nature vis a vis physical infrastructures. And second, the ‘natural monopoly’ character has to be seen *ex ante,* and as systemic and across a sector – which is not the manner competition law typically looks at issues.

Economic governance of data from industrial policy perspective – use of property rights

The earlier sections examined economic issues around data from a micro perspective of individuals and firms. A major concern today is the extreme geo-economic concentration of the digital or data economy. As per UNCTAD’s Digital Economy report, 2019, 90% per cent of the market capitalization value of the world's 70 largest digital platform companies is accounted for by the US and China. All other countries, including developed ones like in the EU, legitimately fear that they may soon be reduced to digital colonies – contributing raw data to US and Chinese digital industry, and consuming their digital products and services. Such a concern is very visible, even paramount, in the digital and AI strategies of these countries, whether of UK, France and India, or of the EU region. In their central objective of developing domestic digital and AI industry, these strategies embody a ‘digital industrial policy’ approach.[44] They are nearly unanimous in their recognition of exclusive hoarding of their society’s data by a few US and China based digital corporations as the key problem to be urgently addressed.

Two approaches have been attempted to address this main problem. First is to provide some new kinds of property rights over data for domestic businesses, which could tip the data balance in their favour. Second is to ensure greater sharing of data across the economy, so that the domestic industry gets access to data

44 cite

currently lying largely with foreign global corporations. This section and the next respectively deal with these two kinds of digital industrial policy efforts.

Two EU initiatives have been based on employing property rights for data. Early, in 1998, EU's Database Directive proposed a *sui generis* property right over certain kinds of data, in a context of general acceptance that data being a representation of facts is not copyrightable. It was hoped that in being the first to provide property rights over certain kinds of data, and explicitly recognise in law copyright over original databases, Europe can attract and foster a stronger computing and digital industry, which was beginning to centre on databases. How far it has succeeded can be assessed from this passage in a paper:[45]

> In 2005, less than ten years after it was introduced at EU level, the European Commission published its first review of the Database Directive, a remarkably self-critical assessment. According to the Commission, "[t]he economic impact of the "sui generis" right on database production is unproven. Introduced to stimulate the production of databases in Europe, the new instrument has had no proven impact on the production of databases". The Commission's report also suggests that the sui generis right has not helped the European industry to overcome its productivity gap vis-à-vis the United States. It points to several other deficiencies of the sui generis right, such as its uncertain contours, and its proximity to a property right in data that might negatively affect innovation and growth. The report juxtaposes the legal situation in the EU with that in the United States, where since the Supreme Court's landmark Feist decision no legal protection for 'sweat of the brow' based databases exists. Nevertheless, as the Commission wryly observes, "there has been a considerable growth in database production in the US, whereas, in the EU, the introduction of 'sui generis' protection appears to have had the opposite effect."

In any case, it is now highly contested whether Database Directive is even applicable to platform-generated data and machine-created data, which constitutes the bulk of big data at the heart of a digital economy.[46] This further erodes the relevance of data rights carved under the Database Directive.

The failure of Database Directive to deal especially with what the EU considers as very important category of machine generated data and 'industrial data'[47] propelled another effort at employing property rights to favour EU's domestic industry. The first decade and half of the new millennium cemented complete dominance of US digital and data companies; at which stage EU became most concerned with siding with its industrial giants on the data issue – shifting its focus to 'industrial data'. A new idea of 'data producers rights' was floated in a 2017 policy document 'Communication on Building a European Data Economy'. It was mostly motivated by a fear that EU's strong auto-mobile industry will be overrun by data-based auto-mobile intelligence services from the US.[48] The latter get built and sustain upon data collected globally, including from Europe. Laying the new data battle-lines as being between those who own and run data-generating machines and data-fed platforms that rendered these machines intelligent, a 'data producers right' was to accrue to the owners or operators of the machines that generate the data in the first place. By moving primary economic rights to data upstream, closer to points of data's origination, EU expected to be able to stand up to to the digital might of the US and its global corporations.

---

45 https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf

46 No investment in 'created data' (CJEU) ..

47 cite

48 cite

The proposal for a data producers' right seems to have been shelved now. It had many problems, like creating additional rights over copyrightable material, and thus *inter alia* adversely affecting important exceptions to copyrights.[49] Data producers' right accrues to the owner of the equipment that generates the data. This leads to significant issues in a digital society context where ownership of digitally enabled equipment can be shared, and unclear, between equipment manufacturer, software licence provider, and the actual user of an equipment.[50] As a full property right, which is exclusive and transferable, it can cause great uncertainty in a context where various kinds of data get used by different actors in multiple ways. While providing economic leverage to some actors, which were hoped to be EU domestic companies, such an exclusive right militates against the need for economy-wide access to, and sharing of, data in the EU.

A third generation of economic rights to data for European businesses are now being shaped. There are three important changes in this new approach. The accent now is on non-exclusive 'right to access' – although the use of 'rights' terminology is yet very muted, if at all – and not old-fashioned exclusive property rights. Second, with platformisation of everything – including manufacturing, the focus shifts to the relationship between platforms and their 'business users' in all sectors (and not only industrial users). Third, and problematically, with discarding of the idea of 'property', the term 'right' has also disappeared, or is very muted. It is discordant that a shift of focus to protecting smaller economic actors rather than bigger ones (like it was through Database Directive or data producers right[51]) has meant abandoning rather than strengthening of an economic rights based framework. To be fair to EU policy makers, this may just represent the lack of needed conceptual evolution – to meet the requirements of this new and fast changing area – and the political will for it. Property rights around data followed easily from an intellectual property rights tradition. No such clear conceptual and jurisprudential lineage has apparently been found, or policy-makers have been innovative and/or bold enough to consider, for a right to access data.

The draft Data Governance Act should have been the place to develop and present such a right.[52] But it chose to take only an enabling approach and not a mandating, rights-based one. Fortunately, as discussed in earlier sections, the draft Digital Markets Act provides an implied individual and collective right to business users to access their individual as well as aggregated data with platforms. Employing the enabling framework provided by the draft Data Governance Act for 'data cooperatives', such an implicit 'access to data' right can be used to fundamentally transform platform businesses. For example, drivers on Uber platform in a city can form a 'data cooperative' (under Data Governance Act provisions) and seek collective access to and use of their aggregate and (relevant) individual data with Uber (employing provisions of Digital Markets Act). They can thus control how Uber uses their collective data, thus completely transforming their collective bargaining power. Such is the power of whoever controls the data behind a platform that such a 'data cooperative' of Uber drivers, or Amazon traders, would practically come to co-own the respective platform businesses.

Instead of having to adopt such cut-paste strategies, what is needed is an explicit general framework based on individual and collective right for platform dependent 'small economic actors' to access their data. As for what 'their data' is, the draft Digital Markets Act defines it very well as "data which is generated through activities by ... business users, including by the end users of these business users".

49 cite
50 cite
51 Trying to respectively favour EU's relatively large digital companies and EU's large industrial companies.
52 As noted earlier, Data Act (2021) promises to do so, and needs to be looked out for.

Economic governance from industrial policy perspective – focus on data sharing

AI is considered as the next industrial frontier by digital and AI strategies of numerous countries – UK, France, Germany and India, as well as EU's. There is much concern that digital and AI power is getting concentrated at two global poles, in the US and China, and that everyone else will lose out and become dependent. Colonisation, and industrial age dependence, arose from outsourced industrial production. Digital age dependence and lock-ins involve outsourcing of data-based intelligence of a country's social and economic systems, and are expected to be much worse. The main objective of the EU's digital policies is of "ensuring the EU's leadership in the global data economy"[53] whereby "Europe become(s) the world's number one data continent."[54]

National digital and AI strategies converge on one key requirement for digital industrialisation[55] – wide spread access to data for domestic digital business.[56] This requires economy-wide sharing of data by those who collect and hoard data. By far most of society's data gets collected over digital platforms that provide different kinds of intermediation and services. These platforms are owned by a handful of global digital corporations. A lot of data also does exist with small or medium businesses, but not nearly at the scale of that collected by sector-wide platforms. The key digital industrial policy question then is; what can ensure that data holders – especially the largest data holders – share their data? This is not easy given that data is considered one of the most, if not 'the' most, important resource and element of competitive advantage.

