Comments on the Report on the Non-Personal Data Governance Framework

January 31, 2021

By: **Aman Nair, Pallavi Bedi and Anubha Sinha**

Reviewed by **: Arindrajit Basu**

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 second version of the Report on Non-Personal Data Governance Framework prepared by the Committee of Experts (hereafter "Report").[1] CIS had also provided inputs to the draft version of the Report published in July 2020. [2]

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 comments below are organised thematically, with references to specific parts of the Report highlighted.

Inputs

1. Examining the economic considerations underpinning the non-personal data governance framework

- **Open Data access is not enough to offset network effects and existing power imbalances in key digital sectors**

The purpose behind sharing and regulating Non Personal Data (**NPD**), as outlined in section 8.iv of the Report, is to ensure "*maximum social and economic benefits from data for the society.*" On the economic side, non personal data sharing is envisioned as a means of spurring innovation and of establishing the anticompetitive nature of the digital marketplace. However, the Report fails to consider that many sectors of the digital market are natural monopolies and that data sharing in the face of existing power structures within the digital marketplace would fail to adequately offset the network effects enjoyed by key players (eg. Facebook, Amazon, Google).[3] As such, in the absence of either breaking up such organisations or -more preferably-

[1] Report by the Committee of Experts on Non-Personal Data Governance Framework- December 16, 2020; https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

[2] Aayush Rathi, Aman Nair, Ambika Tandon, Pallavi Bedi, Sapni Krishna, Shweta Mohandas. 'Inputs to the Report on the Non-Personal Data Governance Framework' September 13, 2020, available at https://cis-india.org/raw/files/cis-inputs-to-report-on-non-personal-data-governance-framework

[3] Fiona M. Scott Morton and David C. Dinielli, "Roadmap for an Antitrust Case Against Facebook" (Omidyar Network, June 2020),

regulating them to the standard of a public utility, data sharing would do little to create any additional competitive benefits.

- **Increased Data collection leads to Data Appropriation**

The Report fails to address issue of data appropriation[4] - whereby the economic benefits of gathering data are preferentially distributed to the corporations collecting data as opposed to the citizen's whose data is being collected. This can take either the form of social media companies amassing hordes of data under the guise of providing 'free services'[5] in return, or private corporations utilising non personal data of a community to improve their good or service that in turn are priced in a manner that is inaccessible to that community.

2. Addressing the societal concerns that arise with sharing Non Personal Data sharing

- **De-anonymization and harm linked with sharing Non Personal Data**

The Report fails to adequately address the concerns regarding re-identification in the case of Non Personal Data. In the absence of an overarching data protection bill (at the time of writing) there is no clear indication as to the standards of encryption and data protection that must be afforded to the various kinds of non personal data. 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

---

https://www.omidyar.com/wp-content/uploads/2020/06/Roadmap-for-an-Antitrust-Case-Against-Facebook.pdf.
"Investigation of Competition in Digital Markets" (Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, 2020), https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf.

[4] Jathan Sadowski, "Companies Are Making Money from Our Personal Data – but at What Cost?," *The Guardian*, August 31, 2016, sec. Technology, http://www.theguardian.com/technology/2016/aug/31/personal-data-corporate-use-google-amazon.

[5] It should be noted that this conception of free is actually false, as the Competition Law Review Committee in July 2019 found that data is subsumed under the current law's definition of price. "Report of the Competition Law Review Committee" (Ministry of Corporate Affairs, Government of India, July 2019), http://www.mca.gov.in/Ministry/pdf/ReportCLRC_14082019.pdf.

[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

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.

Furthermore, the Report ignores that Non Personal Data that has been aggregated can cause unintended harms against individuals or communities- especially when such data is related to a particular geographic region or group of individuals. For example, the illustration on page 9 of the Report outlines examples of non personal data collected by private entities, and notes that a private mobile operator could collect data on individual's locations to ensure that they are in compliance with government quarantine regulations. While it is undoubtedly useful information in the hands of government officials, introducing such data to either the public or to businesses could result in the very real phenomenon of *Lateral Surveillance*.[7] Such Lateral Surveillance could prove incredibly dangerous as it both violates an individual's right to privacy and also disproportionately affects the less privileged.

- **Sharing non-personal data could result in a culture of data maximisation**

The Report fails to recognise that much of the societal harms resulting from the current system of data gathering is caused due to the business model of digital organisations that gain exponentially from such collection. As such a solution of making data available to all does little to remove the incentive for private organisations to collect and monetise as much data as possible. This leads to a clear culture of data maximisation in the private sector, that is a clear departure from the strategy of data minimisation that is being adopted in jurisdictions such as the EU.[8] This could actively hamper purpose limitation and could lead to function creep, especially in cases of public services.

3. Section 7.2-Non-Personal Data Roles- Community

- **Vague and very wide definition of Community**

The Committee has defined a community "*as any group of people that are bound by common interests and purposes, and involved in social and/or economic interactions. It could be a geographic community, a community by life, livelihood, economic interactions or other societal interests and objectives, and/or an entirely virtual community*." The Community (through a non-profit organisation, section 8 company, trust, society should be able to raise

[7] Mira Swaminathan and Shubhika Saluja, "Watching Corona or Neighbours? - Introducing 'Lateral Surveillance' during COVID-19," *Centre for Internet and Society* (blog), May 21, 2020, https://cis-india.org/internet-governance/blog/essay-watching-corona-or-neighbours-introducing-2018lateral-surveillance2019-during-covid201919.
[8] JT Sison, "Data Minimization in the GDPR: A Primer," *Dataguise* (blog), February 15, 2017, https://www.dataguise.com/gdpr-compliance-data-minimization-use-purpose/.

a complaint with a regulatory authority about harms emerging from sharing non-personal data about their community.

