॥न्यायस्तत्र प्रमाणं स्यात्॥

NATIONAL LAW UNIVERSITY DELHI

Prof. (Dr.) Ranbir Singh
Vice-Chancellor

September 11, 2020

**Shri Ajay Prakash Sawhney**
**Secretary**
**Ministry of Electronics and Information Technology**
**Government of India**
Electronics Niketan, 6, CGO Complex,
New Delhi – 110003

**Subject: Submission of Comments on Report by the Committee of Experts on Non-Personal Data Governance Framework**

Dear Mr. Sawhney,

The National Law University Delhi is a publicly funded university established by the government of the National Capital Territory of Delhi on the initiative of the High Court of Delhi, via Act No.1 of 2008 of National Capital Territory of Delhi. The Chief Justice of India is a Visitor to the University and the Chair of the Governing Council. The Chief Justice of High Court of Delhi is the Chancellor of the University.

The Centre for Communication Governance (CCG) was established by the University in 2013 to ensure that Indian legal education establishments engage more meaningfully with information law and policy, and contribute to improved governance and policymaking. CCG is the only academic research Centre dedicated to working on information law and policy in India.

We welcome the opportunity to comment on the Report by the Committee of Experts on Non-Personal Data Governance Framework and commend the Ministry under your leadership for adopting a public and consultative approach to the whole process.

(A University of Law Established in 2008 by Act No. 1 of 2008 of Delhi)
Sector-14, Dwarka, New Delhi - 110 078 ☎ :+91 11 2803 4253 🖷:+91 11 2803 4256
: ranbirsingh2107@gmail.com, vc@nludelhi.ac.in website: www.nludelhi.ac.in

As part of our work, and given how critical it is to provide lawmakers with useful material, we have drafted comments to the Report by the Committee of Experts on Non-Personal Data Governance Framework which are enclosed herewith.

We have previously engaged extensively on the issue of Data Protection and have submitted comments to the *White Paper of the Committee of Experts on a Data Protection Framework for India* lead by *Hon'ble Mr. Justice B. N. Srikrishna* and also submitted detailed comments to the Ministry of Electronics and Information Technology (MeitY) on the *Draft Personal Data Protection Bill, 2018* and to the Joint Parliamentary Committee on the *Personal Data Protection Bill, 2019*.

We hope that these comments are of assistance to the committee in formulating its final report and recommendations on the issue of utmost importance, and we will be will be happy to find any additional material and provide any assistance from our end to the committee.

I also request you to provide an opportunity to my colleagues **Ms. Smitha Krishna Prasad** (smitha.prasad@nludelhi.ac.in; +91-9980175343) and **Mr. Sarvjeet Singh** (sarvjeet.singh@nludelhi.ac.in; +91-9990232298) to present our work before the committee.

With warm regards,

Yours sincerely,

Ranbir Singh

(A University of Law Established in 2008 by Act No. 1 of 2008 of Delhi)
Sector-14, Dwarka, New Delhi - 110 078 ☎ :+91 11 2803 4253 🖷:+91 11 2803 4256
: ranbirsingh2107@gmail.com, vc@nludelhi.ac.in website: www.nludelhi.ac.in


# CENTRE FOR COMMUNICATION GOVERNANCE AT NATIONAL LAW UNIVERSITY DELHI

## COMMENTS TO MEITY ON THE REPORT BY THE COMMITTEE OF EXPERTS ON NON-PERSONAL DATA GOVERNANCE FRAMEWORK[°]

nludelhi.ac.in | ccgdelhi.org | ccg@nludelhi.ac.in

[°] Authored by *Jhalak Kakkar, Kritika Bhardwaj, Shashank Mohan* and *Smitha Krishna Prasad*. Reviewed and edited by *Smitha Krishna Prasad* and *Sarvjeet Singh*.

The Centre for Communication Governance is an academic research centre within the National Law University Delhi and is dedicated to working on information law and policy in India. It seeks to embed human rights and good governance within communication policy and protect digital rights in India through rigorous academic research and capacity building.

We are grateful to the Ministry of Electronics and Information Technology for inviting public comments and suggestions on the Report by the Committee of Experts on Non-Personal Data Governance Framework.

## I. Introduction

There is increasing consensus that a handful of technological platforms (often referred to as "**Big Tech**"[1]), have come to wield significant power today.[2] The extent of this power is visible not just in terms of how such platforms have come to undermine competition[3], but also processes that are integral to a functioning democracy, such as elections[4] and free speech.[5] These businesses have typically grown as a result of almost unrestricted access

[1] Urvashi Aneja and Angelina Chamuah, 'A Balancing Act, The Promise & Peril of Big Tech in India', Tandem Research (2020) <https://tandemresearch.org/assets/Tandem-Research-Big_Tech_report.pdf>.

[2] Lina M. Khan, 'Sources of Tech Platform Power' 2 Georgetown Law Tech. Rev. 325 (2018) <https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Khan-pp-225-34.pdf>; Michael A. Cusumano, David B. Yoffie & Annabelle Gawer, 'The Future of Platforms' 61(3) IT Sloan Management Review (2020) <https://sloanreview.mit.edu/article/the-future-of-platforms/>.

[3] Lina M. Khan, 'The Separation of Platforms and Commerce' 119 Columbia L.R. 973 (2019) <https://columbialawreview.org/wp-content/uploads/2019/05/Khan-THE_SEPARATION_OF_PLATFORMS_AND_COMMERCE-1.pdf>; Narayan Lakshman, 'Why are Amazon, Google, Facebook and Apple facing antitrust issues?' *(The Hindu*, 15 December 2019) <https://www.thehindu.com/sci-tech/technology/why-are-amazon-google-facebook-and-apple-facing-antitrust-issues/article30307356.ece>; Cecilia Kang, Jack Nicas and David McCabe, 'Amazon, Apple, Facebook and Google Prepare for Their 'Big Tobacco Moment' (*New York Times*, 28 July 2020) <https://www.nytimes.com/2020/07/28/technology/amazon-apple-facebook-google-antitrust-hearing.html>.

[4] Scott Dtrow, 'What Did Cambridge Analytica Do During The 2016 Election?' (*NPR*, 20 March 2018), <https://www.npr.org/2018/03/20/595338116/what-did-cambridge-analytica-do-during-the-2016-election>; The Electoral Commission, 'Investigation: Leave.EU Group Limited' <http://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-enforcement-work/investigations/investigation-leaveeu-group-limited>.

[5] Sanja Kelly and others, 'Manipulating Social Media to Undermine Democracy' (Freedom on the Net, 2020) <https://freedomhouse.org/report/freedom-net/2017/manipulating-social-media-undermine-democracy>;

to personal data, and by designing services based on the monetization of such data.[6] In this light, the Report published by the Committee of Experts (the "**Committee**") on Non-Personal Data Governance Framework (the "**Report**") is a welcome step insofar as it seeks to challenge and address the imbalance in bargaining and market power created by such platforms. It is further encouraging to see the Report recognise rights and entitlement of people and communities over the data collected from them.

However, a few of the Committee's proposals for the regulation of non-personal data merit further scrutiny.

These comments set out our concerns with respect to some of the Committee's recommendations.

In this first part of the comments, we address the broad concerns that relate to the scope and objectives of the Report itself. In Part II of this document, we address concerns in the context of the purposes for sharing of non-personal data, the categorisation of non-personal data and the institutional structure and the role of different stakeholders as contemplated in the Report. In Part III we address the specific areas where the Report's proposed framework for regulation of non-personal data intersects with the Personal Data Protection Bill, 2019. In Part IV we highlight certain additional concerns that should be incorporated into further undertakings in relation to any regulation of non-personal data.

## A. Case for Regulation: Lack of Evidence and Legal Basis

While the terms of reference of the Committee were to study issues relating to, and make suggestions for the regulation of non-personal data[7], we note that the Committee largely derives the rationale for such regulation from the 'imbalance in data and digital industry'.[8]

Roger Macnamee, 'Big Tech Needs to Be Regulated. Here Are 4 Ways to Curb Disinformation and Protect Our Privacy (The Time, July 29, 2020) <https://time.com/5872868/big-tech-regulated-here-is-4-ways/>.

[6] Urvashi Aneja and Angelina Chamuah, 'A Balancing Act, The Promise & Peril of Big Tech in India', Tandem Research (2020) <https://tandemresearch.org/assets/Tandem-Research-Big_Tech_report.pdf>.

[7] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 1.1.

[8] ibid paras 3.6, 3.7.

As already set out above, the Report correctly observes that a handful of digital platforms have come to exert substantial dominance in the market owing to a combination of first mover advantage, unrestricted access to data and network effects.[9] However, the specific recommendations of the Committee do not clarify how these will address the problems arising out of such dominance.

A recent report on the dominance of technological platforms in India calls for a graded approach to regulation based on different challenges that such platforms pose.[10] These include competition law reform, platform neutrality as well as interoperability, and long-term interventions such as digital literacy.[11]

To the extent that the Report considers the dominance of a few digital platforms as the basis for regulating non-personal data, we note that such concerns may be better addressed through platform specific investigations to understand the underlying systemic issues[12], which may then be addressed through specific legal and regulatory mechanisms. We note that this also appears to be the approach advocated by other jurisdictions with respect to regulation of non-personal data.[13]

Further, inasmuch as the Report proposes to create a regime for mandatory data sharing between businesses, communities and the government in order to unlock the "social / public / economic" value of data[14], we note that the Committee has not provided justifiable legal basis for some of its recommendations.

[9] ibid paragraph 3.7(iv).

[10] Urvashi Aneja and Angelina Chamuah, 'A Balancing Act, The Promise & Peril of Big Tech in India', Tandem Research (2020) <https://tandemresearch.org/assets/Tandem-Research-Big_Tech_report.pdf>.

[11] ibid.

[12] European Commission, *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' (Feb. 19, 2020) p. 14 <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy#documents>.

[13] ibid.

[14] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020), Key Take-aways- Case for regulating data p 11.

For example, with respect to Public Non-Personal Data, the Committee rightly notes that given that such datasets are derived from public efforts and are funded through public money, they assume the characteristics of a national resource.[15] Further, there is also a public interest element in requiring the state to share the information collected by it, as it increases transparency and accountability.[16]

Similarly, the Committee further correctly observes that given that data held by private entities has been collected from individuals and communities, communities have a legitimate interest in accessing and using such data.[17]

However, it is not clear how the above-mentioned legal basis can be used to justify acquisition of privately held non-personal data by other private entities for the latter's commercial gain. The Report fails to clarify how this mandatory acquisition of data will not conflict with other laws such as the intellectual property rights of private entities over non-personal data stored with them.

The Committee's proposals for a law to mandate horizontal data sharing between businesses appear to be in the nature of *eminent domain* – the right exercised by the state to take over private property in the interest of public utility.[18] However, the principle of eminent domain not only has a legal basis under the Constitution of India itself[19], but also requires that the acquisition of private estate must be for a public purpose.[20] The Supreme Court of India has clarified that such public purpose must be '*primarily public and not primarily of private interest and merely incidentally beneficial to the public*.'[21]

---

[15] Ibid. para 5.2.

[16] Renuka Sane and Rishabh Bailey, 'A Missed Opportunity' (*The Hindu*, 3 September 2020) <https://www.thehindu.com/opinion/op-ed/a-missed-opportunity/article32507522.ece>.

[17] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 5.3.

[18] *KT Plantation Private Limited & Anr. v State of Karnataka* (2011) 9 SCC para 134; Priya S. Gupta., 'The Peculiar Circumstances of Eminent Domain in India' (49) Osgoode Hall Law Journal (2012), 445-489.

[19] Constitution of India, art. 31-A.

[20] *KT Plantation Private Limited & Anr. v State of Karnataka* (2011) 9 SCC para 180; Namita Wahi, 'Land Acquisition, Development and the Constitution' Seminar Magazine (February 2012) <http://india-seminar.com/2013/642/642_ namita_wahi.htm>.

[21] *KT Plantation Private Limited & Anr. v State of Karnataka* (2011) 9 SCC para 180.

Another requirement of such acquisition is the right to receive compensation in exchange for one's personal estate.[22]

In the present context, we note that a requirement to share non-personal data with startups or other businesses for commercial exploitation is unlikely to meet the test of public purpose.

In addition to the above, data is not necessarily property, but is understood differently in different contexts. It is an aspect of personhood in the context of personally identifiable information, but is also understood as property under the intellectual property rights regime.

Therefore, in the absence of any justifiable legal basis, the Committee must reconsider its proposal for mandatory transfer of data by held companies for use by other private companies and businesses.

The Report also fails to consider the impact of such a proposal on innovation and companies' protected intellectual property rights and trade secrets. It further fails to explore any incentives that may be offered to make such data sharing mutually beneficial. These concerns have been set out in detail in Section II.A.3 (Economic Purposes) of this document.