Let us first review what kinds of data sharing already take place without any programmatic, policy or legal intervention. Corporations often share data, and facilitate others to do likewise, when such pooling of data benefits their business interests. They may share data because it supports extensive distributed research in areas where rapid technical advances benefit them. Autonomous automobiles training data has been shared with such a purpose by Google's sister company, Wamyo,[57] and by car manufacturers like Toyato[58]. Such data sharing is restricted only for non-commercial purposes. It helps researchers undertake visual imagery based AI research, which is critical to development of autonomous cars. Second, corporations may pool data among themselves when it is to mutual commercial benefit, but generally with conditions about contribution, usage, etc., and excluding non-contributors. Baidu, the Chinese corporation, runs such a pool of autonomous automobiles training data, the 'Apollo Open Platform'.[59] The data-pooling platform owner is mostly the biggest beneficiary in such cases, because participation may involve various kinds of lock-ins, like adopting Baidu developed standards etc. Dominant corporations may undertake a third kind of sharing, that can overlap with the above two kinds, where data is contributed with an altruistic purpose. Uber's Movement and Facebook's Data for Good initiatives are two examples of such 'data altruism'. But, the kind of data that gets shared is carefully filtered through a competitive advantage lens.

None of these kinds of data sharing serves, to any substantial extent, the industrial policy objective of promoting domestic digital companies *vis a vis* a few monopolisitic global corporations. That dominant global digital corporations holding by far most of society's data would not voluntarily share it – certainly not share the really important data – with their, existing or potential, domestic competitors, should indeed require

53 https://ec.europa.eu/digital-single-market/en/european-strategy-data
54 https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2102
55 A term employed here for adequate development of domestic digital and AI industry but not used in these country strategies. Also see...
56 Data more important to startups than capital … Oxford prof – the economist
57 https://www.infoq.com/news/2019/09/waymo-machine-learning-dataset/
58 https://news.mit.edu/2020/mit-toyota-release-visual-open-data-accelerate-autonomous-driving-research-0618
59 https://apollo.auto/docs/promise.html

no labouring.

Another kind of data sharing involves not giant platform companies but data that remains, or can remain, with smaller domestic businesses. Such sharing does have digital industrialisation potential, and a corresponding role for industrial policies. EU is currently most focussed on this kind of data, and its sharing. But it forms a small part of society's data, as compared to that collected by big platforms. EU is pinning its hope on the next phase of digitalisation where IoT and industrial data will dominate, and where instead of the cloud such data will mostly be processed at the edges, nearer the points of its origin.[60] This may be a valid projection, and it is useful to aim at domestic pooling of such data towards a possible domestic industrial advantage.

Two serious limitations however exist in taking this argument too far, and making it the central plank of a digital industrial strategy. One, it is based on an inadequate understanding of platformisation of all sectors and every kind of economic activity, including manufacturing. Digital platforms structurally transform economic organisation. In the process, they take up key data controls, employing such data for developing increasing intelligence to further re-organise economic systems and control them. In platformising businesses, especially large industrial units, dominant digital corporations do face a different kind of actor on the other side, compared to consumers-facing platform services. It is not just a hapless individual, but a relatively mightier business. The trajectory of digitalisation however shows that data controls will still shift towards platforms. The manner in which European car manufactures are clustering around autonomous mobility platforms like of Alphabet's Waymo (from the US) and Baidu (Chinese) is a good case in point. It is not just a matter of more data being processed closer to edges, or the points of its generations, and not in the cloud. What is important is who controls such data wherever it might physically exist. The emerging digital mobility paradigm seems to favour more of such control being with dominant digital platforms rather than traditional car manufacturers. Policy interventions like the implicit new 'data access right' created for business users of platforms by the draft Digital Markets Act can help shift this balance to some extent. But much more work is required in this regard.

A second problem holds even if successful data sharing arrangements do get created for data with business users that does not get collected by digital platforms. It is not clear whether this too will not benefit dominant platforms more, as an additional source of data that they could not otherwise collect, rather than help create domestic competitors to such platforms. Here a third problem specific to developing countries may also be stated. This strategy of EU focussed on 'industrial data' is premised on its advanced industrial base that is comparable to that of the US. Such industrial conditions do not obtain in most developing countries.

Most of the work on voluntary data sharing arrangements in the EU has been at a programmatic level ,whether it is 'data trusts' in the UK,[61] or Industrial Data Space (now extended to data spaces in other sectors like transport, environment, agriculture, energy and health)[62] and GAIA projects[63] in the EU. Such projects now get provided an enabling framework by the new Data Governance Act, ensuring trusted and neutral data intermediaries, but without any mandatory or rights-based provisions.

The principal objective in these projects is to give maximum control in the hands of data provider.[64] They

60 cite
61 cite
62 cite
63 https://ec.europa.eu/digital-single-market/en/news/stakeholders-dialogue-common-european-data-spaces
64 Also called 'data sovereignty'. GAIA docs

enable robust institutional frameworks for data markets as well as other forms of data sharing. A data markets framework has been provided in India's NITI Aoyog's AI strategy.[65] A new proposal from India's Ministry of Electronics and IT for National Open Digital Ecosystems inter alia seeks sector-side data sharing with appropriate governance frameworks.[66] It is not clear what prevents global digital corporations from employing these new voluntary data facilities to now also collect data otherwise inaccessible to them, and further consolidate their data advantage.[67] If a matter of buying data off data markets, they have the most financial might. Even if some conditions are put for data sharing, their numerous and enormous entrenched advantages will ensure that they corner a net positive gain from any such arrangement. It does not look likely that such voluntary projects will be able to access the most important data with digital giants.

History testifies that just having infrastructure does not mean industrialisation. British Raj infrastructure in India for movement of resources and goods – railways and ports – was developed to denude India's resources, further de-industrialise it, and ensure its economic colonisation. In default of other, stronger, measures for digital industrialisation, a well-developed infrastructure for data markets and voluntary data sharing can similarly be employed by global digital corporations to further consolidate their data, digital and AI advantage. Developing countries face an even higher danger of this happening.

No credible path is presented of how voluntary data sharing arrangements by themselves will help create, say, European digital corporations strong enough to compete with US and Chinese digital giants.[68] The incumbent global digital corporations can not only turn such data pooling to their advantage, the main requirement of getting wider access to all the data collected by them remains unachieved. EU's Data Strategy says that in specific situations compulsory data sharing may be sought, something the proposed Data Act ((2021) will address. It is expected to clarify "data usage rights in business to business and business to government settings".[69] It will also have to lay down the appropriate basis and conceptual framework for such compulsory sharing of data and data usage rights.

Academic discourse has ventured beyond voluntary sharing, recognising its limitations to address the central digital society problem of a few global corporations collecting and controlling the bulk of world's data. Two such proposal are for FRAND (fair, reasonable and non-discriminatory) based data sharing as happens with standards essential patents,[70] and compulsory licensing[71] on the model of health emergency related access to IP protected medicines.

