IT for Change

DGN Policy Brief 02
January 2020

Data Governance Network

Data Sharing Requires a Data Commons Framework Law

Parminder Jeet Singh and Anita Gurumurthy

In a long paper on 'Data and Digital Intelligence Commons',[i] we examined the current non-governance of aggregate non-personal data, and presented a new approach that involves treating a large part of such data as a 'common pool resource'. This policy brief summaries that paper's arguments and takes them forward to propose a concrete policy option; of developing a framework law for protecting and sharing the common resource of aggregate non-personal data.

1. Why would corporations share their data?

Digitalisation was supposed to be a leveller of access, opportunity and resources. UNCTAD's 2019 Digital Economy Report however describes its current tendencies to concentrate economic power. US and China together account for 90 per cent of the market capitalization value of the world's 70 largest digital platforms.[ii] Other countries are very worried about such global digital and artificial intelligence (AI) concentration, which is shaping into a duopolistic race between these two digital superpowers[iii]. AI strategies of UK, France, India, and other countries, all recommend developing a strong domestic AI and data-based industry as the key imperative for retaining global and domestic economic strength and control. Central to these strategies is the need for ensuring wide access to society's data for the domestic industry. The main action item that these AI strategies unanimously propose is taking steps towards increased sharing and availability of data. But the question of how this will happen is not pursued to its logical end. These strategies mostly just hope that wider sharing of data will somehow begin to take place – given some enabling institutional conditions, like security, protection from liabilities, reciprocal access to data, etc.

Such an expectation of voluntary wide sharing of data needs to be tested against some well-known facts. By far the biggest collectors of data today are a few global digital corporations. Exclusive access to, and control over, the data that they collect is increasingly their single most important business advantage. It is not clear why these corporations would share their data, whatever enabling conditions governments provide. They are certainly not waiting to be provided conditions of digital security, protection, reciprocity, etc., to enable sharing. Their own prowess in these areas is much greater than anything a public institution can come up with. They already do share data quite well where needed, and among those valuable to do so, for instance some degrees of sharing of training data for autonomous cars.[iv]

Many of these corporations also voluntarily share some data for public good, two examples being Facebook's 'Data for Good' and Uber's 'Movement' initiatives. Such largesse of course comes only when these corporations' interests are somehow furthered, or at least not compromised. It should, for instance, be clear that they will never be too eager to share data for the purpose of increasing domestic competition in the sectors that these global corporations dominate, or may make future forays into[v]. Improved domestic competition however is the key purpose for seeking data sharing by all the mentioned AI strategies. This paradox remains unaddressed in these policy documents. It cannot be solved without sorting out some core political economy issues around data. Doing so mainly involves allocation of economic rights to various kinds of data for different actors in data value chains.

2. Categories of data

Data is the most valuable resource in a digital economy. To ensure sharing of data it would therefore require explicitly laying out the economic rights and obligations that different actors have *vis-a-vis* data. In order to allocate such rights, data first needs to be classified into some basic relevant categories. Two of these categories are relatively clear, along with the involved economic rights – personal data and public sector data, or in short public data. Personal data is personally identifiable data, about which the individual data subject is supposed to hold most rights. Considerable work in ongoing with regard to such data. Public data is data produced by government agencies, and governments clearly hold the economic rights to it. They are normally exhorted to openly share such data with businesses etc. for enhancing a digital economy. This has its roots in the pre-digital 'open data' movement. Data sharing in the digital age is much more complex, requiring many new kinds of safeguards and enablements. Governments across the world are working on these imperatives. India is especially undertaking considerable activity in this area, developing many kinds of public data infrastructures.[vi]

Data rights of organisations and businesses are much less clear, although in practice digital corporations claim more or less proprietary rights over the data they collect, with its exclusive commercial exploitation. Data that gets fully or substantially created within an organisation may be legitimately its to own and use exclusively. This category of data can be called private data. But what about data that is substantially sourced from outside the collecting entity? What are the respective economic rights to such outside data of data collectors and data contributors, or data subjects? This contestation is developing considerably well for personal data, and this paper will bypass it. But the situation remains absolutely unclear with regard to non-private aggregate non-personal data. This can be anonymised data about groups of people, or arise from outside objects, or even natural phenomena. As the digital economy shifts from being targeted-advertisements focussed to becoming AI-centred, the category of aggregate non-personal data becomes ever more important.