Similarly, other claims made by the Report in support of its recommendations also merit a closer scrutiny. For example, the Report notes that 'abundant availability of data is a primary driver for AI [Artificial Intelligence]' (and therefore, access to non-personal data will help increase its revenue from analytics and machine learning services).[23] However,

[22] *KT Plantation Private Limited & Anr. v State of Karnataka* (2011) 9 SCC para 192; *R.C. Cooper v Union of India* para 112; Namita Wahi, 'Land Acquisition in India: A Review of Supreme Court Cases from 1950 to 2016' Centre for Policy Research (2017), <https://www.cprindia.org/system/tdf/policy-briefs/Land%20Rights%20Report%20Final.pdf?file=1&type=node&id=5891&force=1>.

[23] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 3.2. ii.

there is emerging consensus to suggest that meaningful use of Artificial Intelligence will increasingly depend less on data, and more on how the systems are trained.[24]

The Report similarly also fails to substantiate its recommendation that community rights can only be exercised through a trustee and assumes that the trustee will (always) act in the community's best interests. This concern has been expanded upon below in Sections II.B.3 and II.C.1 below.

### B. Privacy and Surveillance Risks

A large part of the data contemplated to be regulated under the proposed non-personal data regime (as public, community or private non personal data) is actually personal data that has been anonymised or stripped of any identifiers by other means.[25] This raises significant privacy and surveillance concerns, which have not been adequately addressed in the Report.

As the Report itself notes, there is increasing consensus that anonymising data successfully and permanently is rarely possible.[26] Further, the risk of re-identification of personal data increases exponentially in the present context as the proposed regime for non-personal data contemplates indiscriminate sharing of data sets and potentially combining them with data gathered or collected from other sources.[27]

---

[24] H. James Wilson , Paul R. Daugherty and Chase Davenport, 'The Future of AI Will Be About Less Data, Not More' (*Harvard Business Review*, 14 January 2019) <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more>.

[25] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.1. iii.

[26] Scott Berinato, 'There is no such Thing as Anonymous Data' (*Harvard Business Review*, February 9, 2015) <https://hbr.org/2015/02/theres-no-such-thing-as-anonymous-data>; Alex Hern, 'Anonymised' data can never be totally anonymous, says study' (*The Guardian*, July 23, 2019) <https://www.theguardian.com/technology/2019/jul/23/anonymised-data-never-be-anonymous-enough-study-finds>; Gina Kolata, 'Your Data Were 'Anonymized'? These Scientists Can Still Identify You' (*New York Times*, July 23, 2019) **<**https://www.nytimes.com/2019/07/23/health/data-privacy-protection.html>.

[27] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.6. i.

While the Report is cognisant of this risk, it is almost entirely silent on how to prevent or mitigate these risks. The limited recommendation of the Committee is to require individuals to consent to the use of their anonymised personal data at the time of collection of their personal data itself[28], and to allow them to "act upon" harms that may arise as a result of subsequent re-identification.[29]

These recommendations fail to address the concerns arising out of re-identification in any meaningful manner. First, the principle of consent under personal data protection exists to protect the rights of data principals; it is intended to ensure that personal data is processed only after individuals give their express, informed and specific consent for a particular purpose.[30] The requirement of a specific and limited purpose is therefore inherent to the principle of consent. This also follows from the definition of consent under the Personal Data Protection Bill, 2019 ("**PDPB 2019**").[31] Therefore, requiring individuals to consent to blanket use of their data (albeit anonymised) by any entity for any purpose defeats the very idea of consent.

Secondly, the suggestion that data principals can act upon the harms arising out of re-identification disregards the fact that in a regime where the same data set may change multiple hands, a data principal may never get to know that their data has been de-anonymised. Further, the Report fails to acknowledge that re-identification in itself is a serious harm, as an individual's personal data has become known or available to a data principal without her consent and despite not falling under any of other permissible grounds under the PDPB 2019. In *K.S. Puttaswamy v. Union of India* ("***Puttaswamy***")[32], the Supreme Court of India recognized informational self-determination and autonomy of choice as being fundamental to the guarantee of privacy.[33] Therefore, disclosure of

[28] ibid. para 4.1. iv.

[29] ibid. para 4.1. ii.

[30] Council Regulation (EC) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, recital 32, art 7.

[31] The Personal Data Protection Bill 2019, cl 11(2)(c).

[32] *K.S. Puttaswamy v Union of India* (2017) 10 SCC 1.

[33] ibid. paras 169, 190 (Dr. D.Y. Chandrachud, J.).

personal data in the absence of the data principal's knowledge or consent in itself is a violation of the fundamental right to privacy.

The Committee's recommendations appear to entirely absolve the entities exchanging the data in the event of re-identification and places the entire burden of remedying such risks / harms on the individual. There is no requirement for entities acquiring non-personal data to register the fact of having acquired particular data, and state its intended use for it, which can be used to make entities accountable in the event of re-identification or breach. As already pointed out above, the affected individual may not only not be aware of such a breach but may also lack the adequate means to chase and challenge processing of her data by unknown entities for unknown purposes.

Another significant privacy concern arises out of the Committee's recommendation that non-personal data can also be used for sovereign purposes such as national security, and by law enforcement for criminal investigations. At the outset, this sits at odds with the Committee's rationale for regulating non-personal data, viz. to increase transparency, incentivise innovation, and generate economic and social benefits for citizens and communities in India.[34] Further, the legal regime for surveillance under the Telegraph Act, 1885 ("**Telegraph Act**"), the Information Technology Act, 2000 ("**IT Act**") as well as several sectoral legislations already grants the State wide powers with little external checks on how such power is exercised.[35] In addition to these powers, the creation of databases such as the Central Monitoring System ("**CMS**")[36], the National Intelligence Grid ("**NATGRID**")[37] as well as Crime and Criminal Tracking Network and Systems

[34] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 3.9.

[35] Committee of Experts under the Chairmanship of Justice B.N. Srikrishna, 'A Free and Fair Digital Economy, Protecting Privacy, Empowering Indians' ("Justice Srikrishna Committee Report") <https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf>; Chaitanya Ramachandran, 'PUCL v. Union of India Revisited: Why India's Surveillance Law Must Be Revised for the Digital Age' (7) NUJS L.R. (2014).

[36] Sneha Johari, 'Govt's Central Monitoring System already live in Delhi & Mumbai' (*Medinama*, 11 May 2016) <https://www.medianama.com/2016/05/223-india-central-monitoring-system-live-in-delhi-mumbai/>.

[37] Anirudh, 'NATGRID: Should Parliament have a role?' (PRS, June 20, 2011) <https://www.prsindia.org/theprsblog/natgrid-should-parliament-have-role>; Internet Freedom Foundation,

("**CCTNS**")[38] outside the framework of any legislation have already led to the amassing of personal data without any corresponding rights for data principals or accountability mechanisms. Further, information about what data is collected, stored and how it is used is already exempt from disclosure under the Right to Information Act, 2005[39], and is further being sought to be exempted from the purview of the PDPB 2019.[40] The use of non-personal data for national security and related purposes is therefore only likely to strengthen the increasing surveillance architecture by allowing the state to combine personal data collected from other sources with non-personal data. In *Puttaswamy,* the Supreme Court expressly recognised the threat to privacy caused by aggregating distinct silos of information.[41]

In the absence of any judicial or other independent oversight over state surveillance, the threat posed by using non-personal data for sovereign purposes also does not appear to satisfy the tests of necessity or proportionality laid down by the Supreme Court.[42]

Troublingly, the Report does not identify or propose any limitations to the use of non-personal data for sovereign purposes, including restrictions on the combining of data, or requiring data processing to be necessary or proportional.[43] This has serious implications for not only individuals' right to privacy, but also their right against self-incrimination. These concerns are set out more fully below in Section II.A.2 below.

The Committee has understandably attempted to ring fence its proposals by focusing on only the value of and use of non-personal data. However, in doing so, the Report ends

---

'Watch the Watchmen Series Part 1: The National Intelligence Grid' (Sept. 2, 2020) <https://internetfreedom.in/watch-the-watchmen-part-1-the-national-intelligence-grid/>.

[38] Internet Freedom Foundation, 'NCRB finally responds to legal notice on facial recognition, we promptly send a rejoinder' (Nov 8, 2019) <https://internetfreedom.in/the-ncrb-responds/>.

[39] Right to Information Act, 2005, s. 8(1)(a).

[40] The Personal Data Protection Bill, cl 36.

[41] *K.S. Puttaswamy v Union of India* (2017) 10 SCC 1 paras 300, 589-594.

[42] *K.S. Puttaswamy v Union of India* (2017) 10 SCC 1 paras 325, 638.

[43] Anubhutie Singh, Malavika Raghavan, Beni Chugh & Srikara Prasad, 'The Contours of Public Policy for Non-Personal Data Flows in India' (*Dvara Research*, Sept. 24, 2019) <https://www.dvara.com/blog/2019/09/24/the-contours-of-public-policy-for-non-personal-data-flows-in-india/#_ftn1>.

up relegating significant privacy and surveillance risks to mere implementation challenges. However, inasmuch as a large part of the data envisaged as non-personal data is actually anonymised personal data, addressing the significant privacy concerns must be the starting point for any proposal for the regulation of non-personal data. For this and other reasons, any proposal for the regulation of non–personal data must necessarily await the enactment of a data protection legislation in India, and must necessarily be in compliance with such legislation.

### C. No Alternatives Considered

Before responding to the specific suggestions proposed by the Committee, we also note that the Report fails to consider, or even set out other possible alternatives for regulating non-personal data.

In this context, it is useful to note that the strategy for data proposed by the European Commission (the "**Commission**") specifically warns against heavy handed *ex-ante* regulation and proposes voluntary data sharing as the general principle.[44] Instead, the Commission has proposed a host of regulatory interventions, such as prioritising standardisation to facilitate data interoperability and identifying high-quality public sector data sets that can be made available for research and commercial purposes.[45] The Commission has also resolved to explore incentives for horizontal data sharing between governments and businesses with the caveat that '*the general principle shall be to facilitate voluntary data sharing*', except when specific circumstances dictate that access to data should be made mandatory.[46]

Similarly, the United Kingdom's Digital Competition Expert Panel ("**UK Expert Panel**") has also laid emphasis on facilitating data mobility and open standards to enable

[44] European Commission, 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://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy#documents>.

[45] ibid. pp 12, 13.

[46] ibid.

competition and innovation.[47] Recognising that access to data is one of the key barriers to entry in the digital market, the UK Expert Panel also recognised that privately held data may need to be mandatorily shared in certain circumstances, albeit following a thorough analytical assessment and only if less interventionist solutions have failed.[48]

The European Commission further recommends developing common data spaces in certain strategic sectors such as manufacturing, mobility, health and energy.[49] A similar sector-specific approach may be considered by the Committee since that will allow regulatory interventions to be designed according to the needs of the sector and the sensitivity of the data involved. While stricter ex-ante regulation may be required in the health or financial services sector, data pertaining to manufacturing or weather may be made available more easily.

It is not clear why the Committee does not consider or propose adopting open standards, facilitating data portability and interoperability, which will not only lead to controlling network effects but also more choices and options for users.

While the Committee's recognition of competing rights and interests over privately owned data is welcome, it's one size fits all proposal does not appear to be supported with evidence and risks overregulation.

[47] UK Digital Competition Expert Panel, 'Unlocking Digital Competition: Report of the Digital Competition Expert Panel' (March 2019) <https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel>.

[48] ibid.

[49] European Commission, 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://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy#documents> pp 21 - 23.

## II. Analysis of the Proposed NPD Regulatory Framework

### A. Purposes for which NPD can be Shared and the Challenges

The Committee defines Non-Personal Data ("**NPD**") as any data that is not related to an identifiable or identified natural person, or personal data that has been anonymised.[50] The Committee has further classified such data into three categories: i. Public NPD, ii. Private NPD and iii. Community NPD.[51]

The Report proposes a regulatory framework for data-sharing and access to these three categories of NPD for three broad purposes: (i) public interest purposes, (ii) sovereign purposes, and (iii) economic purposes.[52]

#### 1. Public Interest Purposes

The Report states that the government may request data for core public interest purposes including "community benefits or public goods, research and innovation, policy making, for better delivery of public-services, etc."[53] The Report underlines that while public agencies produce significant amounts of data, a large part of the required data is collected by the private sector.[54] In the context of access of the government to NPD for public interest purposes from data businesses, the following issues merit consideration:

##### (a) *A Balance between Incentive-Based Approach and Mandatory Data Sharing*

We welcome the Committee's attempt to enable the government to undertake data-driven and evidence-based policymaking. The Committee is seeking to put in place mechanisms that enable the government to access data for core public

[50] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 13.

[51] ibid. pp. 14 – 15.

[52] ibid. pp. 32 – 36.

[53] ibid. p. 32.

[54] ibid. p. 9.

interest purposes without having to rely on instances of data philanthropy by data businesses.