These no doubt are interesting directions to explore. Both these approaches get applied on the top of, and as exceptions to, well established intellectual property rights. But there are no clearly recognised IP rights for data. The problem with using these approaches – created to meet exceptional conditions – as the mainstream way for data sharing is that they carry with them an implicit acceptance of some kind of property rights of data holders over data with them. Even the 'data sharing services' provision of EU's draft Data Governance Act refers to 'data holders' providing 'their' data to such services. Such ham-handed efforts towards data sharing may have the opposite effect of legitimising the dominant de facto data ownership model. Once the latter is entrenched in popular and legal understanding, it becomes much easier for giant digital corporations

65 cite
66 cite
67 Submission to NODE paper
68 https://www.wired.co.uk/article/eu-tech-data-industrial not clear what the end game is
69 https://techcrunch.com/2020/11/24/europes-data-strategy-aims-to-tip-the-scales-away-from-big-tech/
70 See this paper https://link.springer.com/article/10.1007/s40319-018-00777-7 and US vehicle safety data related
71 Kerbs paper

to resist ‘inappropriate appropriation’ of ‘their data’. This will reduce data sharing, if at all, to some peripheral cases, making no significant dent on the originally identified problem of global data hoarding. The extent, and the systematic manner, of sharing of data with the domestic industry that is needed cannot be achieved in such manners.

Adequate data sharing that can ensure sufficient domestic digital and AI industry requires mandatory data sharing provisions of various kinds. Set in a rights framework, mandatory data sharing is nothing but a collective right to access data. What data, who can access, under what conditions, and on what basis, meanwhile, are important questions that need to be addressed. This calls for a comprehensive framework for collective economic rights to data, which is the subject of discussion in Part 2 of the paper.

Part 2: A community rights framing for economic governance of data

Whose data it is – community as data subject

Data’s greatest value lies in its collective form, which underpins most data analytics and AI. The main solution to breaking global data monopolies has been identified as society-wide data sharing. An effective framework for economic rights to data therefore must include collective rights. In Part 1 we examined how most efforts at economic governance of data, especially as emerging in the EU, fail, or are likely to fail, in sufficiently meeting their objectives because they do not go beyond individualist-property approaches. This second part of the paper attempts a shift towards the required collective rights approach to economic governance of data. It takes its inspiration from the draft report of India’s Committee of Experts on Data Governance Framework.[72] The report states as its basic premise that ‘a community has the right to its data’.

It is worth examining the typical data value chain and the various elements and actors involved in it. First is the actor or thing whom certain data is about, that can be called as the subject of data. Then is the collector of data, who encodes and gathers information about the subject in a machine-readable form. Between the subject and the collector, some analyses pose a ‘data producer’ as the owner of device that encodes data.[73] Downstream from these actors are various processors of data that do combining, analysis, etc. of data to produce insights and intelligence, and further down are businesses that deliver products and services based on such intelligence.

The most important binary in a data value chain is that whom data is about and data describes (data subject) on one side and the collector of data on the other. *De facto* rights on data’s economic value currently vest with data collectors, and for most data they are various digital platforms. Legal attempts have been made to shift the rights balance more towards the data subject. We saw how GDPR and other data protection laws give some implied economic rights to data subjects. The draft Digital Markets Act provides similar rights to platform-dependent businesses on data that is generated through their activity on platforms, which data thereby describes them and their activities. Such businesses too can be called as data subjects for such data. The Digital Markets Act though holds back from describing its provisions as providing a right to access data. It also refrains from calling the implied ‘rights bearers’ as data subjects, when they are businesses and not natural persons. It calls such legal persons just ‘business users’ of platforms. This seems to be because, as a market regulation framework, the Digital Market Act is solely interested in setting a regulatory frameworks for platforms vis a vis its users. It does not get into introducing data access rights. This is proposed to done by a Data Act, 2021. One will have to wait for this Act in 2021 to see if the bearer of data access rights could

72 cite, and disclaim

73 The problems with attempts to give property like rights to data producers was discussed in an earlier section. This section will ignore their claims altogether.

also be a legal person, and whether such legal persons will be described as data subjects – and such subject-hood presented as the legitimate basis for their rights.

Meanwhile, we consider the term 'data subject' much better than 'business users' for the purpose of being allocated a right to access 'data about it'.[74] Legal persons do not have human rights, only natural persons do. A legitimate question is whether legal persons should have economic rights vis a vis data. In this regard, human rights is not the only framework of reference here, there is also the intellectual property rights regime under which corporations claim data related rights. Small businesses also need appropriate explicit data rights in face of data's de facto ownership by data collectors, as well as latter's various claims under IP frameworks. The proposition that primary economic right to data are held by data subjects – whether natural or legal persons – shifts the balance of digital and data power towards smaller economic actors or units – both individuals and small businesses.

There may be good basis to posit that primary economic rights over data should cohere in the data subject, whether natural person, or a business entity, and correspondingly not be with the collector of data. We deliberately speak of 'primary economic rights', and not ownership, recognising that these rights are generally not exclusive. Multiple actors may have a stake in any dataset, and can legitimately lay claim to at least some uses of such data. Exclusive ownership rights may freeze or debilitate, or at any rate put great uncertainty into, the entire data value chain. Fixing primacy of economic rights however is necessary to clarify who holds entitlement and decision-making power in case of overlap or conflict.

Providing economic rights for individual persons, natural or legal still leaves an important problem unaddressed. A great amount of very valuable data exists in forms that are shorn of individual subject identity. Individual rights frameworks definitionally cannot deal with such data. Data collectors may even anonymise more and more of the data they hold just to escape individual rights frameworks, something which they already do. This leaves an enormous gap in the role of economic rights to ensure appropriate economic distribution. Economic rights must also extend to aggregate or group data. These rights can only be collective rights – such data being not possible to be associated with any particular individual person, natural or legal.

The draft Digital Markets Act enjoins against group or aggregate data bring used for causing harm of unfair competition to its subject 'business users' of platforms.[75] The draft Act also provides subject 'business users' an implicit right to access and use such aggregate data that is generated by their or their consumers' activities on a platform. Unfortunately, for natural persons no legal rights seem to yet exist anywhere with regard to their aggregated anonymised data – whether of harm prevention kind or for access and use. The absence of explicit collective rights for data subjects regarding aggregate data about them is a major anomaly in data rights regimes. Even when data is individually identifiable, collectivisation among data subjects may be required for effective bargaining because of the huge power difference between them and data collectors. The EU's draft Data Governance Act, 2020, provides an enabling framework for such collective strategies through 'data cooperatives', but without any specific collective rights.

It remains unclear how the right provided by the draft Digital Markets Act to business users to access their aggregate data on platforms can be exercised other than collectively by the data subject group – all business

74 Definition of subject …

75 The draft Act does not use the term subject for business users, but we are using it as per the preceding discussion on the relevance of this term in allocating primary economic rights.

users together whose aggregated data it is.[76] A community of natural persons should similarly have the right to access aggregate data generated over platforms from the activities of its members. Groups and communities whose activities generate a given aggregate data set can be considered the collective subject of such data. Indian policy documents have called data pertaining to a community of natural persons as 'community data'.[77] The draft report of India's Committee on Data Governance Framework explicitly speaks of a community's right over its data. It does not however deal with groups of legal persons, as done by the EU's draft Digital Markets Act, which in turn does not take note of communities of natural persons.

Primary rights to a community's data (or, community data) should logically accrue to the respective community as its collective subject. But many remain unenthused to explore this line of argument because it is not easy to imagine the appropriate actor or agent that can effectively exercise community rights. While certainly much more complex than for individual rights holders, community rights over different kinds of resources are not unknown. They are well-established in many realms like forests, water bodies, traditional knowledge, flora and fauna, and so on. Building over such precedents, the collective resource of common data can also be placed in a community resource governance framework. This will involve developing necessary rights, obligations, laws and institutions, which enable as well as constrain various actors.

