Inputs to the Report on the Non-Personal Data Governance Framework

September 13, 2020

Authored by **Aayush Rathi, Aman Nair, Ambika Tandon, Pallavi Bedi, Sapni Krishna,** and **Shweta Mohandas** (in alphabetical order)

Reviewed by **Sumandro Chattapadhyay**

The Centre for Internet and Society, India

Preliminary

This submission presents a response by researchers at the Centre for Internet and Society, India (CIS) to the draft Report on Non-Personal Data Governance Framework prepared by the Committee of Experts under the Chairmanship of Shri Kris Gopalakrishnan (hereafter "Report").[1]

CIS, established in Bengaluru in 2008 as a non-profit organisation, undertakes interdisciplinary research on internet and digital technologies from public policy and academic perspectives. Through its diverse initiatives, CIS explores, intervenes in, and advances contemporary discourse and regulatory practices around internet, technology, and society in India, and elsewhere.

CIS is grateful for the opportunity to submit its inputs. The below inputs are organised thematically, with references to specific parts of the Report highlighted.

Inputs

Clause 3.7 (v): The role of the Indian government in the operation of data markets

While highlighting the potential for India to be one of the top consumer and data markets of the world, it also sheds light on the concern about the possibility of data monopolies. The clause envisions the role of the Indian government as a regulator and a catalyst for domestic data markets.

In doing so, the clause does not acknowledge that **the proactive and dominant roles of the Indian government in generation and reuse of data**, based on the existing data collection practices, as well as the provisions that have been given, as under the compulsory sharing provisions in the Report, and would continue to be given by the Personal Data Protection Bill. In reality, the Indian government's role is not just of a catalyst but also of a key player, potentially with monopolistic market power, in the domestic data market, especially due to the ongoing data marketplace initiatives as detailed in published policy and vision documents.[2]

---

[1] Committee of Experts on Non-Personal Data Governance Framework. (2020). Report by the Committee of Experts on Non-Personal Data Governance Framework. Ministry of Electronics and Information Technology. https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf

[2] *See* Ministry of Health and Family Welfare. (2020). National Digital Health Blueprint. Government of India. https://main.mohfw.gov.in/sites/default/files/Final%20NDHB%20report_0.pdf; Tandon, A. (2019). Big Data and Reproductive Health in India: A Case Study of the Mother and Child Tracking System. https://cis-india.org/raw/big-data-reproductive-health-india-mcts

Clause 3.8 (iv): Introducing collective privacy

The introduction of *collective privacy* has initiated an overdue discussion at the policy level to arrive at privacy formulations that account for limitations in the contemporary dominant social, legal and ethical paradigms of privacy premised on individual interests and personal harm. The notion of *collective privacy* has garnered contemporary attention with the rise of data processing technologies and business models that thrive on the collection and processing of aggregate information.

While the Report acknowledges that *collective privacy* is an evolving concept, it doesn't attempt to define either *collective* or what *privacy* could entail in the context of a *collective*. The postulation of *collective privacy* as a legally binding right is bereft with challenges in both domestic and international legal frameworks.[3]

Central to these challenges is the representation of the group of the entity. While the Report illustrates harms that may be incurred by certain *collectives* that *collective privacy* could protect against, these illustrated *collectives* are already recognised in law as rights-holding groups (society members, for example), and/or share pre-determined attributes (sexual orientation, for example).

The Report does not acknowledge that the very technological processes that may have rendered the articulation of *collective privacy* necessary, also are intended to create ad-hoc and newer sets of individuals or groups with shared attributes.[4] In doing so, the Report furthers an ontology of groups having intuitive, predetermined attributes that exist naturally, or in law, whereas the intervention of data collection and processing technologies can determine shared group attributes afresh. Moreover, the Report also ignores that predetermined attributes are static, and in doing so, ignores a vast existing literature speaking to fluidity of identities and the intersectionality of identities that individuals in groups occupy.[5] We fully appreciate the challenges these pose in the determination of the legal contours of *collective privacy*. Much of the Report's recommendations are premised on the idea of a predetermined *collective*, rendering more granular exploration of these ideas urgent.

Further, the Report also puts forth a **limited conception of *privacy* as a safeguard against data-related harms that may be caused to *collectives***. In doing so, it dilutes the conceptualisation of *individual privacy* as articulated in *Justice K. S. Puttaswamy (Retd.) and Anr. vs Union Of India And Ors*. Notwithstanding this dilution, the illustrations also only indicate harms that may be caused by private actors. Any further recommendations should envision the harms that may also be caused by public data-driven processes, such as those incubated within the state machinery.