While we do need to move beyond data philanthropy, a blanket coercive approach of mandatory data sharing by data businesses with the government, as suggested in the Report, raises concerns of distorting market incentives. In certain cases, alternative approaches can be explored, such as encouraging data sharing by data businesses by providing them with various incentives such as tax benefits or preferential access to certain government datasets. In other policy documents, the government has recommended the provision of incentives such as direct and indirect tax benefits, and custom's duty rebates to encourage specific actions by companies.[55]

Additionally, specific circumstances can be evolved where mandatory sharing of data by data businesses may be appropriate. Broad public consultation, detailed discussion with the relevant stakeholder, impact investment and pilots need to be undertaken to evolve an appropriate approach, and find instances where there is a need for mandatory sharing of data for public good.

**(b) *Effective Utilisation of Existing Government Datasets***

The Indian government holds large amounts of data. These datasets are fragmented across various ministries and underutilised in effectively informing the development of public policy and public service delivery. The 2018 – 2019 Economic Survey of India observed, "Governments already hold a rich repository of administrative, survey, institutional and transactions data about citizens, but these data are scattered across numerous government bodies. Utilising the information embedded in these distinct datasets would inter alia enable government to enhance ease of living for citizens, enable truly evidence-based policy, improve targeting in welfare schemes, uncover unmet needs, integrate

[55] Electronic Commerce in India: Draft National Policy Framework (July 27, 2018) <https://www.medianama.com/wp-content/uploads/Draft-National-E-commerce-Policy.pdf>.

fragmented markets, bring greater accountability in public services, generate greater citizen participation in governance, etc."[56]

The government should develop mechanisms to effectively harness the aggregated data available within the various government arms and agencies. It is equally important to do so in a manner that acknowledges and prevents the privacy harms that could impact any individuals whose personal data has been used in the creation of such aggregated datasets. Effective harnessing of data by the government has the potential to increase public welfare. Hence, besides developing regulatory mechanisms to access relevant public interest data from data businesses, the government should redouble its efforts in effectively utilising the data existing within the various government ministries. Additionally, the government, at all administrative levels, needs to invest in building the skill and capacity to analyse and draw relevant insights from the data available within the government.[57]

### *(c) Government Data as a Public Good*

Data gathered by the government on issues of societal interest should be made publicly available and democratised to enhance public welfare.[58] The Economic Survey 2018 -2019 has recommended that such data should be made public goods.[59] The 2012 Open Government Data initiative has been a platform through which the government has shared government datasets. [6061] However, as the

[56] Chapter 4: Data "Of the People, By the People, For the People", Economic Survey 2018 – 2019, Pg 78, <https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/vol1chapter/echap04_vol1.pdf>.

[57] ibid. p. 92.

[58] ibid. p. 83.

[59] ibid.

[60] Open Government Data Platform, https://data.gov.in

[61] Open Government Data (OGD) Platform, India, <https://www.nic.in/projects/open-government-data-ogd-platform-india/>.

Economic Survey 2018 - 2019 has noted, there is scope for much more work and data sharing by the government.[62]

Despite such open data initiatives, various government datasets are not open to the public; despite being funded by public money. Various central ministries and state departments, commission take up various data collection initiatives to undertake needs assessments or conduct impact evaluations of schemes. Many of these studies are made public; however, the underlying data is not made public.[63] Given that these studies are conducted at the instance of government organisations, they should be shared with the public. Such public access will enable analysis by various experts and organisations to draw out relevant insights that may be beneficial for public purposes.

As the government seeks to create mechanisms to access NPD from data businesses for public purposes, the government should lead the way and concurrently enable public access to all non-confidential government datasets. Such public access has the potential to generate a great deal of public, social and economic value. However, public access to government datasets has to be done in a manner that addresses the privacy risks arising from sharing NPD that is anonymised and aggregated personal data.

**(d) *Appropriateness of Data sought as NPD under Public Purpose***

Careful thought needs to be given to the kind of data being sought for public purposes under the NPD regulatory framework and the corresponding privacy implications that may arise. One of the public interest purposes for which the government is seeking access to NPD is for research and innovation. In the Committee report, one of the illustrations for utilisation of data by Indian researchers and government agencies for research purposes and the creation of

[62] Chapter 4: Data "Of the People, By the People, For the People", Economic Survey 2018 – 2019, Pg 83, <https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/vol1chapter/echap04_vol1.pdf>.

[63] ibid. p. 93.

public goods and services is an Indian genome repository.[64] Genetic data is classified as sensitive personal data under the PDPB 2019.[65] Literature has questioned whether genomic datasets can be anonymised and whether the privacy of data principals can be maintained.[66] Hence, it is unclear if data in a genome repository can be considered NPD.

## 2. Sovereign Purposes

The Report recommends government access to NPD from data businesses for sovereign purposes which include "national security, law enforcement, legal or regulatory purposes."[67] In the context of access of the government to NPD for sovereign purposes, we would like to raise the following issues for consideration:

### (a) *Need for Law Enforcement Access to Data through the NPD Regulatory Framework*

Ensuring national security and maintaining law and order are vital functions of the State. While performing these roles, it is understandable that law enforcement agencies require access to various kinds of data. Though the State's need to access data is legitimate, it needs to comply with the legal process and the judicial standards that have been developed and be balanced against an individual's right to privacy.

Data for national security purposes have been enumerated by the Committee to typically include "telecommunications metadata, geospatial or financial data,

[64] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 33.

[65] The Personal Data Protection Bill 2019, cl 2(36).

[66] C. Heeney, N. Hawkins, J. de Vries, P. Boddington, and J. Kayeb, 'Assessing the Privacy Risks of Data Sharing in Genomics, Public Health Genomics' (March 29, 2010) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872768/>; Mark Phillips, Can Genomic Data Be Anonymised?, https://www.ga4gh.org/news/can-genomic-data-be-anonymised/.

[67] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 32.

etc."[68] Currently, there are various legislative mechanisms through which the government and law enforcement agencies can access data, information and computer resources from individuals, companies and other organisations. The legal framework regulating surveillance in India and the procedure for the interception or access to information arises primarily from the Telegraph Act,, the IT Act and the Code of Criminal Procedure, 1973 ("**CrPC**").

Under the Telegraph Act and the IT Act, surveillance can take the form of not only the interception of the content of messages but also access to traffic, static and other metadata.[69] Procedural criminal laws and anti-terror laws such as the CrPC, Unlawful Activities Prevention Act, 1967, Maharashtra Control of Terrorism and Organised Crime Act, 1999 and the Gujarat Control of Terrorism and Organised Crime Act, 2019 contain provisions that allow for interception and the sharing of information with law enforcement agencies.[70] Additionally, various sectoral laws across domains such as income tax and the financial sector enable the government to access data and information.[71]

Therefore, there are existing legislative mechanisms through which law enforcement agencies can access data (both personal and non-personal data) across various domains. These surveillance enabling laws have been the subject of significant criticism for the breadth of powers and access allowed to law enforcement agencies and the lack of adequate oversight.[72] The recommendation

[68] ibid. p. 33.

[69] Telegraph Act, 1885, s. 5 and Telegraph Rules, 1999, r. 419-A,; Information Technology Act, 2000, ss. 69 and 69B, IT Act and Information Technology (Procedure and Safeguards for Interception, Monitoring and Decryption of Information) Rules, 2009.

[70] Code of Criminal Procedure, 1973, s. 91; Unlawful Activities Prevention Act, 1967, s. 46; Maharashtra Control of Terrorism and Organised Crime Act, 1999, s. 14.

[71] Income Tax Act, 1961, chp. XIII; Reserve Bank of India Act, 1934, s. 27, s. 28 and s. 35; Banking Regulations Act 1949. *See* Sunil Abraham and Elonnai Hickok, 'Government access to private-sector data in India' 2(4) International Data Privacy Law 302 (2012) <https://academic.oup.com/idpl/article/2/4/302/676984>.

[72] ML Sharma and Ors. v Union of India, W.P.(Crl.) No. 1/2019; Gautam Bhatia, 'State Surveillance and the Right to Privacy in India: A Constitutional Biography' 26 NLSI Rev. (2014) 127 <https://nlsir.com/wp-content/uploads/2020/07/Gautam-Bhatia.pdf>; Centre For Communication Governance at National Law

of the Committee to enable law enforcement agencies to access NPD through the proposed regulatory framework would significantly broaden the ambit of an already severely criticised government power.

If the approach proposed in the Report is to be followed, as a first step the various categories of NPD that law enforcement agencies need for surveillance purposes should be mapped out. This would involve specifically identifying areas where appropriate legislative mechanisms are not in place to provide access to the data required by law enforcement agencies. Only these specific categories of data should be allowed to be accessed under the NPD framework through a transparent process. Additionally, robust oversight mechanisms and legal processes to access such data should be put in place to ensure adequate safeguards over the government's access to such data.

However, the NPD regulatory framework is not the appropriate mechanism through which law enforcement agencies should be enabled to access NPD for surveillance purposes. Such an approach could result in overly broad and non-specific surveillance powers being vested in the State without adequate safeguards. This scenario could violate judicial standards and principles that have been developed around surveillance and violate the fundamental rights of Indian citizens.

To ensure that the State is not vested with overbroad surveillance powers, it would be more appropriate to amend the relevant sectoral laws. These amendments could enable law enforcement access to specific NPD categories for national security and surveillance purposes. Additionally, the amendments must provide for

University Delhi, 'Response to Call for Submissions on 'Surveillance Industry And Human Rights'' (2019) <https://ccgdelhi.org/wp-content/uploads/2019/02/CCG-NLU-Response-to-Call-for-Submissions-on-Surveillance-Industry-and-Human-Rights_February-2019.pdf>.

a transparent procedure for access to this data by the law enforcement agencies as well as clear accountability mechanisms.

In the medium term, we recommend that a robust legal framework to regulate surveillance be put in place. This framework should govern the functioning of law enforcement and intelligence agencies, their surveillance powers and the categories of information and data they can access, and the accountability measures and the legal processes they have to follow to access data. The Srikrishna Committee has stated that "we also recommend that the Central Government carefully scrutinise the question of oversight of intelligence gathering and expeditiously bring in a law to this effect. Such a law should provide for both parliamentary oversight as well as judicial approval of all requests for non-consensual access to personal data. The key rationale underlying such checks and balances is the need for ex ante access control as well as ex post accountability."[73] Putting in place such a law will allow for the effective protection of privacy and data protection principles. Broad public consultation must be undertaken with stakeholders on the appropriate manner to proceed.

**(b)** ***Privacy Implications***

Besides the above-mentioned surveillance enabling laws being severely criticised for inadequate safeguards to ensure accountability and for the lack of judicial oversight, they have also been criticised for inadequately protecting the various fundamental rights of citizens including the right to privacy.[74] In fact, the constitutionality of the surveillance powers of the State, including under the

[73] Committee of Experts under the Chairmanship of Justice B.N.Srikrishna, 'A Free and Fair Digital Economy' <https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report-comp.pdf>.

[74] Centre For Communication Governance at National Law University Delhi, 'Response to Call for Submissions on 'Surveillance Industry And Human Rights'' (2019) <https://ccgdelhi.org/wp-content/uploads/2019/02/CCG-NLU-Response-to-Call-for-Submissions-on-Surveillance-Industry-and-Human-Rights_February-2019.pdf>; Arushi Gupta, 'The Case Against the Constitutional Validity of Mass Surveillance Programmes' 3(3) Indian Law Review (2019) 225.

Telegraph Act and the IT Act, has been challenged before the Supreme Court and the case is currently pending.[75]

Over the last several decades, the Supreme Court has examined the conduct of surveillance in the context of individuals. The Court has examined a range of cases from the use of traditional surveillance measures, such as domiciliary visits, to the contemporary use of technology for surveillance.[76] The Court's jurisprudence has resulted in amendments to the Telegraph Act and Rules to specify procedural safeguards to be followed by government authorities undertaking surveillance.[77]

More recently, in *Puttaswamy,* while affirming the constitutionally guaranteed fundamental right to privacy, the Court referred to the recent changes in the means of surveillance as a result of the development of new technologies.[78] The Court acknowledged that while the State may have legitimate national security and other interests to monitor communications and collect and process personal data, such actions should be taken in accordance with the Constitution.[79] Hence, any government encroachment on the privacy of an individual must be subject to law, it must be undertaken in pursuance of a legitimate state aim, and the means adopted for such action must be proportional to the objects and needs sought to be fulfilled by the law.[80] Last year, the Bombay High Court outlined the sphere of the State's power to undertake surveillance on an individual and applied the tests

[75] *ML Sharma and Ors. v Union of India*, W.P.(Crl.) No. 1/2019.

[76] *See* Centre for Communication Governance at National Law University Delhi 'Privacy and the Supreme Court' (1st ed., NLU Delhi Press, 2020.)

[77] *People's Union of Civil Liberties v Union of India and Anr*, (1997) 1 SCC 301.