Even as the community gets the primary economic rights to its common resources, community governance frameworks duly consider the needs of all stakeholders involved in any resource system. In order to minimizes uncertainty for everyone, existing and default practices of resource access and use may be impacted only to the extent of specific explicit provisions. Many actors may have usufruct rights – defined as the right to use and enjoy the fruits arising from a resource system but without unduly destroying or diminishing it.[78] In community data governance, such usufruct rights can have useful application vis a vis the rights of data collectors and processors.

Community data consists of anonymised data or otherwise group data of people in a community, and data pertaining to various community owned or associated things (like public infrastructure and natural phenomenon). As describing the community and its activities, such data provides valuable intelligence about it. This intelligence can be employed to harm or benefit the community. The community should have the right to control uses of such data about itself – to prevent harm, and to maximise benefits for itself.[79]

It is in this sense that the Indian Committee on Data Governance Framework asserts that a 'community has the right to its data'. Implied here are primary economic rights, which are non-exclusive. They do not, in any significant manner, effect the rights of other legitimate stakeholders, like the collectors and processors of data, to use and benefit from such data. This is as long as such use does not carry the possibility of harm to the community. Being primarily its resource, a community has the right to allocate and facilitate preferential access to its community data, say for public/community interest purposes, and to its own businesses.

The Committee sees a community's rights over its data in two parts; prevention of harm to itself, and right to benefit from such data by making it available for public services and development of domestic digital

76 The language in the draft Act is unclear, but it appears untenable that every individual business on its own can access complete aggregate data of which its data is only a part – what happens if some others who are a part of a dataset have a different view and do not want to allow such access, or some specific uses following access? Even if one can access and use all such data freely, since this right is common across similarly placed actors, it can be considered as a collective right.

77 Srikrishna report, ecom policy

78 cite

79 If business users can be given a collective right to data that is generated by their activities on platforms, as the draft Digital Markets Act does, there seems to be no reason why a community of natural people should not have collective right to data generated on platforms from activities of the community and its members.

industry. The right against collective harm can be exercised by any non-profit body associated with the community through a complaint to the new Non Personal Data Authority to be set up. Exercising community's right to benefit from its data is a more complex and elaborate process. Considerable size and techno-organisational capabilities are required for a community organisation to undertake it, because data's value operates at scale, and secure real-time data sharing is a sophisticated processes.

The Committee outlines a special process for creating High Value Datasets (HVDs). It requires the consent of the new Authority, and involvement of enough stake and representation from the community concerned. HVDs are built from community data[80] that is mandated to be shared by all collectors of such data. The basis of such mandating comes from an expressly mentioned community's right to its data. Only data taken from community sources is mandated for sharing, and not data from sources private to data collectors. HVDs are to be managed by community trustees who can demonstrate sufficient techno-organisational capabilities to do so. Any entity registered in India can seek access to data with HVDs, which thereby act as a public good in the nature of data infrastructures. The Committee proposes an appropriate law to enable mandated sharing and other recommendations.

Government bodies may also get involved in the process of creating and sustaining HVDs. In all such roles, government bodies will act on the behalf of the community concerned, exercising its community rights in trust. They will be fully accountable and appropriately constrained as per the public trust doctrine. To keep track of who collects community's data, in what manner, and towards what purposes, large businesses – above certain thresholds – involved in data collection will have to register as 'data business'. Their processes of data collection, processing and use will have to made transparent.

Most recommendations of the Committee can be derived from principles emerging out of various existing governance frameworks for community resource in different areas. The Committee makes an explicit reference to such general principles for community resource governance. Before discussing some community resource frameworks that provide the basis for claiming and operationalising a 'community's right to its data', the next section examines how IP rights relate to community data. It is important to beforehand exclude any claims of private property rights over such data.

IP rights and community data

Copyright is available for human created things as due remuneration and incentive for creativity. Data being a reflection or recording of facts, it is almost universally accepted that there is no copyright on data. Data holders argue that they need intellectual property (IP) protection because collecting and curating data takes a lot of effort and investment. In its Feist decision, US Supreme Court refused copyright protection to data on grounds of 'sweat of the brow'.[81] However, originality of a database design can get copyright protection, though not the data contained therein. EU's Database Directive of 1998 affirmed copyright for original database design, and also provided limited *sui generis* protection to data itself if the condition of sufficient effort in compiling data is met. Subsequent court judgements in the EU have clarified that most platform- and machine-generated data would not meet such a condition.[82] Such auto-generated big data being the main bone of contention in a digital society, most consider *sui generis* protection provided by EU's Database

80 This term existed in an earlier draft of Committee's report but is not in the latest one. It perhaps is a part of staying away from defining categories of non personal data as attempted in the earlier draft. But with many descriptions of community and its data in the report, the term 'community data' is implicitly there.

81 cite

82 cite

Directive as already outdated in the digital age.[83]

Regarding copyright protection for database design, two issues are important to note. First, the threshold of originality for database design is fairly high.[84] Most database designs may be fairly obvious given the nature of data and the uses it is put to in mainstream digital society operations. Second, even if an original database design was to qualify for protection, extraction of data from such a dataset without copying the database design is not protected.

Community rights on raw data formatted in obvious ways therefore do not infringe any copyright provision. In mandated sharing, the Indian Committee only seeks extraction of data over pre-provided fields and structure, and not any original database *in toto* that a company may have creatively designed. No copyrights are therefore liable to be violated.

The IP right most mentioned by data collectors is trade secrets. India does not have any specific trade secret law or provisions. Some protections may however be provided on the basis of common law. Trade secret is a relational protection, for instance with respect to a company's employee or contractor. It does not provide any *in rem* property rights over its subject, meaning property rights that apply in relation to the entire world or for all actors. Any trade secret claim is therefore meaningless against lawful demand for sharing of data, as envisaged under the community rights framework.

Since India does not have a trade secret law, it is argued that mandatory data sharing may violate India's trade secrets related commitments under article 39 of TRIPS of WTO. Such an international commitment does not directly apply as law inside a country; it needs to be adopted in national law which India has not done. Still, below is a brief analysis of how trade secrets relate to the kind of data that get included under community data frameworks.

As per TRIPS, to be considered a trade secret, data should not be "generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question". The kind of data sought for sharing under community data framework may not be 'generally known'; because unlike open information about a community, data is often not generally known in 'forms' in which it exists with digital companies. The key issue however is whether or not such data is 'readily accessible' to others within the circles involving similar businesses. To understand the applicable meaning of 'readily accessible' one may refer to the corresponding formulation in US trade law which is 'readily ascertainable'[85]

'Ascertainable' means 'capable of being ascertained or found out'.[86] Community data is such data whose source and/or subject exists out there in the community, and access to which is public and generally unhindered.[87] This is almost definitionally so, as describing what all data comprises community data. Such community data and its publicly-accessible sources/subjects can be contrasted with what may lie inside

83 cite

84 cite

85 Unless WTO negotiators pushing for trade secret protections (which incidentally were resisted by India till the very end) among whom US negotiators were most prominent deliberately jacked up the threshold for WTO TRIPS over and above that for corresponding US law. This being unlikely, we may take it that 'easily accessible' may be read in the meaning og 'easily ascertainable'.