---

[3] Taylor, L., Floridi, L., van der Sloot, B. eds. (2017) Group Privacy: new challenges of data technologies. Dordrecht: Springer.
[4] Mittelstadt, B. (2017). From Individual to Group Privacy in Big Data Analytics. Philos. Technol. 30, 475–494.
[5] *See* Taylor, L., Floridi, L., van der Sloot, B. eds. (2017) Group Privacy: new challenges of data technologies. Dordrecht: Springer; Tisne, M. (n.d). The Data Delusion: Protecting Individual Data Isn't Enough When The Harm is Collective. Stanford Cyber Policy Centre. https://cyber.fsi.stanford.edu/publication/data-delusion

Clause 4.1 (iii) and Recommendation 1: Defining Non-Personal Data

The Report proposes the definition of non-personal data to include (i) data that was never related to an identified or identifiable natural person, and (ii) aggregated, anonymised personal data such that individual events are "no longer identifiable". In doing so, they have attempted to extend protections to categories of data that fall outside the ambit of the Personal Data Protection Bill, 2019 (hereafter "PDP Bill"). The Report is cognizant of the fallible nature of anonymization techniques but fails to indicate how these may be addressed.

The test of anonymization in regarding data as non-personal data requires further clarification. Anonymization, in and of itself, is an ambiguous standard. Scholarship has indicated that anonymised data may never be completely anonymous.[6] Despite this, the PDP Bill proposes a high threshold of zero-risk of anonymization in relation to personal data, to mean "such irreversible process of transforming or converting personal data to a form in which a data principal *cannot be identified*". From a plain reading, it appears that the Report proposes a lower threshold of the anonymization requirements governing non-personal data. It is unclear how non-personal data would then be different from inferred data as described within the definition of personal data under the PDP Bill. **This adds regulatory uncertainty making it imperative for the Committee to articulate bright-line, risk-based principles and rules for the test of anonymization**. Such rules should also indicate the factors that ought to be taken into account to determine whether anonymization has occurred and the timescale of reference for anonymization outcomes.[7]

The recommendation also states that the data principal should "also provide consent for anonymisation and usage of this anonymized data while providing consent for collection and usage of his/her personal data". However the framing of **this recommendation fails to mention the responsibility of the data fiduciary to provide notice to the data principal about the usage of the anonymized data while seeking the data principal's consent for anonymization.** The notice provided to the data principal should provide clear indication that consent of the data principal is based on their knowledge of the use of the anonymized data.

Clause 4.8 (i), (ii): Function of data custodians

The Report does not make it clear who may perform the role of *data custodians*. The use of *data fiduciary* indicates the potential import of the definition of 'data fiduciary' as specified under Clause 3.13 of the PDP Bill. However, this needs to be further clarified.

---

[6] Rocher, L., Hendrickx, J.M. & de Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10, 3069 . https://doi.org/10.1038/s41467-019-10933-3

[7] Finck, M. & Pallas, F. (2020). They who must not be identified—distinguishing personal from non-personal data under the GDPR. International Data Privacy Law, 10 (1), 11–36. https://doi.org/10.1093/idpl/ipz026

Clause 4.8 (iii): Data custodians' "duty of care"

As is outlined in the following section on data trustees, it can be difficult for a singular entity to maintain a *duty of care* and undertake actions with the *best interest* of a community when that community consists of sub-communities that may be marginalised.

Further, 'duty of care', 'best interest', and 'absence of harm' are not sufficient standards for data processing by data custodians. **Recommendations to the effect of obligating data custodians to uphold the rights of data principals, including economic and fundamental rights need to be incorporated in the framework.**

Clause 4.9: Data trustees

The committee's suggestion that the "most appropriate representative body" should be the data trustee—that often being either the corresponding government entity or community body— is reasonable at face value. However, in the absence of any clear principles defining what constitutes "most appropriate" there are a number of potential issues that can appear:

Lack of means for selecting a data trustee: The report makes note of the fact that both private and public entities can be selected to be data trustees but offers no principles on how these data trustees can be selected, i.e. whether they are to be directly selected by the members of a community, and if so how. Any selection criteria or process prescribed has to keep in mind the following point regarding the potential lack of representation for marginalised communities that could arise from a direct selection of a data trustee by a group of people.