Understanding the nature of (non-private) aggregate non-personal data, and establishing various economic rights to it, is key to enabling wide sharing of data. This is required for a sustainably productive and just digital economy. Since such data comes from groups of people or communities, and/or objects or phenomena associated with them, we call it community data. This categorisation is useful to see such data as distinct from personal, public and organisational or private data.

3. Default data regimes

Ownership of data is not only established by law; it can also be formed by practice[vii]. Digital corporations often provide free or subsidized services in order to collect enormous amounts of data, and then treat it as their property. They are now making their AI engines open source and freely available, a technology that is very expensive to develop currently. This results in extensive 'data enclosures' as more and more actors are enticed to bring their data to these engines to obtain AI insights for themselves. In the bargain, these digital corporations begin getting almost unlimited access to society's data of every kind. As practices concretise default ownership of society's data to be with its collectors – especially of aggregate non-personal data, this becomes the *de facto* property law for such data. As a digital economy shapes to its fullness, enveloping all sectors, and getting enmeshed with all social and economic areas, such default data property law will *per force* have to be accepted by everyone. It will be too late to change anything – with too much invested in and staked on entrenched data practices and models. As they say, in the digital area, code is law and architecture is policy.[viii]

The world seems to be walking blindfolded into this trap, of a digital economy and society largely run by a few global digital corporations that own most of the key digital resource of society's data. By owning society's data, they own most of society's intelligence that such data generates. A digital economy consists of sector-wide intelligent systems that begin by connecting, then coordinating, and finally controlling all other actors in the sector[ix]. The hierarchically highest position of the owner of the intelligence in a system in relation to its other actors is self-evident. Further, intelligence has this unique centralising quality, whereby one system controller – or a monopoly – is exponentially more intelligent than multiple ones. These basic features of data based intelligence are behind how the digital economy is currently developing

– in the absence of appropriate countervailing policy measures. It is concentrating digital power and wealth, which is getting organised around two competing global digital poles of the US and China.

Policy-makers have good reasons to be extremely concerned with these developments. It is not just about promoting a strong domestic digital industry, but also justified fears of long-term loss of overall economic and political autonomy and sovereignty. Urgent policy steps are required to decentralise digital and data power.

Various national AI strategies rightly identify wider data sharing as being key to a sustainable and just digital economy. This requires effective institutional means that are based on allocation of legal economic rights to, and obligations for, different data actors, and includes mandatory sharing, as appropriate. A hard-nosed political economy and institutional approach to data rights and data sharing is necessary. The key institution of resource pooling and sharing is a 'commons', of which many kinds exist. This is a useful area to explore, in a grounded institutional manner.

4. Political economy of data

Regarding information and knowledge commons, it was the large corporations that wanted strong legal economic rights around knowledge, while many public interest actors sought looser regimes. The case seems to be reversed for society's data. Digital corporations want to avoid any discussion on legal economic rights to data, because they possess the technical means to collect and exclusively appropriate society's data, and exclude others. Public interest actors may, therefore, have to become the main advocates for allocating economic rights around data, in order to ensure that society's data can be pulled from behind the technology walls and shared widely, arriving at an arrangement that is fair to all.

Digital corporations like society's data to be treated as an open access resource, freely available to everyone. This notion underpins their current data practices. They reason that others can as well collect the same data that they do. This argument is important to rebut. Collection of society's data is fully linked to applications and platforms that increasingly perform natural monopoly-like infrastructural functions in a society; like, Google's search engine and maps, Facebook's social networking platform, and Amazon's e-commerce platform. Unless one first owns such digital infrastructures one cannot collect the required data. And without such data one cannot start a digital business to get to be a sector-wide data collection platform. In order to be shared widely, society's data may, therefore, need to be obtained from the incumbents employing legal means, as appropriate, and on suitable terms.

Data cannot simply be considered an open access resource. It is embedded into corresponding digital systems, and is a rivalrous system resource.[x] Operating closely in conjunction with rival and excludable resources, that include physical resources and system control and management, data is neither adequately non-rival nor non-excludable. Data, and intelligence derived from it, are inalieanbly linked to the respective group or community about which the data is. Such a deep linkage is recognised for personal data; it being considered the extension of one's person-hood. Community data is similarly an extension of a community, which should have inherent primary rights over such data as its collective data subject. This leads to a differentiation between the rights over community data of actors within the community, and representing it, and those outside.