86 https://www.vocabulary.com/dictionary/ascertainable#:~:text=adj%20capable%20of%20being%20ascertained,determined%20or%20limited%20or%20fixed

87 Whether the subject of anonymised data of several individuals can be considered to be the community or networked individualism consisting of several individuals still in their respective private shells is an important sociological, philosophical and ideological questions that we will bypass here.

private spaces and domains to which access is accordingly controlled and constrained.[88] Dominant data companies often argue that they do not claim exclusive rights to such outside 'readily ascertainable' data; competitors are welcome to also collect it for themselves – but why ask for data that has been collected by them? Firstly, such an argument itself shows that community data is 'easily ascertainable', and therefore it cannot be treated as trade secret.[89]

But the question remains, being 'readily ascertainable', when others can indeed collect same or similar data on their own, why institute mandated sharing of community data?[90] This brings us to a different kind of social welfare maximisation and cost minimisation issue, that gets associated with natural monopolies. Natural monopolies are such facilities that are very wasteful to independently duplicate or reproduce by all competing businesses in a sector. There may not as such be anything very difficult, technically challenging, or innovative required to reproduce it, just a lot of expense. To avoid wasteful expenditure of a society's resources, such facilities may be retained as monopolies but as closely regulated (privately-run) utilities or publicly-owned infrastructures. A good example is that there generally is just one electricity supply company licensed for an area. A data related example is; detailed real-time data about the state of a given city road, in terms of the condition of the road, speed of traffic flow etc., held by a transport platform like Uber. It is quite possible for another similar business to independent ascertain and collect the same or similar data, but such data silos are a wasteful expenditure of society's resources, and sharing them maximises social welfare.

The kind of data that is considered 'community data' is relatively 'readily ascertainable', it being only a matter of sufficient effort. Society would not want wasteful duplication of effort and resources, whereby an data infrastructure approach in the form of shared community data is preferred. The situation may be different for data of which sources and subjects are exclusively 'inside' and subject to private realms, or bear other similar lawful constraints of access and ascertainability. The Indian Committee does exclude from its mandated data sharing remit such data that may not be considered 'easily ascertainable' (by other similar businesses) in this manner; for instance data about a firm's internal processes. Therefore, while trade secrets are not meaningful in the context of legal demands for access to data, the Committee's draft report does observe the spirit of trade secret law by mandating sharing of only such data which is 'out there' in the community and therefore 'easily ascertainable'.

There being no clear property rights in relation to data, private contracts are a major way for digital corporations to protect unwanted access to and use of their data. Public law, like the one under which demands for data sharing will be made in a community rights framework, trumps over any private law arrangements like contracts. There have been concerns about how mandatory data sharing requirements may affect India's vast and important business process outsourcing industry, a big part of which services data coming from abroad. The Committee's report clarifies that only the initial collector of data may be subject to data sharing requirements and not those who process data under contract with other parties.

---

88 Shopping malls and digital platforms are in this sense public, and largely unconstrained in accessibility, notwithstanding certain private rules that may apply to them. 'Publicness' is therefore more about open and generally available and unconstrained 'nature of an interaction', than strict ownership of the 'space where it may take place'. An interaction between a person and commercial platform is in this sense public, but not between two persons inside a private realm. This does not make the relevant data necessarily public, but subject to public rules, of privacy protection etc. It is the an interaction as the source of data that is judged to be public/community, and not necessarily the produced data itself.

89 There is a clear contradiction in dominant digital corporations arguing both ways; (1) that the same data can be collected by others and they do not stop them, and (2) that such data is their trade secret (which it cannot be if the data can readily be collected by others by putting in some effort).

90 Competition law asks this question. It does not consider data as a shareable essential facility if the data involved is available to be independently collected. See the discussion on the section of competition law.

Acquiring data for data infrastructures – eminent domain or community rights

General data about a society's characteristics, activities, etc., in a sector is of infrastructural nature for digital or AI businesses in that sector. Such data is increasingly necessary for any such business to function effectively. For every business to have to assemble the whole infrastructure anew for itself, and maintain it, is socially wasteful. The main reason for digital and AI concentration, globally, and within countries, is that a few first-starters quickly build the needed data infrastructure, out of freely collected data, and then monopolise it. Data being a dynamic, ever evolving, cumulative asset, growing very rapidly as the services-footprint of a firm expands, the data gap between the incumbents and any start-up becomes almost impossible to bridge.[91] Opening domestic competition to dominant mega-corporations – mostly foreign – will require providing essential data in any sector as infrastructure to all businesses on a non-discriminatory basis. (Positive discrimination in favour of small domestic start-ups and businesses is a different matter.)

EU's new digital policy document, 'Shaping Europe's Digital Future', states that "European technological sovereignty starts from ensuring the integrity and resilience of our data infrastructure'.[92] The EU, however, still only focusses on developing systems to enable pooling of voluntarily shared data. As we discussed at length, these efforts do not adequately touch the real hoards of sectoral data that lie with sectoral platforms, almost all of which are US or Chinese companies. Mandating these platforms to share sector essential data alone will lead to sufficiently provided, and thus meaningful, data infrastructures that can sustain development of domestic digital and AI industry.

Infrastructure is traditionally built by governments, even if sometimes managed under their oversight by private players. Physical infrastructures get built through huge financial expenditure. Data infrastructures however are different; even if a government had all the money to buy the needed data most of it is just not available on sale. Monopolising infrastructural data is the key competitive advantage of dominant global corporations, which they are not going to sell for any reasonable price. Further, data infrastructures are dynamic, requiring real time continuous feed of data collected from sites of various community activities, much of which comprise service interactions over private digital platforms.

Building the public data infrastructures needed to support a vibrant domestic digital industry requires all companies collecting relevant data to share it, on a continuous basis.[93] [94] Data fortunately being a non-rival good, the collecting companies do not lose the use of data that they share for data infrastructures. These companies, in any case, do not have any IP rights over such data. Since most of this data gets collected as a by-product of their service development and delivery, there is no special effort made. Mandatory sharing of data therefore would normally not result in any social loss of disincentiving and thus reducing production of the needed data.[95]

To the extent that collection of some kinds of data involves very evident special and committed investments, these costs can be appropriately reimbursed when mandating sharing of such data for building public or community data infrastructures. A company having many micro-weather data collection stations, or one

91 Currently, in this early phase of digitalisation, startups may still come up in some new unexplored service areas capturing new kinds of data which may be be available with dominant firms. However, as digitalisation proceeds and stabilizes such new areas will disappear and all key data infrastructures will be in the hands of a few dominant firms, or closed clubs of them.

92 https://ec.europa.eu/info/sites/info/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf

93 The other, less palatable, option will be for the data mines, the digital platforms, themselves being publicly owned.

94 Data infrastructures will also contain a lot of data with governments. We are keeping such public data out of our analysis throughout the paper because the paper's focus in on allocation of economic right to data in a society, and public data is a very different kind of special case, requiring altogether different analyses.

95 cite

employing satellites to collect land, crop patterns, etc, related data, are two such examples. Still employing the infrastructure model, such companies can operate as regulated utilities which would involve mandatory data sharing against remuneration, with some kinds of administered price, as happens for electricity generating companies.

Governments can mandate sharing of the data needed for public data infrastructures by exercising their power of eminent domain, which allows them to appropriate private property against due remuneration. The first issue with such an approach is that, as discussed at length, most data is no one's private property. More importantly, eminent domain acquisition will turn data over to complete government ownership and control. Its accountability in managing and using such data will only be as per normal norms associated with public functions.[96] Amidst strong fears, and world-wide indications, of the emergence of a data authoritarian state, state's unchecked access to a society's non-personal data can be dangerous. It is personal data that mostly gets discussed in this regard, but non-personal data provides a great amount of granular, real-time, general intelligence about people and groups, and various facets of society. Such intelligence can be grossly misused.