Issues of having a single data trustee for large scale communities and when dealing with marginalised communities: The report assumes that in instances wherein a community is spread across a geographic region, or consists of multiple sub-communities, then the data trustee will be the closest shared government authority (for example, the Ministry of Health and Family Welfare, Government of India being the data trustee for data regarding diabetes among Indian citizens).

This idea of a singular data trustee assumes that the 'best interests' of a community are uniform across that community. This can prove problematic especially when dealing with data obtained from marginalised communities that forms a part of a wider dataset. It is entirely possible to imagine that a smaller disenfranchised community may have interests that are not aligned with the general majority. In such a situation the Report is unclear as to whether the data trustee would have to ensure that the best interests of all groups are maintained, or would they be responsible for ensuring the best interests of the largest number of people within that community.

There are power differentials between citizens, government agencies, and other entities described by the Report. This places citizens at risk of abuse of power by government entities in their role as trustees, who are effectively being empowered through this policy framework as opposed to a representative mechanism. **It is recommended that data trustees be appointed by relevant communities through clear and representative mechanisms. Additionally, any individual should be able to file complaints regarding the discharge of**

community trust by data trustees. This is necessary as any subsequent rights vested in the community can only be exercised through the data trustee, and become unenforceable in the lack of an appropriate data trustee.

Any legislation that arises on the basis of this report will therefore have to not only provide a means for selecting the data trustee, but also safeguards for ensuring that data collected from marginalised communities are used keeping in mind their specific best interests—with these best interests being informed through consultation with that community.

Clause 4.10 (iii): Data trusts

Section 4.10 (iii) notes that data custodians may voluntarily share data in these data trusts. However it is unclear if such sharing must be done with the express consent of the relevant data trustee.

Clause 4.10 (iv): Mandatory sharing and competition

The fundamental premise of a mandatory data sharing regime seems increasingly distant from its practical impacts. The EU which earlier championed the cause now seems reluctant to further it on the face of studies which skews towards counteractive impacts of such steps. **Such steps could apply to huge volumes of first-party data companies collect on their own assets, products and services, even though such data are among the least likely to create barriers to entry or contribute to abuses of dominant positions.**[8] This is hence likely to bring in more chilling effect on innovation and investment than a pro-competition environment. The velocity of big data also adds to the futility of such data sharing mandates.[9] It is recommended that a sectoral analysis of this mandate be undertaken instead of an overarching stipulation.

The Report suggests extensive data sharing without addressing the extent of obligation on the private players to submit to these requests and process them. The availability of meta-data about the data collected may be made easily accessible under mandates of transparency. However, the access to the detailed underlying data will be difficult in most cases due to the current structure of entities functioning in cyberspace, evidenced by the lack of compliance to such mandates by Courts of Law in the EU. Such a system can easily eliminate the comparative advantage of smaller players, helping larger players with more money at their disposal enabling their growth and throttling the smaller players. It could have serious implications on data quality and integrity through the sharing of erroneous data. Access to superior quality digital services in India may also have to be compromised. If

[8] European Commission (2020). Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: A European strategy for data. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX:52020DC0066

[9] Modrall, Jay. (2019). Antitrust risks and Big Data. Norton Rose Fullbright. https://www.nortonrosefulbright.com/en-in/knowledge/publications/64c13505/antitrust-risks-and-big-data

this regime is furthered without amends to address these concerns, it might end up counter productive.

Clause 5.1 (iv): Grievance redressal against state's role

This clause acknowledges the vast potential for government authorities and other bodies to abuse their power as data trustee. In addition, it should describe the **setting up of impartial and accessible mechanisms for citizens to complain against such abuse of power and appropriate penalties, including the removal of the data trustee.**

Chapter 7, Recommendation 5: Purpose of data-sharing

Recommendation 5 leaves scope for "national security" as a sovereign purpose for data sharing. This continues to be in line with the trend of having an overarching national security clause, as in the Personal Data Protection Bill, 2019. **There could be provisions made to enable access to data for sovereign purposes without such broad definition, replacing it based on constitutional terms which will limit it to the confines laid down in the Constitution.** This will effectively curb any misuse of the provision and strongly embed the proposed regulation of non-personal data on constitutional ethos. This can also prevent future conflicts with the fundamental rights.