Owning and controlling a group's or community's data is to control intelligence about it – intelligence of the most intimate, detailed and granular kind. There is a strong moral basis for a group or community to own and control systemic intelligence about and over itself. This further establishes the case for society's or a community's data being a 'common pool resource', like pastures and forests, rather than an 'open access resource' like ordinary generalisable knowledge. A community's data requires various protections and circumscribing that generalisable knowledge may not. The manner in which data is actually employed in a digital economy, its value is subject to over-exploitation, depletion, congestion and pollution, characteristics typical of 'common pool resources'.

5. Common property regimes for community data

Elinor Ostrom has been the foremost theoretician on common pool resources.[xi] The Institutional Analysis and Development (IAD) framework popularised by her gets applied to many areas; from natural resources to knowledge commons, infrastructure, property rights, donor-recipient relationships, public housing, non-profits, social dilemmas, peace and national building, and foreign aid.[xii] Society's data should be a natural inclusion in this list. The mentioned long paper on 'Data and Digital Intelligence Commons' subjected a typical digital economy system to an IAD framework. It found that the dominant model of exclusive data appropriation performs poorly on all the six criteria proposed by Ostrom for evaluating a resource governance system. These are economic efficiency, fiscal equivalence, redistributional equity, accountability, conformance to values of local actors, and sustainability.[xiii]

Common pool resources are governed under common property regimes. Ostrom proposed eight design principles for common property regimes: define clear group boundaries; match rules governing use of common goods to local needs and conditions; ensure that those affected by the rules can participate in modifying the rules; make sure the rule-making rights of community members are respected by outside authorities; develop a system, carried out by community members, for monitoring members' behaviour; use graduated sanctions for rule violators; provide accessible, low-cost means for dispute resolution; build responsibility for governing the common resource in nested tiers from the lowest level up to the entire interconnected system. [xiv]

It will be interesting to explore how these principles can be employed to develop sharing regimes for society's data, which – and a conceptual and policy paralysis over which – we earlier identified as the key digital economy issue today.

Application of a common property regime framework to develop data sharing arrangements will certainly be highly contextual to the unique nature of data, and data derived intelligence, as common resources. The real shift here is to move the default ownership of a society's data from the data collecting entity to the community concerned. Data is not like physical things, nor it is like created, generalisable knowledge. The term ownership is employed here somewhat loosely, as representing primary economics rights to specific kinds and uses of data. The actual spectrum of data-related economic rights, privileges, obligations and exclusions will be a complex system, evolving over time upon this foundational principle of a community's ownership over its data. It will be underpinned by policy, legal and business developments extending to many years, if not decades.

7. Need for a data governance framework

Such a new data governance approach will not necessarily disrupt everything that is currently known and practised in the digital economy. It would in fact bring considerable legal certainty, which is to everyone's benefit. A lot of data will still remain private data of the concerned corporations. The latter will also be given appropriate incentives to continue to collect and process important data of the society. These, in some cases, could be in terms of exclusivity for specific uses of data for a limited period, as appropriate. It has also to be figured out how the relative data rights of the data collecting business and the society or community shift with increasingly higher levels of processing of data; and at which point derived data becomes mostly or considerably private to the business concerned. A practical, least disruptive, path of transition to new models can be worked out.

For a start, what is needed is a framework law that institutes society's data as a common pool resource, laying the basic outlines of a common property regime for community data. The term 'framework law' refers to a legislative technique used to address cross-sectoral issues; framework legislation lays down general principles and obligations, and leaves it to implementing legislation and competent authorities to determine specific measures to be taken to realize them.[xv] It provides a larger enabling legal and institutional framework. A framework law on community data can enable and support ongoing development of specific rules and institutions in different areas regarding various kinds of data, flexibly addressing emergent digital realities.

Such framework laws have been developed or proposed for some other common pool resources, like water; for instance, the European Water Framework Directive (2000)[xvi], and a draft Water Framework Law developed by the Planning Commission of India's Working Group on Water Governance for the Twelfth Plan.[xvii]

Apart from proclaiming the default community ownership of relevant data, such a framework of community data governance will provide the principles for allocation of different economic rights to different data actors, and for enabling appropriate regulation of a data economy. This is especially necessary for developing the required institutions for data sharing, that are practical and can actually deliver; going beyond the simplistic notions of data as an open access resource and unrealistic expectations of voluntary sharing of the data that is most required to be shared.