It is therefore much preferred that the data required for public data infrastructures be collected under the framework of a community's right to its data – the fundamental proposition put forward by the Indian Committee. This can employ and build on well-developed commons management thinking and principles (see Elinor Ostrom's work, for instance).[97] An appropriate framework for governance of community data can draw its inspiration and lineage from other community resource governance frameworks like for forests, water, and traditional knowledge. Even if government agencies directly get involved in data infrastructures – as they will need to in various forms, because data infrastructures operate at scale and require various kinds of intensive technical and other capabilities – they will hold the community data involved in trust for the community. These government agencies will be guided and constrained by the doctrine of public trust. It would involve strong community participation, accountability and responsiveness, for instance, in relation to preventing harm to a community, or how private parties access data infrastructures. The principle of subsidiarity[98] should be employed in how various national, state and local agencies, as also community organisations, get involved with data infrastructures. A community rights and governance framework will result in much more distributed and decentralised data infrastructures.

The Indian Committee recommends that data infrastructures involving 'High Value Datasets' (HVD) will be proposed, developed and maintained by appropriate trustees of communities whose data is involved in such data infrastructures. This represents a very significant shift from the industrial age tradition of public infrastructures to digital age community infrastructures. It implies a different kind of governance structure, openness, participation and accountability.[99] As large sectoral databases are increasingly decentralised and federated – the National Digital Health Blueprint[100] recommends a federated system for Indian health data involving decentralised data storage and processing at hospital, district and state levels – the corresponding HVDs as data infrastructures can also be governed in a decentralised manner. This will ensure adequate participation of relevant communities, enabling them to effectively exercise their collective rights.

96 See ref
97 Commons papers
98 What could be done and governed at a lower level of organisation may not be done at a higher level. Federated data infrastructures. Framework law for water India, appropriate body is the most local
99 Networked publics paper
100 cite

Governance of community resources and its application to data

> *Natural resource governance refers to the norms, institutions and processes that determine how power and responsibilities over natural resources are exercised, how decisions are taken, and how citizens – women, men, indigenous peoples and local communities – participate in and benefit from the management of natural resources.*
>
> Natural Resource Governance Framework
> of Commission on Environmental, Economic and Social Policy [101]

The International Covenant on Economic, Social and Cultural Rights recognises the right of all people to freely deal with their natural resource for their own ends.[102] The Indian Constitution in its directive principles of state policy lays down that the "State shall, in particular, direct its policy towards securing, (1) that the ownership and control of the material resources[103] of the community are so distributed as best to subserve the common good; and (2) that the operation of the economic system does not result in the concentration of wealth and means of production to the common detriment". The Indian Supreme court has clarified that; "material resources of the community in the context of reordering the national economy embraces all the national wealth, not merely natural resources... Every thing of value or use in the material world is material resource ...". [104]

Such material resources could be tangible resources like forests and water, or intangibles like spectrum and traditional knowledge. Traditional knowledge has been subject globally and nationally to governance frameworks very akin to those for natural resources. Since traditional knowledge is a social phenomenon, 'community resources' can evidently involve both natural and social commons. It can plausibly be argued that data about a community, and about things and phenomenon associated with it, also constitutes a 'social commons' of the community. And, that it should be governed like a society's 'natural commons'. Both traditional knowledge and 'community data' are products of socio-cultural processes, representing in them social and cultural patterns. It is the very nature of socially reproducing yet creative human civilisation that its socio-cultural patterns are both 'facts out there' (community data) and collective 'creations' that continually evolve (called community knowledge in their conscious forms). Community knowledge is explicit humans-embedded information, while data provides intelligence that, in the digital society context, is mostly autonomous and machine-based. Both evidently are a community's 'social commons'.

Within such a 'community resource' framing, legitimate private claims get admitted; like for such data that is not sourced from a community in general but from things, spaces, etc. that are private to the data collector. All governance frameworks for community resources provide multiple private claims and rights for various actors, which may or may not overlap. A traditional knowledge framework, for instance, may support private claims of families that are hereditary custodians of some traditional knowledge, like Ayurveda medicines.

India's Mine and Minerals Act sets up District Mineral Foundations as non-profit trusts for ensuring 'benefit

---

101 https://www.iucn.org/commissions/commission-environmental-economic-and-social-policy/our-work/knowledge-baskets/natural-resource-governance

102 https://www.ohchr.org/en/professionalinterest/pages/cescr.aspx

103 The manner in which courts have interpreted the term 'material resources' in 39 (b) to include "every thing of value or use in a material world" (State of Karnataka v Ranganath Reddy AIR 1978 SC 215), it can be considered to include informational or data resources, since data is certainly of immense value in the contemporary material world. Precedents for such interpretation also exists in international law.

104 cite

sharing' and participation in governance for communities associated with mineral resources. India's Forest Rights Act[105] provides 'community forest resource' rights to relevant communities for accessing and using various forest resources,[106] and participating in their governance. Joint Forest Management is an "'arrangement' between the village community (i. e. the beneficiaries), NGOs and the State Forest Departments", whereby "access and usufruct rights are limited to people who organise themselves as groups such as co-operatives or village forest committee and in no case shall access ... be given to individuals".[107] Under the Forest Rights Act and the Biological Diversity Act, relevant communities can make claims to resources through a formally laid out process. Informed prior consent of the community concerned is a frequently invoked concept in governance of these resources.

An official legislative proposal was made to the Indian Parliament in 2015 for a framework law for management of the commons resource of water. It recognised;

> Water as a Common heritage and Resource, held in Public Trust – (1) Water is the common heritage of the people of India, held in public trust, for the use of all, subject to reasonable restrictions, to protect all water and associated ecosystems. In its natural state, such as river, stream, spring, natural surface water body, aquifer and wetland, water is a common pool resource, not amenable to ownership by the state, communities or persons. (2) The state at all levels holds water in public trust for the people and is obliged to protect water as a trustee for the benefit of all: Provided that the responsibility of the state as public trustee shall remain even if some of the functions of the state in relation to water are entrusted to any public or private agency.

The proposed law provided for participatory management of the commons of water through Water Users Associations.

Developing countries have been voicing the need at global fora for protecting their traditional knowledge embedded in cultural practices of various communities. Some attempts have been made to fit IP frameworks over this very different context.[108] As mostly involving collective cultural practices, traditional knowledge is very unlike typical subjects of IP rights. The main objectives of protecting and promoting traditional knowledge are also quite different. For these reasons, many countries have developed *sui generis* frameworks for traditional knowledge best suited to their contexts and needs.[109] The Intellectual Property Rights Policy of the State of Kerala (2008), in India, incorporates one such framework.[110] It asserts that all traditional knowledge, including traditional medicine, the practice of which sustains livelihoods, must belong to the domain of 'knowledge commons' and not to the 'public domain'. This means that there cannot be unchecked use and exploitation of the resource of traditional knowledge. Such uses will be controlled by the community to prevent harm, and for its own best benefit. The policy recognizes private claims of families closely associated with some kinds of traditional knowledge. All traditional knowledge will be held under commons licenses, taking cue from open source software practices. Such knowledge will be available for free for non-commerical use, against attribution, with the condition that any development made by using

105 Scheduled Tribes and Other Traditional Forest Dwellers (Recognition of Forest Rights) Act, 2006

106 These rights are of usufruct kind, meaning the right to use and enjoy the benefits and fruits of forest resources without depleting the overall resource system. This looks quite akin to managing an intangible resource commons.

107 Circular on Joint Forest Management (1990). http://awsassets.wwfindia.org/downloads/lecture_notes_session_9_1.pdf

108 WIPO, TK

109 Intellectual Property Rights: Need for a *Sui Generis* Regime for Non-codified Traditional Medicine in India, https://www.imedpub.com/articles/intellectual-property-rights-need-for-a-suigeneris-regime-for-noncodified-traditionalmedicine-in-india.pdf

110 https://www.wipo.int/edocs/lexdocs/laws/en/in/in048en.pdf

such traditional knowledge gets put back into the 'knowledge commons', and it cannot be patented anywhere. Commercial uses of traditional knowledge have to be appropriately negotiated with commons license rights holders. The government was to set up a Traditional Knowledge Authority which will register practices and practitioners of traditional knowledge. Participatory governance of traditional knowledge will be ensured through Traditional Knowledge Users Cooperatives.