From the definition of community, it appears that as per the Committee, a community is one homogenous of individuals with similar interests and purposes. It does not take into account the heterogeneous and at times disparate nature and interest of the concerned individuals. As highlighted by Dvara Research in its submission to the first report[9], *'the Report assumes that members of communities or groups share some socially constructed, physical and/or behavioural characteristics and individuals are aware of their membership to a community or group.'* Such an assumption is not accurate and it does not recognise the competing interests of different actors within a particular community. As articulated by Vidushi Marda in her submission to the first report, "*A 'community' can be formed along different criteria. Some of these criteria flow through pre-identified groups or are visible and explicit, like the community arising from a particular geographic area, while other communities may come into existence unbeknown to the very individuals that make up that community through the process of algorithmic profiling and analytics.*" The Report does not take into account the power dynamics within communities, the social inequalities and how the power asymmetry within members of a 'community' could exclude them from participating in the discussion on the sharing of community data.

- **Section 7.7- Data Trustee**

Section 7.7 defines Data Trustee '*either a government organisation or a non-profit private organisation for the creation, maintenance and data-sharing of the High-value Datasets in India.*' The data trustee has to ensure that the HVDs are used only for the interests of the community and that to further ensure that the persons do not suffer any harm on account of re-identification of their non-personal data.' The Report has proposed the creation of HVD and has defined HVD (Section 7.6) as a dataset that is beneficial to the community at large and shared as a public good.

We appreciate that the Committee has recognised the concerns raised with the previous report regarding the lack of clarity on the process of establishing a data trustee, and has now specified that the Authority will set out the guidelines to determine the appropriateness of the HVD and the data trustee. However, the Report does not define the term 'public good' or 'harm' It also does not provide any guidance on the activities which could be regarded as harmful.

[9] Dvara Research, "Response dated 13 September 2020 to the Report by the Committee of Experts on Non-Personal Data Governance Framework released by the Ministry of Electronics and Information Technology in July 2020," https://www.dvara.com/research/wp-content/uploads/2020/10/Our-Response-to-the-Report-of-the-Committee-of-Experts-on-Non-Personal-Data.pdf

4. Section 7.4(iv)- 'Duty of care' of data custodian

Section 7.4 (iv) of the Report places a duty of care on the data custodians. It states that the data custodian has a responsibility to ensure that no harm occurs to persons / groups of persons by re-identification of non-personal data. It further also states that the Committee distinguishes between 'active misuse of NPD' and 'accidental misuse of NPD'. However, the Report does not define or provide any guidance on what actions would constitute 'accidental misuse of NPD' and/or 'active misuse of NPD.' The Report also does not define 'harm.'

We also reiterate our earlier comment to state that 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.

5. Section 7.10 -Non-Personal Data Authority

- **Composition of the Authority**

Section 7.12 (i) of 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 community 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.

- **Roles and Responsibility of the Authority**

The Report is unclear about the roles and responsibility of the Authority. It does not specify whether the Authority will also have penal powers as prescribed for the Data Protection Authority under the PDP Bill

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 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. There is also no clarity on whether the Authority will also have quasi-judicial powers. if it will be a judicial authority or a regulator with powers of a civil court vested upon it. These should be clarified. The functions

of the Authority should be clearly spelled out as well e.g. SEBI, IRDA are sectoral regulators with clearly spelled out functions in their respective legislations[10]

6. Section 9.3 - Copyright Law

- **Failure to recognise copyright in underlying data of datasets**

The Report claims that creation of high-value datasets will not infringe copyright of the raw datasets, as only pre-designated sub-fields will be extracted. However, on page 8, the Report highlights two examples of non-personal data where the underlying data are videos and photos of activities in the public sphere (and captured by private parties). The underlying works in a database of such non-personal data are copyrightable. It is not clear how extraction of pre-designated subfields will apply or translate into high-value dataset without infringing on copyrights of the private parties.

- **Consider advocating use of limitations and exceptions in copyright law to limit ownership in datasets and underlying data**

Limitations and exceptions are certain types of provisions in copyright law which enable users to use copyrighted works without permission from the copyright owner. The Report fails to recognise this public interest enabling mechanism of copyright law, which is very much in alignment with the regulatory goals of the Committee in terms of increasing innovation and data-sharing. Many AI techniques involve 'learning', which in turn requires technical reproduction of data. This reproduction is widely permissible in copyright laws, especially if done in the context of research and for non-commercial purposes. The Committee should recognise and support such existing uses in research, and expand the ambit of the use of non-personal data in copyright law by supporting its permissionless use in commercial purposes, as well.[11]

[10] SLFC.IN's comments on Non-Personal Data Governance Framework, available at https://sflc.in/our-comments-nonpersonal-data-governance-framework-report, accessed on January 22, 2021

[11] Arul Scaria, "Should Indian Copyright Law Prevent Text and Data Mining?", available at https://spicyip.com/2019/08/should-indian-copyright-law-prevent-text-and-data-mining.html accessed on January 30, 2021