8. A framework law for community data

We attempt here to make the proposal for a framework law relatively concrete, with the intent to provide a clear picture to policy-makers of what a common property regime for community data would look like. A suggested title for the proposed framework law is 'National and Community Data (Protection and Sharing) Act'. Some indicative points for its preamble part are as follows:

- Aggregate data of and about a community is to be considered a 'common pool resource' on the lines of a community's natural resources, like its flora and genetic resources. These natural resources arise from the nature, close to and inter-twined with a community. Data arises directly from the society, its relationships and structures, and can be considered a common social resource.

- Data about groups and communities, and from other commons, is a key resource of a digital economy. It is currently freely appropriated by anyone in a position to collect it and then employed as private data and property. No legal regime exists around collection and use of such data, which allows its free and unregulated appropriation. This has strong protection or security as well as economic implications for any community, apart from a range of other social, political and cultural implications. This Act provides a legal framework for regulation, including enablement, of collection and use of such commons data emerging from a society or community, and belonging to it.

- The purpose of this Act is twofold: (1) To protect and secure all such aggregate non-personal data, and the communities, people and 'things' related to it: (2) To ensure, duly protected, wide sharing and availability of such data domestically, and otherwise, as required and found appropriate, for public interest purposes, including for India's digital development, and rapid growth and sustenance of India's digital industry.

- This Act is an enabling framework for groups and communities, and their respective representatives and trustees, to regulate the use of non-personal community data. Such enablement includes ensuring sharing and availability of data. The Act seeks to provide communities and authorities a suite of regulatory and enabling tools that are necessary for adequate and appropriate governance of a data-based society and economy. How such tools are actually applied in the best public interest, and the interest of the community concerned, will be contextual, and subject to relevant laws, regulations and rules as laid down from time to time.

- This Act does not cover personal data – defined as personally-identifiable data, but includes its anonymised aggregate forms. The Act also does not include or cover data that is private to an entity as having been entirely or substantially produced by it, and/or arising from within it.

Below are the kind of provisions that will go into the substantive part of the proposed framework law.

- Aggregate non-personal data that arises from people, groups and communities, or from objects belonging to them, or from natural things or phenomena generally associated with them, belongs to the respective group or community, in terms of its economic, public, social, political and cultural value.

- Such a group or community will be considered the collective owner of such community data. It will have the right to determine the use, and the manner of use, of such data, and in general be primarily entitled to the value generated from it. This is apart from various economic rights that the group or community may decide to allocate or transfer to different entities, for incentivising data collection and/or processing, or for other purposes.

- The terms 'belong' and 'ownership' are employed in the sense of a party having primary economic and other rights to the data concerned, and being the principal one in matters of decisions about such data, with an understanding that data can have various kinds of rights and uses simultaneously associated with it, which can accrue to different parties.

- The group or community concerned will exercise its collective ownership, or primary rights, over such community data through the agency of a suitable trustee. Such a trustee would be representative of that group or community. For various geographic communities, the closest governmental unit, as practical and appropriate, may act as the trustee to regulate – through a legitimately constituted body – the collection and use of such community data, including its sharing.

- Data generated and collected within a nation or country collectively belongs to that nation or country, whereby it has overall controlling rights over such data, without unduly affecting private data rights. All such national data is subject to corresponding national sovereignty – which can be termed as data sovereignty, as are other things, people, etc., within, and arising from, the territory of that country.

- Principles and rules will be developed to assign the appropriate representative and trustee for a group or community to operate its community data rights.

- Entities that collect community data will be deemed custodians of such data. They will be subject to legitimately issued rules, regulations and directions of the corresponding group or community representative or trustee for collection, use and sharing of such data.

- Such rules and regulations will be laid out separately for different kinds of data, and for different areas of data collection, use and sharing. These can range from simply requiring intimation of collection of data to prohibitions on collecting certain kinds of data, through various other regulatory, enabling and sharing possibilities.

- The provisions of the Act become valid and operative as, when and where the corresponding laws, regulations and rules are developed under it, pertaining to specific categories of data, sectors, data-uses and actors. In default, existing data practices will be deemed not to have violated the provisions of this Act.

- An independent National Data Trust, and its sub-units, will be set up under this Act. It will have quasi-judicial authority to decide and adjudicate on application and appropriateness of various data regulations and rules under this Act. The Trust will provide necessary principles and guidelines, from time to time, that will help clarify the application of the Act in different contexts, including new emerging situations.