[78] *K. S. Puttaswamy v Union of India* (2017) 10 SCC 1 , para 328 (Dr. D.Y. Chandrachud, J.).

[79] Centre For Communication Governance at National Law University Delhi, 'Response to Call for Submissions on 'Surveillance Industry And Human Rights'' (2019) <https://ccgdelhi.org/wp-content/uploads/2019/02/CCG-NLU-Response-to-Call-for-Submissions-on-Surveillance-Industry-and-Human-Rights_February-2019.pdf>.

[80] *K. S. Puttaswamy v Union of India* (2017) 10 SCC 1, para 180 (Dr. D.Y. Chandrachud, J.); *See also* Vrinda Bhandari and Karan Lahiri, 'The Surveillance State: Privacy and Criminal Investigation in India: Possible Futures in a Post-Puttaswamy World' 3(2) Univ. of Oxford Human Rights Hub Journal 15 (2020) <http://ohrh.law.ox.ac.uk/wordpress/wp-content/uploads/2020/05/U-of-OxHRH-J-The-Surveillance-State-Privacy-and-Criminal-Investigation-1.pdf>.

of legitimacy and proportionality laid down in *Puttaswamy*, to the interception orders issued under the Telegraph Act, and held that in the particular case the order for interception could not be substantiated in the interest of public safety as it did not satisfy the test of "principles of proportionality and legitimacy".[81] Though this privacy jurisprudence has been developed in the context of personal data, it has relevance in a discussion around the sharing of NPD (anonymised personal data) with law enforcement agencies, given the fluidity between personal data and NPD (anonymised personal data).

Multiple privacy concerns arise concerning the reidentification of personal data from NPD that is shared with law enforcement agencies:

**(i) Reidentification of Anonymised Data**

As the Committee themselves repeatedly acknowledges in the Report, anonymisation of personal data is not absolute and is by no means irreversible.[82] Anonymised data can be reidentified into the constituent personal data.[83]

**(ii) Cross-referencing of Databases**

Data is more useful in the insights it yields when it is combined with other data.[84] Hence, law enforcement agencies are likely to cross-reference NPD data shared with them with other databases to garner richer

[81] *Vinit Kumar v Central Bureau of Investigation*, 2019 SCC OnLine Bom 3155; *see also* Sangh Rakshita and Nidhi Singh, 'Right to Privacy: The Puttaswamy Effect' (*CCG-NLU*, Jan. 22, 2020). <https://ccgnludelhi.wordpress.com/2020/01/22/the-right-to-privacy-the-puttaswamy-effect/>.

[82] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 16.

[83] UK Information Commissioner's Office, *Anonymisation: Managing Data Protection Risk Code of Practice* (2012) <https://eprints.soton.ac.uk/345821/1/anonymisation_code.pdf>.

[84] Chapter 4: Data "Of the People, By the People, For the People", Economic Survey 2018 – 2019, Pg 78, <https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/vol1chapter/echap04_vol1.pdf>

insights. Cross-referencing multiple databases increases the possibility of the reidentification of personal data.[85]

Thus, there is an element of fluidity between NPD and personal data. Once the government has access to NPD, there are limited mechanisms to check how the data is being used. The State can potentially deanonymise NPD that have been accessed under these surveillance powers and identify individuals, potentially impinging their right to privacy.

This raises serious concerns about the privacy of individual's personal data against government surveillance. Assuming, the NPD framework is enacted in its proposed form and continues to include this mechanism to access data for sovereign purposes, adequate statutory mechanisms will have to be developed to ensure that the route of access to data suggested by the Committee for law enforcement agencies through the NPD regulatory framework does not dilute the already limited protections in place. These statutory mechanisms will include:

(i) **Scope of the Surveillance Power**

Assuming, the proposed NPD framework is enacted, to protect against government abuse of their surveillance powers, it will be imperative to specify the extent of government access to NPD and the kind of data processing that can be undertaken, drawing on the judicially developed principles of proportionality and necessity. Besides this, restrictions may need to be put in place to limit the deanonymisation of NPD into personal data.

(ii) **Oversight Mechanisms**

[85] *See* Boris Lubarsky, 'Technology Explainers: Re-Identification of "Anonymized" Data' 1 Georgetown. Law Tech. Rev. 202 (2017) <https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/>.

If the NPD framework is enacted in its proposed form, adequate judicial[86] or other independent oversight mechanisms will have to be put in place to oversee the exercise of surveillance powers to access NPD for sovereign purposes. These oversight mechanisms will have to ensure compliance with the tests of necessity and proportionality as laid down by the Supreme Court.[87]

(iii) **Compliance with Regulatory Framework**

Under the PDPB 2019, government agencies can be exempted from complying with the provisions of the law with respect to the processing of personal data.[88] This exemption has been widely criticised. It is argued that even when authorised to access personal data, law enforcement agencies should comply with the data protection principles such as data retention and purpose limitation to the extent possible. Additionally, as per principles of natural justice, individuals should have a right to be provided post facto notice, and an opportunity to appeal being put under surveillance and the related impingement of their fundamental rights. Similarly, in case of surveillance undertaken under the proposed NPD framework, law enforcement should to the extent possible comply with any obligations under the NPD framework that other entities have to adhere to with respect to the protection of the data principal.

**(c) *Potential Misuse of NPD accessed under Public Interest Purposes for Sovereign Purposes***

Mechanisms need to be put in place to prevent the potential misuse of NPD accessed for public interest purposes for sovereign purposes, specifically for national security purposes. There may be instances where the government and its

[86] *K.S Puttaswamy v Union of India*, 2019 (1) SCC 1, paras 409 and 513.6.

[87] *K.S. Puttaswamy v. Union of India,* (2017) 10 SCC 1.

[88] The Personal Data Protection Bill 2019, cl 35.

agencies seek to access NPD through the public interest purpose mechanism to circumvent the oversight and accountability mechanism in place for access to NPD for surveillance purposes or later divert NPD previously accessed under the public purpose mechanism for use by government agencies for surveillance purposes.

Frameworks and principles will have to be evolved to ensure clarity on the categories of data that can be accessed by the government under the public interest mechanism and prevent the diversion of NPD accessed for public interest purposes for law enforcement purposes. This will ensure that the government does not misuse the potentially less regulated public purpose process to access NPD for surveillance purposes.

### 3. Economic Purposes

The Report states that data can be requested by startups, businesses, data trustees and governments from data businesses and data businesses may in certain circumstances need to either mandatorily share the data for no compensation, fair monetary remuneration or market compensation.[89] The objective of enabling such access to data is with a view to encourage competition and provide a level playing field, and encourage innovation through startup activities.[90] Hence, the Committee is setting up this framework as a way of correcting the imbalance in the market created by the emergence of a few dominant players.[91]

While we welcome the Committee's recognition of the need to rectify the prevalent imbalance in bargaining and market power created by platforms and data businesses, the mechanism of mandatory access to the NPD of data businesses, as suggested by the Committee, may be an inappropriate mechanism to address these concerns. The Report

[89] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 34-35, para 7.3.

[90] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 32.

[91] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 8.

argues that one of the key drivers of the value of these digital and data businesses is their ability to collect and analyse data of users, which in turn leads to network effects that help them grow and become dominant players in the market.[92] A pivotal argument that the Committee seems to be making is that granting startups with access to data will level the playing field. However, the Committee has not relied on any evidence to support this proposal. It may be argued that such a proposal may have a negative impact on innovation.

Literature suggests that data is one of the several aspects that influence a company's ability to succeed.[93] Consequently, the question remains whether the proposed NPD framework enabling mandatory access to data is the most effective mechanism to address this challenge. As discussed in Section I.A, an appropriate response may be a graded approach to regulation across various domains such as reform of competition law, platform neutrality and platform interoperability.[94] Broad public consultations need to be done by the Committee, with a spectrum of data businesses and other stakeholders, on whether access to data is a challenge that startups and other data businesses are facing and whether the proposal of the Committee is the appropriate response to address the prevalent market imbalances.

Below we discuss the mechanisms suggested by the Committee to access NPD from data businesses and the concerns that may arise:

***(a) Requests for Access to Raw Data***

Data businesses have to share the metadata about the various data they have collected. When a relevant stakeholder identifies a specific dataset they want access to, they may request the data custodian for the detailed underlying raw

[92] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 8.

[93] Urvashi Aneja and Angelina Chamuah, 'A Balancing Act, The Promise & Peril of Big Tech in India', Tandem Research (2020) <https://tandemresearch.org/assets/Tandem-Research-Big_Tech_report.pdf>.

[94] ibid.

data.[95] The Report states that “The principle of raw data is standards compliant, machine readable and fidelity as collected. The raw data will be made available in usable formats, and only an open, reviewed licence-free standard can be used.”[96] If the data custodian services this request for data, then the process is complete. However, if the data custodian refuses to share access to the data, then the relevant stakeholder can make a request to the NPDA.[97]

The NPDA will evaluate the request from the perspective of social, public and economic benefit.[98] If the NPDA finds the request to be genuine and that the sharing of the data can result in these benefits, then the NPDA can mandate the data custodian to share the factual or raw data.[99] In such a scenario, the Committee envisages creating a shared public database so that the data can be accessed by all to spur innovation.[100]

In the Committee's description of the process, there is no mention of compensation for access to this NPD. Hence, it can be inferred that the expectation is that the data business will provide this data to the startup or business for free. Later in the Report, it is stated that “Only the raw/factual data pertaining to community data that is collected by a private organization need to be shared, subject to well-defined grounds at no remuneration.”[101]

This process of data access laid out above raises two concerns:

**(i) Genuineness of Request**

[95] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 37.

[96] ibid p. 25.

[97] ibid p. 37.

[98] ibid p. 38.

[99] ibid p. 38.

[100] ibid p. 35, 38.

[101] ibid. p. 37, para 7.4.

It is unclear what specific principles will guide the NPDA's assessment on the genuineness of a request for data. Suppose such a mechanism is adopted in the NPD regulatory framework, detailed principles will need to be specified in the statute for the NPDA to rely on when making such an assessment.

**(ii) Incentives for Innovation**

While the Committee recommends such mandatory access from the perspective of spurring innovation, it fails to consider the potential of the impact of such a proposal on market incentives for innovation and data collection. The Report acknowledges that while ensuring markets function properly, adequate incentives for new business creation must be safeguarded.[102] The proposed framework of mandatory data sharing may create disincentives for data businesses to enter the market and invest in growth, which could in turn, hamper the effective development of the market.

Additionally, the Report recognises that symmetric data sharing obligations on all data businesses may not always work for small businesses, and may even be detrimental to them.[103] Hence, the Report envisages that the framework law will have to have provisions on threshold size for data sharing and potentially graduated sharing obligations.[104] This may disincentivise business expansion and encourage businesses to stay below certain data thresholds.

Companies invest significant human, conceptual and financial capital to collect data and build entire businesses around them. Inherent in these datasets is a company's ability to innovate, which may be constrained if they have to share the

[102] ibid. p. 9.

[103] ibid. p. 20.

[104] ibid.

datasets with other companies mandatorily. Requiring mandatory sharing could dampen businesses' incentive to invest in the collection, structuring and maintenance of these datasets. The question that then arises, in this context, is that given the potential costs to innovation, is this the most effective way to create a level playing field? Would the costs to innovation by this mechanism be outweighed by the economic potential for growth that could be unlocked for other companies/communities, as argued by the Committee?

Besides this, the Committee's proposal may limit the ability of data businesses to realise the value of their data. In the Report, the Committee iterates three mechanisms by which organisations realise the value of their data: direct monetisation, internal investments and mergers and acquisitions (M&A).[105] It has been argued by the industry that the regulatory mechanism being proposed by the Committee may negatively impact a startup or data businesses' ability to attract investments.[106] Investor's may shy away from investing in startups or data businesses who may at any time be required to share their data with a competitor and thus, potentially lose their edge in a highly competitive market. This may further disincentivise innovation.

The Committee needs to undertake broad public consultations, across a spectrum of data businesses and other relevant stakeholders, to identify the primary bottlenecks and challenges faced by startups and data businesses. If the sector is examined holistically, what all needs to be done to ensure the growth and development of the digital economy? Such a consultation will enable the Committee to understand whether the startup sector needs such a regulatory intervention and whether the market-wide benefits of such a proposal would outweigh the above discussed potential costs to innovation.

[105] ibid. p. 6.

[106] BSA, the Software Alliance, 'Submission on the Report by the Committee of Experts on Non-Personal Data Governance Framework' (Sept. 10, 2020) <https://www.bsa.org/files/policy-filings/09102020indiabsanpd.pdf>.

### *(b) Access to Value Added Data*

The Committee has discussed mechanisms for access to three levels of value-added data and the corresponding remuneration that may be provided to a data business.