Platform companies have leveraged their position in society to take on an ever-greater number of quasi-public functions, exercising new forms of unaccountable, transnational authority. It is not difficult to imagine that this trend can continue to non-platform companies, or even taken forward by these very entities which also have access to a large chunk of non-personal data. A strict division between sovereign purposes and core public interest purposes seems difficult. However, it is imperative to have a clearer definition of core public interest purposes and sovereign purposes. The broad based definition may facilitate reduced accountability. Separating government actions from sovereign purposes could bring forth the power imbalance between the State and its people, while in the case of the non-governmental entities, it will facilitate encroachment of government functions by private players. Both these cases may not consider the best interest of the data generators, or the people at large.

Clause 7.1 (i): Data needs of law enforcement

Clause 7.1 (i) allows for acquisition of data governed by this framework for crime mapping, devising anticipation and preventive measures, and for investigations and law enforcement. While this may be necessary to be granted to law enforcement in certain cases, this should happen only with an express permission of a court of law. Blanket executive access allows higher possibility of misuse by the people involved in law enforcement.

Clause 7.2 (iv): Use of health data as a pilot

The clause suggests the use of health sector data as a pilot use-case. This is highly undesirable due to the inherent nature of high sensitivity of the larger part of data related to the health sector. The high vulnerability of such data to harm the data principals should act as a deterrent in using this as the pilot use-case. Given the mass availability of data related to the health sector due to the pandemic, it creates further points of vulnerabilities which can be illegally monetised and misappropriated. It is recommended that this proposal be scrapped altogether.

Clause 7.2 (iii): Power of government bodies

As per this clause, data trustees or government bodies (who could also be acting as data trustees) can make requests for data sharing and place such data in appropriate data infrastructures or trusts. This presents a conflict of interest, as a data trust or government body can empower itself to be the data trustee. Such cases should be addressed within the scope of the framework.

Clause 8.2 (vii): Level-playing field for all Indian actors

In terms of this clause the "Non-Personal Data Authority (Authority) will ensure a level playing field for all Indian actors to fulfil the objective of maximising Indian data's value to the Indian economy". The emphasis on ensuring a level playing field for only Indian actors instead of non-discriminatory platform for all concerned actors irrespective of the country/nationality of the actor has the potential of violating India's trade obligations under the WTO. Member states of the WTO are essentially restricted from discriminating between products and services coming from different WTO Members, and between foreign and domestic products and services unless they can avail of exceptions. There is also no clarity on what constitutes 'Indian Actors', would a Multi-National Corporation with its headquarters in a foreign State, but its subsidiaries in India also come within its ambit.

Clause 8.2 (x): Composition of the Authority

Clause 8.2 (x) states that the Authority will have some members with relevant industry experience. However, apart from this clause, the report is silent on the composition of the Authority. The report recognises that Authority will need individuals/organisations with specialised knowledge, i.e. data governance, technology, latest research and innovation in the field of non-personal data), however, it does not mention or refer to the role of civil society organisations and the need for representation from such organisations in the Authority.

The report frequently alludes to non-personal data being used for the best interest of the data principal and therefore, it is essential that the composition of the Authority reflect the inherent asymmetry of power between the data principal and the State. Considering that the

Authority will also be responsible for sharing of community data and with determining the code of conduct for sharing of such data, it is important that the Authority also has adequate representation from civil society organisations along with groups or individuals having the necessary technological and legal skills.

Clause 8.2 (iii) and (vi): Roles and Responsibility of the Authority

A majority of the datasets in the country comprise of 'mixed datasets', i.e. it consists of both personal and non-personal data. However, there is lack of clarity about the coordination between the Data Protection Authority constituted under the PDP Bill and the Non-Personal Data Authority with regard to the regulation of such datasets. The Report refers to the European Union which provides that the Non-Personal Data Regulation applies to the Non-Personal Data of mixed datasets; if the Non-Personal Data part and the personal data parts are 'inextricably linked', the General Data Protection Regulation apply to the whole mixed dataset. However, it is unclear whether the Report also proposes the same mechanism for the regulation of mixed datasets.

Further, the contours of the enforcement role of the Committee should be specified and clearly laid down. Will the Committee also have penal powers as prescribed for the Data Protection Authority under the PDP Bill? Also, will the privacy concerns emanating from the risk of re-anonymisation of data be addressed by the NPD Committee or by the DPA under the PDP Bill. Ideally, it should be specified that any such privacy concerns will fall within the domain of the DPA as the data is then converted into personal data and the DPA will be empowered to deal with such issues.