- The Trust will lay out what constitutes shareable community data for common use in a given context, and what incentives, if any, may be provided to collectors of data in case they add any explicit and substantial value to such data. The required manner of sharing data will be laid out. The Trust may determine which community data – with or without some value added to it – must be shared for free, with which data users, and under what conditions; which value-added community data may be required to be shared on FRAND (fair, reasonable and non-discriminatory) terms; and, what kind of data may be subject to normal fair and regulated data markets. Special and differential provisions can be made, as appropriate, for businesses below a capital, or data collection, threshold.

- The National Data Trust will have an implementation unit that will develop and support the required data infrastructures, and other mechanisms, for safely sharing various kind of community data.

Notes

i. Singh, Parminder Jeet (forthcoming). Data and Digital Intelligence Commons (Making a Case for their Community Ownership). IT for Change, as part of the Data Governance Network.

ii. UNCTAD. Digital Economy Report 2019 – Overview. Retrieved from https://unctad.org/en/PublicationsLibrary/der2019_overview_en.pdf

iii. In her first speech as the Managing Director of the International Monetary Fund, Kristalina Georgieva spoke of a ‘digital Berlin Wall’ that forces countries to choose between technology systems. See https://www.imf.org/en/News/Articles/2019/10/03/sp100819-AMs2019-Curtain-Raiser

iv. The Verge (2019), Waymo Gives Away Free Self-Driving Training Data -- But With Restrictions. Retrieved from https://www.forbes.com/sites/bradtempleton/2019/08/22/waymo-gives-away-free-self-driving-training-data-but-with-restrictions/#dc9fe5923bdc

v. UK’s AI strategy of 2018 proposed data trusts as the major framework for data sharing. The 2019 report of the Digital Competition Expert Panel set up by the UK government, ‘Unlocking Digital Competition’, however observes that “the scope for data trusts to stimulate competition within existing markets would currently appear to be limited...”. The report is at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf

vi. Singh, Parminder Jeet (2017). Digital Industrialisation in Developing Countries. Commonwealth Secretariat, Part 3. Available here: https://itforchange.net/sites/default/files/1468/Digital-industrialisation-May-2018.pdf

vii. Purtova, N. (2015). The illusion of personal data as no one’s property. Law, Innovation and Technology, 7(1), 83-111. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2346693

viii. Lawrence Lessig (2000). Code is Law and Law is Code. Retrieved from https://cyber.harvard.edu/works/lessig/pcforum.pdf

ix. Singh, Parminder Jeet (forthcoming). Data and Digital Intelligence Commons (Making a Case for their Community Ownership). IT for Change, as part of the Data Governance Network.

x. Purtova, Nadezhda, Illusion of Personal Data as No One’s Property (October 29, 2013). Law, Innovation, and Technology, Volume 7, Issue 1, 2015. Available at SSRN: https://ssrn.com/abstract=2346693

xi. Ostrom, Elinor. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge University Press.

xii. See, for example, Ostrom, E. (2002). Aid, incentives, and sustainability: an institutional analysis of development cooperation. Main report. Sida. Retrieved from https://www.oecd.org/derec/sweden/37356956.pdf

xiii. Ostrom, E. (2011), ibid.

xiv. Walljasper, J. (2011). Elinor Ostrom’s 8 Principles for Managing A Commons. On The Commons. Retrieved from https://www.onthecommons.org/magazine/elinor-ostroms-8-principles-

managing-commmons

xv. Food and Agriculture Organisation (2009). Guide on Legislating for the Right to Food. Retrieved from http://www.fao.org/fileadmin/templates/righttofood/documents/RTF_publications/EN/1_toolbox_Guide_on_Legislating.pdf

xvi. European Union (2000). The EU Water Framework Directive. Retrieved from https://ec.europa.eu/environment/water/water-framework/index_en.html

xvii. Planning Commission of India (undated). The Draft National Water Framework Act: An Explanatory Note. Working Group on Water Governance for the 12th Plan. Retrieved from http://www.planningcommission.nic.in/aboutus/committee/wrkgrp12/wr/wg_wtr_frame.pdf

Data Governance Network

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

About Us

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

Acknowledgements

The authors will like to recognise the role of the entire IT for Change in this work. At IT for Change, knowledge production is a collective enterprise. But for it, this work would have been impossible. The extensive research assistance provided by Jai Vipra is especially appreciated. The author expresses gratitude to the members of the Data Governance Network, some of those who attended its research meetings, and an anonymous peer reviewer, for very valuable comments.

IDFC Institute
3rd Floor, Makhija Chambers, 196 Turner Road,
Bandra(W), Mumbai 400050

/idfcinstitute @idfcinstitute /IDFCInstitute