*Non-trivial value-add:* The Committee states that where the value-add to the data is non-trivial with respect to "the value or collective contribution of the original community data and collective community resources used, (or otherwise for reasons of over- riding public interest) data sharing may still be mandated but on FRAND (fair, reasonable and non-discriminatory) based remuneration."[107] Hence, even where non-trivial value-add has taken place, data businesses can be mandatorily required to share the data. However, in these cases, they will be compensated with FRAND based remuneration.

*Increasing value add:* With increasing value add the Committee has noted that "it may just be required that the concerned data is brought to a well-regulated data market and price be allowed to be determined by market forces, within general frameworks of openness, fairness etc."[108] It is unclear whether in this situation, the data business can be mandated to share the data through a market mechanism or whether the data business can decide what it wants to do.

*High level of value add:* In the scenario of high level of value-add the Committee has stated that it would largely be left to the data business to decide how it wishes to use the data.[109] It is relevant to note here that, in this context, the Committee observed that economic privileges are now appropriately inherent to the business.[110]

---

[107] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 37, para 7.4.

[108] ibid.

[109] ibid.

[110] ibid.

The following issues arise for the consideration of the Committee:

**(i) Ascribing Value to Data**

The Report acknowledges that it is difficult to place value on data. There are several challenges with ascribing a price or value to data.[111] This is partly driven by the fact that data can be easily duplicated and sold multiple times and can generally be considered a non-rival good (though with several exceptions).[112] The value of a dataset differs from one organisation to another, and hence the amount each organisation may be willing to pay would also vary.[113] Thus, setting one price for access by all organisations may result in the price in some instances being either too high or too low. Hence, it may be more appropriate to determine the price or value of data given the various contextual factors in a transaction rather than determining a fixed price or value for all the potential parties interested in accessing the data.

Does this imply that a FRAND remuneration determination by the NPDA should be variable depending on the organisation seeking to access the data? While such an approach may work in a market mechanism, it may be challenging to implement in a regulatory context. Such a variable pricing model may not be appropriate in the situation of the NPDA fixing a FRAND remuneration, nor may the NPDA have the necessary technical expertise to make such a valuation.

Besides this, the second and third category of value-added data would likely be shared through market mechanisms. The Committee alludes to the setting up of data markets. Some private data marketplaces have been set up across the world,

[111] J. Heckman and others 'A Pricing Model for Data Markets' iConference 2015 Proceedings (2015) <https://www.ideals.illinois.edu/bitstream/handle/2142/73449/207_ready.pdf?sequence=2>; Avanish Kushal, Sharmadha Moorthy and Vikash Kumar 'Pricing for Data Markets' <https://courses.cs.washington.edu/courses/cse544/11wi/projects/kumar_kushal_moorthy.pdf>.

[112] Erwin Kuhn, 'Let's Talk About Data Pricing — Part I' (*Medium,* August 13, 2019) <https://blog.oceanprotocol.com/lets-talk-about-data-pricing-part-i-bbc9cf781d9f>.

[113] Everett Leonidas and Whit Walker, 'A Practical Guide to Pricing Data Products, DatastreamX' <https://cdn2.hubspot.net/hubfs/573334/Downloadable_Content_(WP_or_Guides)/DataStreamX_Data_Product_Pricing_Whitepaper.pdf>.

however, their functioning has been impacted by challenges of seller paralysis and by a lack of market structure.[114]

Both the mechanisms for valuing data and setting up of data marketplaces are in iterative stages globally. Public consultation would have to be undertaken by the Committee with stakeholders to determine the appropriate approach to ascertaining the value of each of the three categories of value-added data as well as the setting up of data markets and exchanges.

**(ii) Advantage to Larger Market Players**

The FRAND and market mechanisms for accessing value-added data may end up disproportionately benefiting larger market players. Such players have the capital to spend on acquiring data from a large variety of market players, in comparison to the start ups, who have limited capital at their disposal. The more that is charged for the data, a narrower set of market players may be able to afford access to the data. On the other hand, if the data is priced more accessibly, it may allow larger market players to buy up vast volumes of data. This could negatively impact the ability of startups to compete in the market effectively.

Public consultation would have to be undertaken by the Committee with stakeholders across the spectrum from larger companies to startups to understand the potential implications of adopting the mechanisms recommended by the Committee.

## B. Categorisation of Non-Personal Data

The first recommendation the committee has made is with regard to the definition of NPD. The committee suggests that a general definition of NPD can be adopted. According to this definition, NPD would broadly include two sets of data:

[114] Erwin Kuhn, 'Let's Talk About Data Pricing — Part I' (*Medium,* August 13, 2019) <https://blog.oceanprotocol.com/lets-talk-about-data-pricing-part-i-bbc9cf781d9f>.

- Data that never related to an identified or identifiable natural person, such as data on weather conditions, data from sensors installed on industrial machines, data from public infrastructures, and so on.

- Data which were initially personal data, but were later made anonymous. Data which are aggregated and to which certain data transformation techniques are applied, to the extent that individual specific events are no longer identifiable, can be qualified as anonymous data.

Beyond this general definition, the Committee has defined three categories of Non-Personal Data – 1) Public Non-Personal Data 2) Community Non-Personal Data & 3) Private Non-Personal Data.

The Committee has also defined a new concept of 'sensitivity of Non-Personal Data', stating that even Non-Personal Data could be sensitive from the following perspectives – 1) It relates to national security or strategic interests; 2) It is business sensitive or confidential information; 3) It is anonymised data, that bears a risk of re-identification

This section of the comments will primarily address the classification of NPD described above, and the concerns that arise from such a categorisation.

The issues of sensitivity, and related issues of consent and anonymisation of NPD derived from personal data are addressed in Part III of this document.

### 1. Public Non-Personal Data

Public NPD includes any NPD "collected or generated by the governments, or by any agency of the governments, and includes data collected or generated in the course of execution of all publicly funded works". An exception has been made for data that is required to be confidential under law.[115]

[115] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.2.

The Report also provides two examples of such NPD – (a) anonymised data of land records, public health information, vehicle registration data etc., and (b) data regarding pollution levels in the city collected by a university, based on a publicly funded project.

The objective of categorizing NPD in this manner seems to be two-fold - first, to help guide means of data sharing (whether mandatory or voluntary), and second, to understand and protect the underlying rights and interests of the subjects from whom the data originates.

On the first, the Report rightly suggests that the Government should improve on existing Open Government Data initiatives[116]. The Open Government Data initiative, and specifically the National Data Sharing and Accessibility Policy, have been in place since 2012. However, there has been some criticism around the effective implementation of these policies in a manner that increases transparency and informs citizens and different stakeholders better.[117]

In terms of the rights and interests of data principals, the framework for regulation of NPD will need to be drafted in keeping with privacy and data protection principles, and particularly the PDPB 2019. Barring some exceptions, the PDPB 2019 applies to all government collection and processing of personal data. Limitations on such collection and processing, including limits on the purpose for which data can be processed, the time such personal data can be retained will be applicable to the different government agencies that undertake collection and processing of personal data. This personal data will often be the basis on which public NPD is generated.

The examples of public NPD provided in the Report itself indicate that public NPD could comprise of NPD that is generated from personal data, therefore creating a significant overlap between public NPD and community / private NPD (described below).

[116] ibid para 7.4. i.

[117] Srinivas Kodali, 'Not Open or Accountable: The Government Open Data Use License Is Flawed' (*The Wire*, Jan. 25, 2017) <https://thewire.in/business/open-data-license-government>.

More clarity is required from both the personal and non-personal data frameworks on how government agencies will need to implement these regulations. There is also a need to develop capacity in these agencies accordingly.

Further, specifically on public NPD, the Report suggests that since public NPD is derived from public efforts, the datasets created can be treated as a 'national resource'. In light of the overlap between public NPD and other types of NPD, this creates concerns regarding the manner in which underlying rights and interests of the data principals from whom the data has been obtained, especially in the case of NPD derived from personal data.

The Report also recognises that data is non-rivalrous, and therefore different from natural resources, since multiple people can use and benefit from data at the same time. The Report states that value of the data can be consumed by the relevant community as well as third parties, without degrading the value of such data to the relevant community. The Report however, does not go into a detailed examination of how the concept of 'national resource' has been treated under Indian law, and the pre-requisities applicable when government agencies acquire and use such resources. We recommend that a detailed study of the legal and other implications of such an approach is undertaken before formalising the same in regulation. It may also be useful to undertake a comparative study to examine the terminology and concepts used by other jurisdictions in this regard. For instance, the EU Strategy for Data uses the term "publicly-held data" and does not go into issues of ownership of data.

As discussed in Part I.A above, (personal) data is not property and can be understood differently in different contexts. The Committee's approach seems to limit the value of data to economically measurable value only whereas the data principals in each given case might value different aspects of their data over its economic value. For instance, the Committee's approach may not account for situations where the individual members of the community value (information) privacy over the potential economic benefits of the use of their data by a third party, whether the State or a private actor. It would be difficult to anticipate and identify the different types of values that one may find important in a

particular set of data, about themselves or a community they belong to. More so, given the concerns that have been discussed in Part II.B.3 and II.C of these comments regarding the identification of communities, and the role of data custodians and data trustees.

While the Report does provide that the concept of 'beneficial ownership' should be used to allocate primary economic and other statutory rights over data, it is not clear how these different rights will be balanced in operation, given the number of stakeholders with interests, and the difficulty in identifying the source of NPD in many instance.

### 2. Private Non-Personal Data

Private NPD is defined as any NPD that is "collected or produced by persons or entities other than government, the source or subject of which relates to assets and processes that are privately-owned by such person or entity, and includes those aspects of derived and observed data that result from private effort".[118]

This definition is vague, and does not explain the concepts of 'source and subject', or 'privately-owned assets and processes'. For instance, where the NPD is derived from personal data of a data principal, or community data, it is not clear how ownership of any aspects of such a source or subject can be attributed to the private entity.

As with public NPD, there is also significant potential for overlap between private NPD and community NPD. In this regard the Report simply states that community NPD shall not include any private NPD, leaving many questions about how the community's rights will be identified and exercised unanswered. These issues have been discussed in the section on community NPD below, and in the respective sections on different stakeholders such as data trusts, data principals and data trustees.

[118] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.4.

### 3. Community Non-Personal Data

The third category of NPD, borrows from the report that the Justice Sri Krishna Committee presented along with the draft Personal Data Protection Bill, 2018[119]. In that report, the committee of experts recognised that there is a need for collective protection of the privacy of an identifiable community. Taking these concepts of community data and collective privacy forward, the Report introduces the concept of community non-personal data, i.e. NPD whose source or subject pertains to a community of natural persons. The Report states that this community NPD could include anonymised personal data, and NPD about inanimate and animate things or phenomena. A community is defined as any group of people that are bound by common interests and purposes, and involved in social and/or economic interactions. The Report states that such a community could be a geographic community, a community by life, livelihood, economic interactions or other social interests and objectives, and/or an entirely virtual community. The Report also states that raw/factual data collected by telecom, e-commerce, ride-hailing companies would be treated as community non-personal data.

Some of the concerns in the Report's definition of community pNPD and its use are discussed below.

#### *(i) Defining a Community*

It is clear from existing scholarship that with the rise of big data analytics, aggregated data of various individuals can provide insights into the behaviour of a group or a community with relative ease.[120] Individual habits and characteristics may be taken to represent a class of similar individuals and once this information has been converted into datasets, common traits may emerge between individuals who have no awareness of

[119] Committee of Experts under the Chairmanship of Justice B.N. Srikrishna, 'A Free and Fair Digital Economy, Protecting Privacy , Empowering Indians' ("Justice Srikrishna Committee Report") <https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf>.

[120] Lanah Kammourieh and others, 'Group Privacy in the Age of Big Data' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 3.

being bound by such similarities.[121] Even when such data has been anonymised, risks of profiling, pattern recognition, and activity tracking may infringe the privacy rights of groups or communities.[122] Policies and decisions are made on the basis of such inferences and patterns, which may negatively impact groups or communities as a whole.[123]

The pertinent question to answer even before we prescribe rights to these groups or communities is what constitutes a group or a community? The Report provides a very broad definition of 'community'. It states that to form a community a group of people must be involved in social and/or economic interactions and must be bound by common interests. As discussed above, it classifies raw or factual data collected by tech companies like e-commerce or ride-hailing companies or such data collected by telecom companies as community data. With examples of such broad communities, the Report does not clarify how all users of a telecom or an e-commerce company will be bound by 'common interests' or address the potential for conflict of interest within such disparate communities.

The Report only states that for community non-personal data, the data principal responsible for exercising key rights will be the source/subject community from which such data has been collected.[124]

### *(ii) Algorithmically created Groups or Communities*

Beyond the consideration of certain identifiable communities like users of telecom or e-commerce services, or a group of people affected by a particular disease for example diabetes, the Report does not comprehensively engage with the concept of collective or group privacy from the lens of big data analytics.

---

[121] ibid.

[122] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 1.

[123] ibid.

[124] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.7. iv.