The international Convention on Biological Diversity (CBD) governs genetic resources of flora and fauna associated with a nation and its communities. Closer to traditional knowledge in being collectively associated with specific communities, genetic resources are even more clearly not governable under IP frameworks. Occurring naturally, it is not possible to consider them in any way as intellectual products or 'creations of mind' – whether individual or collective. The Nagoya Protocol of the CBD employs *sui generis* formulations to institute national and community rights over their genetic resources. Such rights, on one hand, build over traditional knowledge frameworks, flora and fauna being closely associated with a nation/community's cultural practices. On the other hand, such rights connect to a nation's ownership over its natural resources, and its right to manage them for its people's best benefit.[111]

The Nagoya Protocol calls for 'prior informed consent' of, and 'benefit sharing' with, the nations and communities associated with the flora and fauna concerned. It also lays out the means for interested parties to legitimately access the resources involved. These concepts of 'consent', 'benefit sharing' and appropriate 'access to resources' closely mirror ideas regarding the community resource of data expressed in the report of the Indian Committee on Data Governance Framework, as discussed in next section.

Giving effect to the global Convention on Biological Diversity (CBD), India's Biological Diversity Act lays out rights for communities to benefit from their genetic resources, and participate in their governance, through Biodiversity Management Committees at local levels. Courts have recognized biological resources as "definitely the property of a nation where they are geographically located, (and) ...also the property, in a manner of speaking, of the indigenous and local communities who have conserved it through centuries".[112] Prior approval is required for foreign parties to access any biological resource occurring in India or knowledge associated therewith, whether for "the purpose of research or for commercial utilization or for bio-survey and bio-utilisation".

With data becoming an important new resource, governance frameworks for community resources are getting extended to data connected with the community resources that they cover. The hottest topic currently in the framework of Nagoya Protocol of the CDB is whether gene sequencing data of flora and fauna gets covered by it, especially in terms of the Protocol's 'prior consent' and 'benefit sharing' requirements. Countries of the South, including India, assert that as involving a use or 'utilisation'[113] of genetic resources, gene sequencing data and its derivatives are included. Northern countries, and the International Chambers of Commerce, claim otherwise, arguing that the CBD describes genetic resources as 'material' and not in the form of information. Nagoya Protocol says "...utilization of genetic resources as well as subsequent applications and commercialization shall be shared in a fair and equitable way with the Party providing such resources that is the country of origin of such resources...".[114] It is difficult to see gene sequencing data as not

111 Common Article 1 of the International Covenant on Civil and Political Rights and the Covenant on Economic, Social and Cultural Rights.

112 http://nbaindia.org/uploaded/pdf/Judgment%20in%20W.P.%20No.3437%20of%202016%20(Divya%20Pharmacy%20Case.pdf

113 A term included in the CBD's and NPs original language.

114 https://www.cbd.int/abs/text/articles/?sec=abs-05

being a ‘utilisation’ of genetic material.

Considerable discussion currently focusses on whether open sharing of gene sequencing data itself constitutes sufficient ‘benefit sharing’ or such data should be subject to additional conditions of prior consent, remuneration for the country of origin, etc. About open sharing too, there is a debate whether it is enough to put such data in public domain (the need for which most agree on), or open licensing is required with appropriate conditions of attribution of provenance, putting derivatives back in the common pool, etc.[115]

It appears discordant that when a community has commons rights to data about its flora and fauna similar community rights do not exist for data about its people (the CBD does not include human genetic material because of its natural preservation focus). The difference seems to be simply because flora and fauna, and other natural resources, are seen in a common heritage perspective, but social artefacts and resources get ‘primarily’ treated individualistically in a property framework.

The Human Genome Project very early adopted a radically open access approach, whereby new genomic data has to be put into public domain within 24 hours of its discovery.[116] However, issues are now arising about whether specific communities from which certain kinds of human genomic data is sourced require various protections, prior consent, rights of decision making, and ‘benefit sharing’.[117] This shift from an open access to community rights approach is significant.

If human genetic data can require community governance approaches, so may also larger health data. Many initiatives have been undertaken and papers written for treating health data as a public good.[118] Going beyond universally open public good to a community rights approach, WHO’s Pandemic Influenza Preparedness Framework recognises possible issues of sensitivity to open publication of data for genetic material originating country. It also mentions special benefit sharing, recognition and acknowledgement, and participation in decision-making, for the ‘originating country’.[119] When WHO’s Executive Board discussed its digital health strategy in January 2020, the advocacy group Public Health Movement made a statement seeking a community rights approach to health data as against treating it just as a public good.[120] Access to COVID-19 Tools Accelerator (ACT-A), is a G20 initiative for Global Collaboration to Accelerate the Development, Production and Equitable Access to New COVID-19 diagnostics, therapeutics and vaccines. Its Framework for the Governance of Personal Data discusses possible harms to communities and community privacy, engagement and involvement with communities in decision-making on the use of data, and sharing benefit arising from the use of data with the community concerned.[121]

Society’s health data is the first being recognised as requiring a community resource governance framework. Data about education, agriculture, transportation, urban development, and so on, too will soon be viewed through a similar lens. All such data involves important public interest aspects – although the nature and intensity of public interest may vary, as it also evolves and changes. ‘Community data’ in every sector can, in general, be recognised as a commons resource, best subject to community governance frameworks. Depending on contextual and evolving public interest requirements in different sectors, as needed, data

115 The gene sequencing survey paper
116 Havana principles
117 cite
118 cite
119 http://whqlibdoc.who.int/publications/2011/9789241503082_eng.pdf?ua=1
120 cite
121 https://www.finddx.org/wp-content/uploads/2021/01/ACT-A-Dx-data-governance-framework_15.01.2021.pdf

infrastructures based on common pooling and common access can be developed from such community data. The Indian Committee takes such a graduated approach, proposing that India begins with a health data infrastructure as a pilot project for community rights based data governance and management.

A community resource governance framework for data

A review of governance of some community resources in the last section showed that while different community resources require different treatments, there exist many ideas and principles that are common to their effective governance. Five such guiding principles can be distilled: (1) a community's right over resources that are collectively associated with it, ((2) prior informed consent of the community for use of its resources, (3) benefit sharing with the community, (4) transparency in recording community resources to prevent misuse and enable easy access of the legitimate kind, and (5) community's participation in governance of community resources, including through trusts etc. These established – though also yet evolving – principles at the global and national level for governing community resources provide a good basis for a governance framework for community data.