Data processors who make use of advanced data analytics do not segregate groups on the basis of traditional notions like shared perception or pre-existing commonalities between groups of people.[125] They classify groups based on the use of algorithms post collection of their data. Patterns in raw data only emerge once algorithms have sifted through them and segregated information based on which community or group needs to be targeted.[126] For example, e-commerce companies may divide users into different groups and personas based on various metrics such as - age, gender, location, browsing history etc. to determine whether they are interested in shoes or home appliances, and target them accordingly.

The challenge is that such practices of data analytics make the grounds on which groups are identified increasingly imperceptible.[127] Individuals comprising these groups or communities may never gain the knowledge that they had been classified together based on similar characteristics.[128] In fact, often data analysts themselves might not be aware of why certain groups are classified as they are.[129]

Martin Tisné summarises this conundrum succinctly by stating that the challenge arising from problems of data privacy and discrimination is that the public does not know which group they are part of as only the alogrithm has any knowledge of that. He says that, '*solutions lie in hard accountability, strong regulatory oversight of data-driven decision making, and the ability to audit and inspect the decisions and impacts of algorithms on society.'*[130]

This presents us with a unique challenge, wherein on one hand broad definitions of a community may not be workable due to the conflicts which may arise within such a

[125] Lanah Kammourieh and others, 'Group Privacy in the Age of Big Data' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 3.

[126] ibid.

[127] ibid.

[128] ibid.

[129] ibid.

[130] Martin Tisné, 'The Data Delusion: Protecting Individual Data Isn't Enough When the Harm Is Collective' (2020) Stanford Cyber Policy Center <https://cyber.fsi.stanford.edu/publication/data-delusion>.

community. And on the other hand, due to advancements on data analytics, it is practically impossible to identify groups which may warrant protection before their data is processed.

To address this challenge, we recommend that the government, while drawing up the data governance framework for non-personal data take a holistic view of collective privacy and lay down a mechanism so that all Indians can protect themselves against collective privacy harms. The definition of a group or community should consider how these collectives get profiled and targeted using modern technology and the practical mechanics of how they might exercise their collective rights over their non-personal data.

We believe that the concept of data trusts as put forward by the Open Data Institute[131] and scholars such as Aline Blankertz,[132] Anouk Ruhaak,[133] and Sylvie Delacroix and Neil D. Lawrence[134] might help in addressing some of these challenges arising out of regulating group or community non-personal data.

### C. Institutional Structure and the Roles Envisaged under the NPD Framework

#### 1. Data Trustees

The Report proposes new stakeholders in the form of data trustees who will exercise data rights on behalf of specific communities. However, the Report leaves out the basis and principles for who could constitute an appropriate data trustee to future legislation on non-personal data. The Report only provides that the data trustee must be the closest and most appropriate representative body for the concerned community and should utilise community non-personal data in the 'best interest' of the community. It envisages relevant

[131] Jack Hardinges, 'Data Trusts in 2020' (*Open Data Institute*, March 17, 2020) <https://theodi.org/article/data-trusts-in-2020/>.

[132] Aline Blankertz, 'Designing Data Trusts: Why We Need to Test Consumer Data Trusts Now' (2020) Stiftung Neue Verantwortung <https://www.stiftung-nv.de/en/publication/designing-data-trusts-why-we-need-test-consumer-data-trusts-now>.

[133] Anouk Ruhaak, 'Data Trusts: Why, What and How' (Nov. 12, 2019) <https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34>.

[134] Sylvie Delacroix and Neil D. Lawrence, 'Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance' (2019) Vol 9, No 4 International Data Privacy Law <https://academic.oup.com/idpl/article/9/4/236/5579842>.

government departments as data trustees for certain communities. For example, the Report proposes that the State Government of Manipur can be a trustee on data on Meitei language.

As the source/ subject community is the data principal for community non-personal data and they exercise key rights over such data,[135] it becomes essential to nominate a party to exercise data rights on behalf of the entire community. The Report suggests data trustees as such parties/ agents for exercising data rights on behalf of the entire community. However, it does not lay down the legal relationship or the technical architecture for such data trustees to perform their duties. The Report does not lay down what it means by 'closest and most appropriate representative body' or phrases like 'best interest of the community'.

We recommend that the proposed framework provide details on the principles on which data trustees may be appointed, their legal relationship with communities, and information on the technical architecture required for them to exercise data rights on behalf of communities.

***Government Entities as Data Trustees***

The Report suggests that government entities be assigned as the data trustees for certain community data. An example of this is the Directorate of Urban Land Transport acting as the data trustee for all traffic data collected by multiple ride-sharing platforms.

The Report also recommends mandatory data sharing for sovereign, public interest, and economic purposes. It proposes controlled access to public, community, and private non-personal data to individuals and organisations for defined purposes and with appropriate safeguards. In seeking access to and sharing of community non-personal data, the report suggests that data trustees work alongside the Non-Personal Data Regulatory Authority ("**NPDA**") for enforcement and enablement.

[135] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.7. iv.

Questions around the independence of various stakeholders within the proposed non-personal data governance framework arise if the government acts as the data trustee, custodian of public non-personal data, and the regulator in the form of the NPDA.

For example, if the government wants to collect certain non-personal data related to the spread of AIDS disease in India from a particular community, and the data trustee for such community is the Ministry of Health and Family Welfare, such a community will have no choice or control over the transaction. They will mandatorily have to share such non-personal data with the government. Since the regulator in this case, the NPDA, will also be appointed by the government, there will be no recourse left for such a community. By having a government entity as a data trustee, the source/subject community in this case effectively loses rights over its non-personal data as envisaged by the Report.[136]

These concerns are exacerbated by the fact that privacy harms like profiling and behavioural tracking may result even from the processing of non-personal data or data that has been anonymised.[137]

The Report recommends that important community data for different sectors, which is pre-identified by data trustees or the governments, may be directly sought by such data trustees or governments from private players for mandatory sharing. Such community data may be placed in appropriate data infrastructures or data trusts and made available to all relevant parties.[138] In such cases where the government is the data trustee and the sectoral regulator, questions around whether it will be in a position to independently act in the best interest of communities arise. This may also hamper the data rights enjoyed by communities as data principals of their non-personal data.[139]

---

[136] ibid.

[137] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 1.

[138] *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) paras 4.9. iii, 7.3. ii (n 135).

[139] ibid para 4.7. iv.

For example, if the government decides that certain community data must be demanded from private players to open up market competition, it may not be in the appropriate position to determine whether such data sharing may negatively impact the collective privacy rights of such communities.

Additionally, as discussed above, mandatory requirements to share community non-personal data, either for public interest purposes or economic purposes might result in the violation of the right to self-determination (and consequently the right to privacy) as held by the Supreme Court's judgement in *Puttaswamy*.[140] For mandatory sharing of community non-personal data for sovereign purposes like national security, the non-personal data governance framework will need to fulfil the four-part test for restricting privacy rights as propounded in *Puttaswamy*.[141]

Another role of the data trustees as proposed by the NPD Report is the enforcement of soft obligations (via the NPDA) on data custodians like transparency, reporting mechanisms, and regulation of data practices while sharing community non-personal data. Concerns regarding the definition of soft obligations have been discussed in Section II.C.6 of this document. In addition, questions of independence lso arise, especially in situations where the government acts as both data custodian and data trustee. The Report fails to account for such possibilities and provide for alternatives to avoid conflicts of interest and the resultant dilution of standards of independence.

While protection and safeguarding of community data rights will require some level of representation, the Report does not consider alternatives or explain how this model of data trustees will be best suited for such purpose. Alternate models in the form of data

[140] (2017) 10 SCC 1.

[141] ibid.

cooperatives,[142] data collaboratives,[143] and research partnerships[144] could be adopted to address the challenge of collective privacy and community non-personal data.[145] We recommend that the government explore various models of community representation before formal adoption. Running pilot tests to evaluate the usefulness of various models may also help in arriving at the most appropriate solution.

We recommend that communities or groups must have complete autonomy to decide who they nominate as their data trustees or stewards for exercising their data rights on their behalf. We recommend that government agencies should not take up the role of data trustees in the proposed framework, as they are already data custodians and also the regulators in the form of the NPDA.

### 2. Data Principal

In the context of NPD, the Committee has explained that a data principal can be an individual, company or community.[146] The Report states that in the case of NPD that is derived from the personal data of an individual, the data principal for the personal data will continue to be the data principal for the NPD.[147] With regard to a data principal, the following issues should be considered:

#### *(a) Challenges with Adopting the Consent Model*

[142] A mutual organisation owned and democratically controlled by its members, who delegate control over their data.

[143] A new collaborative approach to data sharing where private companies, research institutions, and government agencies come together to share data for solving public problems. Stefaan Verhulst, Andrew Young, and Prianka Srinivasan, 'An Introduction to Data Collaboratives' (GovLab) <https://datacollaboratives.org/introduction.html#section1>.

[144] Groups or communities collectively deciding to share their data with research institutions or organisations.

[145] Open Data Institute, 'Data trusts: lessons from three pilots' (2019) <https://theodi.org/?post_type=article&p=7888>.

[146] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 19, para 4.7.

[147] Ibid. p. 23.

The Committee has concluded that the consent provided by the data principal for the collection, processing and use of personal data cannot automatically be applied to non-personal data. Therefore, the Committee has suggested that the data principal should provide consent for the anonymisation of their data and the use of such anonymised data at the stage of providing consent for the collection and use of their personal data.[148] This gives rise to two issues:

**(i) Difficulty in Withdrawing Consent**

The Committee itself alludes to the need for consent to be 'capable of being withdrawn' as per the PDPB 2019.[149] The notion of consent, in the context of data protection, has evolved to include the ability to withdraw consent with respect to the further use and processing of data.[150] Such withdrawal of consent has to be followed by the deletion of the personal data. However, once personal data is anonymised and integrated into non-personal datasets, it is potentially challenging for the data to be removed from the dataset, if the data principal later revokes their consent.[151]

**(ii) Non-compliance with the Requirement for the Consent to be Specific**

The Committee refers to the need for consent to be 'specific' as per the PDPB 2019.[152] However, typically, companies using NPD for analytics and AI systems do not necessarily anticipate the various uses for which the NPD could be

[148] Ibid. p. 14

[149] The Personal Data Protection Bill 2019, cl 11(2)(e); Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 17.

[150] Eugenia Politou, Efthimios Alepis and Constantinos Patsakis, 'Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions' 4(1) Journal of Cybersecurity (2018) 1-20 <https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056>.

[151] Ibid. pp 7-9.

[152] The Personal Data Protection Bill 2019, cl 11(2)(c); Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 17.

relevant.[153] Hence, it is unclear how 'specific' consent can be sought from the data principal at the time of collecting their personal data for the purposes of the anonymisation of their data and the use of such anonymised data.

The challenges to applying the model of consent to NPD are discussed further in Part III.C of this document.

***(b) Anonymised Data and the Associated Risk of Harm***

As noted by the Committee, it is difficult to anonymise data permanently. Additionally, in the context of the NPD regulatory regime, the risks of reidentification of personal data is heightened, as the regime contemplates the widespread sharing and combining of datasets. Such reidentification of the personal data of the data principal negatively impacts their privacy, and opens them up to various harms. The Committee allows the data principal to act upon and seek to address the harm that may arise as a result of the reidentification of personal data.

Some of the challenges relating to anonymisation of data and the risk of re-identification are discussed further in Part III.A of this document.

Detailed public consultation needs to be undertaken with the relevant stakeholders on regulatory and technical strategies to mitigate the risks of such harm.

***(c) Need for Clarity around the Notions of 'Best Interest' and 'Duty of Care'***

The Report enjoins the data custodian to act in the best interest of the data principal.[154] The Report further goes on to state that "Such community 'best interest' will need to be channelled or communicated to data custodians by data

[153] Eugenia Politou, Efthimios Alepis and Constantinos Patsakis, 'Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions' 4(1) Journal of Cybersecurity (2018) 1, 4 <https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056>.

[154] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 19, para 4.8.

trustees on behalf of the data principal community. It could be in the form of data advice, recommended data practices requirements/guidelines, etc. but must always meet the cannons of the best interests of the data subject community or the data principals."[155] Besides this, the data trustee has an obligation to act in the best interest of the community.[156]

While it is notable that the Committee has underlined the importance of the data custodian and the data trustee acting in the best interest of the data principal, the Report does not enumerate what that would entail. Protecting the interests of the data principal and balancing their best interests with those of the other stakeholders will be a key element of any regulatory framework around NPD. Hence, the concept of best interest and its application in the context of NPD needs to be explained in detail to ensure the creation of an effective regulatory framework around NPD.

Additionally, the Report directs the data custodian to have a duty of care[157] towards the concerned data principal community while handling their NPD. The Report states that "this concept of 'duty of care' is a general set of obligations, which can in time be specified better, by regulatory guidelines, practices, rules, legislations etc. This report does lay down some such duties in the form of anonymisation standards and requirements, protocols and means for safe data sharing, etc. The duty of care should be operationalized through the "best interest" standard, and the prevention of harm to communities and individuals from the processing of Non-Personal Data."

By putting off the enumeration of the relevant standards for duty of care to subsequent regulations, the Report has lost a key opportunity to discuss what an

[155] ibid.

[156] Ibid. p. 23, para 5.1.

[157] *See* Usha Ramanathan, *Tort Law in India*, Annual Survey of Indian Law (2001), 14; James C. Plunkett, The historical foundations of the duty of care, (41) (3) Monash University LR (2015) 716; James Plunkett, The Duty of Care in Negligence (Hart Publishing 2018) <https://www.bloomsburycollections.com/book/the-duty-of-care-in-negligence/ch1-introduction>.

effective regulatory framework to protect data principals could look like and receive feedback on the same through the ongoing public consultation. Amongst all the stakeholders in the NPD regulatory framework, it is the data principal that is the most vulnerable, and it is a discussion of their interests that has received the least focus in the recommendations of the Committee. Detailed public consultations need to be undertaken by the Committee to discuss and enumerate the contours of the principles of best interest and duty of care.

## 3. Non-Personal Data Authority

The Committee has recommended the establishment of a Non-Personal Data Authority (NPDA) to regulate NPD. In this section, we explore whether there is a need for a separate regulatory authority, the composition of the NPDA, the absence of an adjudicatory mechanism, the potential conflict of interest concerns and the need for the establishment of an independent and autonomous NPDA.

### *(a) Regulatory Design*

#### (i) Need for a Separate Regulator

The Committee recommends the creation of a separate regulator - the NPDA.[158] In the Report, the Committee has examined options such as self-regulation, sectoral regulators addressing various aspects of NPD and the potential for the DPA proposed under the PDPB 2019 to address NPD in coordination with the Competition Commission of India.

There may be many synergies in creating a combined regulator to regulate both personal and non-personal data. Besides reducing the burgeoning number of regulators in India, a combined regulator would ensure that we avoid the development of fragmented regulations and inconsistent regulatory standards. One regulator making regulatory decisions, in turn, will allow for more certainty for

[158] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 38.

stakeholders in the space. Markets and market players thrive on regulatory certainty. However, providing such regulatory certainty becomes challenging when there are a multiplicity of regulators overseeing an overlapping regulatory domain. Besides this, having a combined regulator reduces the opportunities for regulatory arbitrage by data businesses.

However, as highlighted by the Committee, the regulatory objectives for the regulation of personal data and non-personal data are at variance with each other.[159] While the regulatory objective of the PDPB 2019 is to protect individual privacy and prevent personal harm, that under the proposed NPD Policy is to unlock the social and economic value of data.[160] The regulatory objectives of the personal data and non-personal data regulatory frameworks are in tension with each other. In Section I.A of this document, we have expressed our concerns with regard to the regulatory objectives set out in the Report. However, if these objectives are to remain as currently enumerated in the Report, it may be ineffective to have a combined regulator overseeing both these domains.

However, several challenges may arise if a separate DPA and NPDA are established. For instance, the Committee envisages that the NPDA will have to work within the framework of the PDPB 2019, and in consultation with the DPA, to mitigate the risk of reidentification of anonymised data. However, experience suggests that coordination between regulators has been challenging.[161] Additionally, who would create regulations around personal and non-personal data that are inextricably linked - the DPA or the NPDA? In this context, it is relevant to note that the Report acknowledges that there has to be harmonisation of data – related directories and disclosures required for personal data and non-personal data, across regulators, so that businesses supply the same information only

[159] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 41.

[160] ibid. p. 40.

[161] *See* Press Trust of India, 'Better coordination needed among financial sector regulators, says RBI' (*Business Standard*, December 31, 2018) <https://www.business-standard.com/article/finance/better-coordination-needed-among-financial-sector-regulators-says-rbi-118123100840_1.html>.

once.[162] This will ensure that businesses do not find the regulatory compliance unnecessarily burdensome.

**(ii) Navigating Overlapping Fields of Regulation**

Past experience suggests that regulators have struggled to navigate overlapping fields of regulation effectively.[163] In the situation that both the DPA and NPDA are established, thought will have to be given to the coordination mechanisms that may have to be set up. It has been recommended that regulators with overlapping regulatory domains enter into a Memorandum of Understanding to establish the roles and responsibilities of the authorities and the procedures for cooperation and information sharing between them.[164] However, since this is a developing approach in the Indian context, it remains to be seen how effective such a mechanism is in facilitating regulators in navigating overlapping regulatory domains.

**(iii) Intersection between the Regulatory objectives of the NPDA and CCI**

A regulatory objective of the NPD policy framework is to prevent the dominance of certain data businesses in the market due to the network effects they have accrued[165], address lack of sufficient levels of competition[166] and address market failures.[167] These regulatory objectives of the NPDA would potentially overlap with the regulatory domain of the Competition Commission of India ("**CCI**"). It is unclear

[162] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 29.

[163] *See* Press Trust of India, 'Better coordination needed among financial sector regulators, says RBI' (*Business Standard*, December 31, 2018) <https://www.business-standard.com/article/finance/better-coordination-needed-among-financial-sector-regulators-says-rbi-118123100840_1.html>.

[164] Report of the Committee to study the Financial Data Management Legal Framework in India, Department of Economic Affairs, Ministry of Finance, October 2016, p. 15, https://dea.gov.in/sites/default/files/FDMC%20Report%20along%20with%20draft%20bill_0.pdf

[165] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) pp. 7-8.

[166] Ibid. pp. 41-42.

[167] Ibid. p. 41.

why the Committee is creating this overlapping regulatory field, what regulatory powers it plans to empower the NPDA with to deal with these challenges, beyond the powers already vested in the CCI, and why it believes that the NPDA would be the appropriate regulator to deal with these issues.

**(iv) Overlap between the NPDA and Sectoral Regulators**

There is likely to be overlap between the regulatory domain of the proposed NPDA and the other sectoral regulators. Data businesses are envisaged as a horizontal classification by the Committee and not an independent industry sector. Businesses across sectors which collect data above a specified threshold will be categorised as a data business.[168] The compliance requirements for data businesses will apply whether or not another sectoral regulator regulates them.[169] The Committee states that the sectoral regulators can "ride on top of the data business compliance requirements i.e., use these compliance requirements as a base and add any sector specific data disclosure requirements."[170]

Given this, there will be a need for coordination between the NPDA and the relevant sectoral regulators.[171] The statute and associated rules will have to be crafted carefully, keeping this potential for overlap and the consequent potential for inconsistencies in regulation or regulatory uncertainty in mind. The Report does acknowledge that the threshold requirements for data businesses may vary with time, context and need and will be fixed and intimated by NPDA, if needed, in consultation with sector regulators.[172]

Additionally, mechanisms would have to be developed, so the regulated data businesses have clarity on the various regulations that apply to them. The Report

[168] Ibid. p. 27.

[169] Ibid. p. 30

[170] ibid.

[171] Ibid. p. 27.

[172] Ibid. p. 28.

envisages using the mechanism of a policy switch.[173] Since this is a relatively novel suggestion, it remains to be seen how effective this will be in ensuring businesses visibility on all the relevant regulations that apply to them in relation to NPD across regulators.

After adequate public consultation, when the regulatory objectives around NPD have been further refined and clarified and keeping in mind the challenges addressed above, an appropriate design for the regulator can be put in place.

***(b) Composition needs to be Specified***

The Report does not describe the composition of the NPDA and the qualifications it envisages for its members. It only states that the NPDA will have some members with relevant industry experience.[174] There needs to be public consultation with the relevant stakeholders to map out the composition of the NPDA. Besides individuals with past industry experience, the NPDA should be composed of individuals with relevant legal and regulatory experience as well as the relevant technical expertise.[175]

***(c) Lack of Grievance Redressal Mechanism***

While the Report lays out the enabling and enforcing role of the NPDA[176], it does not discuss a mechanism for adjudication to appeal decisions of the NPDA and grievance redressal mechanism to address various harms.

**(i) Grievance Redressal for Harms**

[173] Ibid. pp. 71-72.

[174] Ibid. p. 43.

[175] Centre for Communication Governance at National Law University Delhi, Comments on the Draft Personal Data Protection Bill, 2018, https://ccgdelhi.org/wp-content/uploads/2018/10/CCG-NLU-Comments-on-the-PDP-Bill-2018-along-with-Comments-to-the-Srikrishna-Whitepaper.pdf.

[176] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) p. 41.

Data principals need access to a grievance redressal mechanism to address various harms and other concerns that may arise in the context of the NPD regulatory framework. For instance, a grievance redressal mechanism is required to address collective privacy harms arising from the reidentification of anonymised data.

**(ii) Adjudicatory Mechanism to Appeal Decisions of the NPDA**

No adjudicatory mechanism has been specified which can be used to appeal decisions made by the NPDA. For instance, the Report lays out a data sharing mechanism for NPD, where the NPDA is involved in decisions around data-sharing. Data businesses have to share the metadata about the various data they have collected. When a relevant stakeholder identifies a specific dataset they want access to, they may request the data custodian for the detailed underlying data.[177] If the data custodian services this request for data, then the process is complete. However, if the data custodian refuses to share access to the data, then the relevant stakeholder can make a request to the NPDA.[178]

The NPDA will evaluate the request from the perspective of social, public and economic benefit.[179] If the NPDA finds the request to be genuine and that the sharing of the data can result in these benefits, then the NPDA can request the data custodian to share the factual or raw data.[180] Alternatively, if the NPDA finds that the benefits are not real, they can deny the request. No adjudicatory mechanism has been specified with regard to these decisions of the NPDA.

Hence, an appropriate appeal process would need to be set up for the stakeholders to appeal various decisions of the NPDA's. This tribunal would have to be composed of the appropriately qualified legal and technical experts.

[177] Ibid. p. 37.

[178] ibid.

[179] ibid. p. 38.

[180] ibid.

### *(d) Conflict of Interest arising from the Government's role as Several of the Stakeholders under the NPD and the need for an Independent NPDA*

The NPD Committee Report envisages the government as not only the regulator in the form of the NPDA but also a data custodian and data trustee. The NPDA will be responsible for regulating the data principal, data custodian, data trustee and data trusts. The government acting as a data trustee of various community and public data, as data custodian for various public data and as the regulator may give rise to several scenarios of conflict of interest. Additionally, given the government's role as an accessor of data from data businesses and communities, there may be conflict of interest situations that may arise where the government's position may pressure the NPDA to rule in favour of the government's request for data. We discuss some of these scenarios below.

Since, it is likely that, at times, the actions of a government data trustee or government data custodian will need to be inquired into by the NPDA, situations of conflict of interest will need to be mitigated by ensuring that the NPDA is an independent regulator.

#### (i) Conflict of Interest in the Relationship between the NPDA and Data Trustees

Any request for access to community data by a data trustee/community if not acceded to directly by data custodians, would require determination by the NPDA. The Report states that the NPDA will base this decision on the guiding principles laid down in the NPD legislation and in consultation with the appropriate data trustee, in cases involving community NPD.[181] The Report states that this will be further clarified in codes of conducts brought out by the NPDA.[182] Also, the Report

[181] ibid. p 42.

[182] ibid.

states that for a lot of community data, the corresponding government entity or community body may act as the data trustee.[183]

In this scenario, it is unclear why the NPDA is consulting with the data trustee when deciding on whether a data business has to share the community data in cases involving community non-personal data, and what role the trustee will play. It raises the question of whether it would be appropriate for the trustee, who will be an interested party to the decision, to be involved in such decision-making by the NPDA. This becomes particularly important where a data trustee is a government entity, as it would raise serious concerns of conflict of interest, given that the NPDA is also government-appointed.

Further, the Report states that data trustees can recommend to the NPDA the enforcement of soft obligations on data custodians, including "transparency and reporting mechanisms, or stronger ones involving regulation of data practices, within the framework to be specified by legislation".[184] It is unclear why data trustees will be enabled to make recommendations to the NPDA on obligations to be placed on data custodians. This becomes particularly important in scenarios where the data trustee is the government, as it would raise serious concerns of conflict of interest.

**(ii) Government as both Data Trustee and Custodian and Conflict of Interest of the NPDA**

In certain scenarios, the government can be both the data trustee and the data custodian as well as the entity requesting access to community NPD. Firstly, would such a scenario be required to go through the process specified in the NPD regulatory framework? Secondly, such a scenario would give rise to concerns

[183] Ibid. p. 20.

[184] Ibid. p. 21.

about a conflict of interest. It is important to note that data trustees have to communicate their community's 'best interest' to the data custodian.[185]

For instance, if the government wants to collect sensitive NPD of a community and the data custodian and trustee is a central ministry, the data trustee may hand over the information on receiving the request from the government. The community in question may have limited recourse to challenge whether such a decision was in the community's best interest, especially given that the regulator is appointed by the government.