The Indian Committee recognizes the above five principles emerging from other community governance regimes, positing them as underpinning and justifying its recommendations. To quote the draft report:

- Principle 1 [above] enables the community to establish its rights over [its] NPD.
- Principle 3 enables a community to seek sharing of NPD about it collected by various parties (through data trustee creating an High Value Dataset). At the same time, such community rights may not interfere with a data collector's own access to and use of the concerned data resources.
- Principle 4 guides the concept of [data business and] High Value Datasets.
- Principle 5 underpins community's right to govern its NPD through appropriate data trustees. (Additions made are in square brackets)

The above omits general principle 2 of prior informed consent. Apparently, the Committee might have wanted to avoid giving the impression that collecting or processing Indian NPD will require any kind of prior permission, given the Committee's stated intention to keep the regulatory burden as light as possible. The principle of prior informed consent however connects to the Committee's recommendation of taking consent from individuals before anonymising their data. It is also implied in having to register as 'Data Business' before undertaking any large-scale NPD activity in India. The general principle 4 also provides the basis for the requirement of transparency from any Data Business regarding various data processes. General principle 5 on community's right to govern its resources may support the localization requirements for a new category of 'sensitive NPD'

Every resource system is unique, and so is the nature of the community associated with it. Having applied concepts from other community resource frameworks to governance of community data, it must also be discussed how the resource of data is different and unique. Data mostly works well only at scale, and in a multi-tiered fashion, also forming many, often dynamic, cross-connections with other data-sets, possibly in very different realms and of distant provenance. This has raised questions about what could be the relevant community for any given dataset, as compared to other community resource systems where association with a community, and latter's geographic locality, are relatively well-marked. Data collection normally bears a direct connection with digital service delivery. Consumption of digital services, especially in non-virtual sectors like transportation, health, education, agriculture etc., can to a fair degree be associated with certain communities. Such relevant communities can be demarcated with significant meaningfulness, including as appropriate in a geographic manner, even if admittedly with a lot of overlaps and boundary vagueness. For

example, residents of a city consuming ride hailing, food delivery, educational, or health related digital services. In some cases there could be more specific user groups, like Parents Teachers Associations for educational services. In default of such more specific groups, the entire city, or its different localities, and the respective representative structures may be invoked. In other cases, the act of data generation/contribution itself indicates a distinct community, like farmers in a region generating/contributing farming related data.

With a significant degree of factual materiality ascertained for a community that can be associated with a certain set of community data, rest of the gap gets filled by the concept of legal fiction whereby an assertion is accepted as true for legal purposes even if unproven (or known to be false). The important legal purpose here is to develop an appropriate, useful and workable framework of governance for general data about a society in a manner that fulfils some key societal objectives discussed throughout this paper. As long as the results of legal fiction hold ground, the legal fiction can be accepted. In this case the result is a very plausible and effective system of data sharing and the data infrastructures – described in this paper, and laid out by the Indian Committee – that is much needed for a robust and fair digital society. Even if not mathematically precise, there is enough basis for the proposed new regulator to make decisions on what can be described as a community's data, whether a community trustee is an appropriate one, if appropriate and enough community participation exists for governance of an HVD data infrastructure, and whether the claim of community harm is raised by a legitimate party, and if it is valid. As long as these requirements can be met in a relatively satisfactory manner, the proposed community governance framework, and its concepts of community and community data, hold ground.[122]

Adopting for community data the definition of natural resource governance quoted at the beginning of the last section, we may say;

> *Community data governance refers to the norms, institutions and processes that determine how power and responsibilities over community data are exercised, how decisions are taken, and how citizens – people and communities – and businesses participate in and benefit from the management of community data.*

A most important difference between society's data and other resource systems under community governance frameworks is that society's data is increasingly at the heart of world's economy. Other discussed community resource systems – even if more fundamental to human survival – are relatively at the periphery of the mainstream economic system and its contentions. Society's data is an end-to-end resource across the society and economy. It is fundamentally transforming social and economic systems in a manner whereby control of data comes to be at the top of value chains. Any system for economic governance of data will therefore be contested much more strongly than those for the other discussed resources. It will in fact be the fundamental political economy contest of the coming times.

It must be clarified that (1) not all data is to be subject to a community resource framework, and (2) even for data so subject extensive private claims and rights will exist for use, benefit, harm preventions, etc. A community-commons framing for the resource of data merely complements the current exclusive use of individualist property approaches, whose limitations were discussed in part 1 of this paper.[123]

---

122 As appropriate data governance frameworks take hold, which include community rights, data holders will organise their data more in ways that community rights can be assessed and implemented.

123 New research by MIT IDE co-lead, Alex "Sandy" Pentland, and his co-authors, claims that we must move "from an individualized, asset-based understanding of data control, to a collective system based on rights and accountability." The new system will have "legal standards upheld by a new class of representatives who act as fiduciaries for their members" much as credit unions do for their owner-members. http://ide.mit.edu/news-blog/blog/qa-sandy-pentland-data-cooperatives

This paper is unable to enter into a complete elaboration of how a community resource framework for data will provide private actors with various kinds of data-related privileges and rights. We may just touch upon one good basis for data rights of data collectors in a community data framework. Data collectors can have usufruct rights, to use community data and benefit from its fruits, without diminishing or destroying the resource of community data in terms of its value to the community (this will include privacy harm, economic exploitation, etc.). Such rights are well recognized in various community resource frameworks, for instance in India's Forest Rights Act.

As happened with land, industrial/financial, and intellectual resources, it will take decades for the required frameworks of economic governance of data to mature. This paper suggests what can be an appropriate point of departure and overall conceptual framework. The basis for such a community-commons starting point for economic governance of data arises from (1) big data's economic value being most in its aggregate or collective forms, and (2) 'data sharing' being the key plank of digital industrial strategies for nations to make digital progress in a globally just and fair manner. Putting the concept of 'data sharing' in a rights framework directly bespeaks a collective right to a data commons. Without providing a collective rights framework to it, adequate and appropriate 'data sharing' cannot happen, which means that the necessary data infrastructures cannot be developed, or they would not be effective. And without appropriate economy-wide data infrastructures, a vibrant and fair digital economy would not exist. This is the key takeaway from the draft recommendations of India's Committee on Data Governance Framework. It is hoped that the EU's proposed Data Act in 2021, which is expected to lay out data access rights, too will recognize these cardinal imperatives.

Endnote

The Indian Committee on Data Governance Framework has focused on the rights of communities of natural persons as the basis for developing sectoral data infrastructures needed to support a domestic digital and AI industry. It has not addressed the important issue of data rights of platform dependent small economic actors, something that some recent draft laws in the EU deal with very effectively and thoroughly. These draft laws provide (implicit) rights and enabling legal frameworks for platform dependent actors, which, if innovatively used, can transform the platform business model. The Indian Committee takes a community rights approach whereas the EU has mostly relied on enabling provisions, including collective ones like for data cooperatives, towards forming what it calls as sectoral 'data spaces' – similar to India's sectoral data infrastructures. But what remains unclear are the means of access to the the main hoards of society's data held by a few global corporations, that are required to feed European data spaces. Without access to these data hoards, no data space or data infrastructure can be reasonably complete and effective. The Indian framework ensures this through community rights based mandated data sharing. Another very significant benefit of the Indian community rights approach is that it is simultaneously able to address the issue of collective community harms from data abuse. This remains a blind spot in the European framework, it being centred on business-to-business relationships, in so far as it goes beyond personal data protection.

As seen from its projects like GAIA and Industrial Data Space, governance of sectoral data spaces in the EU is with large domestic data generators and users. Recent draft laws in the EU provide a very useful and detailed enabling framework for third party institutions that can underpin such cooperative governance, including providing checks against misuse of data intermediation. The governance of data infrastructures in India will be with representatives of the communities that are the collective subjects of community data, a participatory tenet that arises from India's community rights approach. However, adequate guidelines are not

provided on how such governance will actually be organized, and how various checks and balances ensured. Without effectve institutional frameworks, community trustees may employ their power inappropriately for undue selective benefits.[124] It is equally feared that governments can come to exercise too much unchecked power in society's data management and governance. The state's basic role as the backstop for any economic rights framework, and for ensuring appropriate economic infrastructure, meanwhile remains important. It is therefore a matter of institutional depth, balance and maturity.

The EU and Indian approaches have a lot to learn from one another. These together can hopefully provide a comprehensive framework for economic governance of data.

---

124 The Indian Committee's report recognises the problem of community trustees becoming too powerful, but leaves it to the proposed new regulator to consider such issues.