In any of these instances described above, the possibility of bias by the NPDA in its any of regulatory decision making involving the government as a data custodian, trustee or accessor, would negatively implicate the principles of natural justice. Since it is likely that actions of a government data trustee or government data custodian will need to be inquired into by the NPDA, situations of conflict of interest will need to be avoided by empowering the NPDA to be an independent regulator. These scenarios underline the importance of ensuring the independent functioning of the NPDA. An NPDA structured as an independent regulator will have the ability to check the government and mitigate the concern around the occurrence of such instances of conflict of interest.

### *(e) Structuring of the NPDA as an Independent Regulator*

Given the discussion in the previous section, to ensure an effectively functioning NPD regulatory framework, that all stakeholders trust, the NDPA will have to be structured as an independent regulator. The idea of an independent regulator is to provide a level playing field for public and private actors and ensure independence from government influence.[186] An independent regulator allows for the protection

[185] Ibid. p. 19.

[186] Sakshi Balani and Harsimran Kalra, 'Parliamentary Oversight of Regulators' PRS Legislative Research (2012) p. 1 <https://www.prsindia.org/parliamenttrack/discussion-papers/parliamentary-oversight-regulators>.

of the regulatory domain from interference by the executive (elected representatives and bureaucrats).[187]

Independence in the functioning of the NPDA is of utmost importance to ensure that stakeholders such private data businesses, communities and private trustees have confidence in the regulator's impartiality in regulatory decision making, adjudication and enforcement. Few regulators in India have been able to evolve into being autonomous in their decision making and independent from the government's influence.[188] Some mechanisms for enabling autonomy are: (a) having a selection process and selection committee for selecting the members of the NPDA, that is at arms length from the government, (b) having a fixed tenure for NPDA members with limited grounds for dismissal, (c) curtailing of the executive's power to give directions to the NPDA, and (d) ensuring the accountability of the NPDA to Parliament.

The government is likely to be one of the biggest data custodians, data trustees and data accessors under the NPD framework. Consequently, a clear moat needs to be drawn between the government in its role as the data custodian, trustee and accessor and its role as the regulator (NPDA) responsible for framing the regulations, adjudicating access requests for data and enforcing access requests. It is imperative that the NPDA be structured as an independent regulator, as it will:

- Direct the government in its role as a data custodian on whether it needs to share data when requested by individuals or organisations.

[187] *See* 'Regulatory Policies in OECD Countries - From Interventionism to Regulatory Governance' OECD (2002) <https://read.oecd-ilibrary.org/governance/regulatory-policies-in-oecd-countries_9789264177437-en#page6>.

[188] *See* Devesh Kapur and Madhav Khosla, 'State regulation in India – the art of rolling over rather than rolling back' (*The Print*, April 4, 2019) <https://theprint.in/opinion/state-regulation-in-india-the-art-of-rolling-over-rather-than-rolling-back/216647/>.

- Conduct oversight over the government in its role as a data trustee and ensure that it is discharging its fiduciary duty and acting in the best interest of the data principals/community.

- Adjudicate on the government's ability to access data as a data accessor from data businesses/custodians for a sovereign, public interest or economic purpose.

The appointment process for most regulators in India involves significant involvement of the executive. One mechanism for establishing this moat between the NPDA and the government is limiting the government's role in the appointment of the members of the NPDA. A selection committee appointing the NPDA members, independent of the executive, could empower the NPDA to be independent. The selection committee could be composed of Members of Parliament, judicial members and field experts.

However, with this independence afforded to the regulator, adequate transparency and accountability mechanisms need to be put in place to ensure not only the effective functioning of the regulator but also the appropriate development of regulation in consonance with the policy embedded in the statute.[189] One such accountability mechanism is the regulator being made accountable for their actions to Parliament. Within Parliament, the relevant sectoral Standing Committees can be one of the appropriate mechanisms to conduct oversight over the regulator. In the case of the NPDA, the relevant Standing Committee would be the Standing Committee on Information Technology. The Standing Committee can examine the various actions of the NPDA, including the regulations issued, decisions made and enforcement of their decisions. Additionally, the NPDA could be required to submit reports on its functioning to the Standing Committee for

[189] OECD, 'Regulatory Policies in OECD Countries - From Interventionism to Regulatory Governance', (2002) <https://read.oecd-ilibrary.org/governance/regulatory-policies-in-oecd-countries_9789264177437-en#page6.>

review, and members of the NPDA may be required to appear before the Standing Committee to explain the NPDA's functioning.

### 4. Data Trusts

The Report defines data trusts as institutional structures, comprising specific rules and protocols for containing and sharing a given set of data. Data trusts may contain data from multiple sources that is relevant to a particular sector, and required for providing a set of digital or data services. There are various questions on the subject which the Report leaves unanswered. Questions around the legal structure of data trusts, the legal relationship of the trust with other stakeholders like data custodians and data trustees, and the technical architecture of data trusts remain unanswered by the NPD Report.

Another crucial question the report does not address comprehensively is the difference between data trustees and data trusts and their intersection. If data trustees are tasked with managing community non-personal data, can they not be responsible then for facilitation of data sharing between communities and all other stakeholders? Or if data trusts are imagined as institutions for the facilitation of data sharing, then can they not additionally take up the role of data trustees i.e. the management of data rights on behalf of communities?

Additionally, this regime of data trusts and data trustees seems centered around communities and their non-personal data. But, the NPD Report doesn't clearly state whether data principals of public and private non-personal data could benefit from the data trust infrastructure or not.

These are critical questions to answer and must be taken into account while developing the non-personal data governance framework.

Data trusts have been the discussion of a growing body of academic work in recent years.[190] Unfortunately, the Report does not consider or take account of this growing body of academic work.

***Loss of Control over Data***

With advancements in big data analytics the processing of even anonymised or non-personal data can result in privacy harms.[191] This coupled with the fact that average users do not have the requisite knowledge or awareness to make informed decisions about the use of their data creates a unique challenge for privacy protection. As *Delacroix* and *Lawrence* put it, 'very few of us have the time or know-how to understand—let alone control—what parts of our data we are happy to share, and under what terms. The systematic monitoring of one's data presupposes resources that most simply do not have.'[192]

There is a power asymmetry between data processors and those from whom such data is collected. This asymmetry arises from all-pervasive data collection activities of data processors, which allows them to invest time and expertise in complex data processing. In comparison to this, users or individuals from whom such data is collected, knowingly or unknowingly provide such data to many entities, including 'invisible' data brokers. They neither have the expertise nor the time to make informed decisions based on each data processor's data collection policies.[193] When it comes to non-personal data, such complexities get further enhanced as it becomes more difficult to identify the source of such data.

[190] Aline Blankertz, Anouk Ruhaak, and Sylvie Delacroix and Neil D. Lawrence (n 132, 133, and 134).

[191] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol. 126, Springer, Oxford 2017) ch 1.

[192] Sylvie Delacroix and Neil D. Lawrence, 'Bottom-up data Trusts: disturbing the "one size fits all" approach to data governance' (2019) Vol 9, No 4 International Data Privacy Law <https://academic.oup.com/idpl/article/9/4/236/5579842>.

[193] ibid.

There are four key reasons why users/data subjects/data principals are not able to make informed decisions about use of their data: a) complex privacy policies, b) lack of bargaining power, c) context-dependent decisions based on user convenience, and d) the collective dimension of data, where data about an individual/group exposes information on similar individuals/groups.[194]

To mitigate these, many scholars have suggested the introduction of novel entities called data trusts for the management of both personal as well as non-personal data on behalf of data subjects/principals.[195] If combined with the concept of collective privacy, these recommendations could help in the protection of data rights of all data principals in the non-personal data ecosystem. They might be worth exploring while framing a robust non-personal data governance framework.

***Data Trusts as Protector of Data Rights***

While there are many definitions of data trusts proposed by various scholars,[196] according to the simple and lucid definition by *Blankertz* data trusts are intermediaries that aggregate user/ consumers' interests and represent them vis-à-vis data using organisations.[197] *Blanketz* argues that data trusts will be capable of using more technical and legal expertise for greater bargaining power, to negotiate with organisations on the conditions of data use. They will be able to assign rights of access, audit practices, and support enforcement.[198]

This is in line with the Report's suggestions for the appointment of data trustees for community non-personal data. But the Report doesn't make it very clear whether such trustees will only use data trusts for enabling data sharing or not. Further, the Open Data Institute ("**ODI**") defines data trust as 'a legal structure that provides independent

[194] Aline Blankertz, 'Designing Data Trusts: Why We Need to Test Consumer Data Trusts Now' (2020) Stiftung Neue Verantwortung <https://www.stiftung-nv.de/en/publication/designing-data-trusts-why-we-need-test-consumer-data-trusts-now>.

[195] Aline Blankertz, Anouk Ruhaak, and Sylvie Delacroix and Neil D. Lawrence (n 132, 133, and 134).

[196] ibid.

[197] ibid.

[198] ibid.

stewardship of data'.[199] The main responsibilities of such stewards will be deciding who gets access to data, under what conditions, and to whose benefit.[200] The ODI envision data trusts as independent organisations that hold data and have a legally binding responsibility for ensuring that data of its members is used for their benefit. They do take inspiration from legal trusts, but are not such traditional forms of trusts themselves.[201]

*Ruhaak* imagines data trusts to be formed when individuals hand over their data rights to a trustee, who then holds and governs the data on behalf of a group of beneficiaries for a specific purpose.[202] In such a structure, a fiduciary responsibility will be imposed on a trustee and they will not be permitted to profit from such an exercise.[203] This definition is more akin to a traditional legal trust as compared to ODI's or *Blankertz's* definition of data trusts.

*Delacroix and Lawrence* propose data trusts as bottom-up mechanisms, wherein data subjects/principals choose to pool their data rights within the legal framework of a trust.[204] In their proposal, data subjects/principals are both the settlors (authors) and the beneficiaries of the trust. Trustees are placed under a fiduciary responsibility towards the data subjects/principals (i.e. the beneficiaries) and are compelled to manage their data rights.[205]

In this system of bottom-up trusts, wherein data subjects/principals choose to participate voluntarily as compared to a top-down approach where they are mandated to participate in the system, an ecosystem of different trusts exist.[206] Numerous trusts, with a mix of

---

[199] Jack Hardinges, 'Data Trusts in 2020' (*Open Data Institute*, March 17, 2020) <https://theodi.org/article/data-trusts-in-2020/>.

[200] Open Data Institute, 'Data trusts: lessons from three pilots' (2019) <https://theodi.org/?post_type=article&p=7888>.

[201] ibid.

[202] Anouk Ruhaak, 'Data Trusts: Why, What and How' (Nov. 12, 2019) <https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34>.

[203] ibid.

[204] Sylvie Delacroix and Neil D. Lawrence (n 192).

[205] ibid.

[206] ibid.

publicly and privately funded initiatives, with different constitutional terms, allow data subjects/principals to choose from diverse approaches to data governance.[207] Like ODI's approach, even *Delacroix* and *Lawrence*'s approach to data trusts take inspiration from legal trusts, but tweak them to suit needs of the data economy.

The Report does not take into account these useful resources for a deeper exploration of data trusts as novel mechanisms for protecting digital privacy rights. Such creative regulatory tools will help level the playing field[208] between data custodians and data principals. The proposed model of data trustees and data trusts in the Report also seem focussed on exploitation of the economic rights of communities and the protection of community non-personal data. However, the Report does not clearly offer such solutions for data principals of public or private non-personal data.

***Challenges with Data Trusts***

The different kinds of data trusts discussed in this section are still concepts in academic theory. As this idea is fairly new in the field of data governance, practical examples are few and far between.[209] To translate such a regulatory tool from academic studies to practical reality, certain principles will need to be ensured.

Firstly, there is a need for mechanisms that will ensure that these trusts genuinely represent the best interests of their members/beneficiaries.[210] Without such strong alignment in goals, data trusts might enable further exploitation of data, rather than bringing about data governance reform. The genuine representation of interests can be

[207] ibid.

[208] Anouk Ruhaak (n 202).

[209] One long standing example of a data trust is the UK Biobank, which was set up in 2006 and is a data steward for the genetic data and samples of 500,000 people, 'About UK Biobank' <https://www.ukbiobank.ac.uk/about-biobank-uk/>.

[210] Aline Blankertz (n 194).

ensured by placing a legal obligation (possibly in the form of an enforceable trust deed, borrowing from traditional legal trusts) on the trust of a fiduciary duty.[211]

Secondly, data trusts will require money to operate, funding models to ensure the independence of trusts and that they serve their members'/beneficiaries' interests will need to be developed.[212]

The Report suggests that its model of data trusts be managed by public authorities, new neutral bodies (which it does not define), cooperatives, or industry associations. If data trusts are to play a larger role in the data economy, each funding model will have its own challenges. Publicly funded data trusts might lead to larger control of non-personal data by the government; privately funded trusts may tip the balance in the favour of corporations,[213] and neutral bodies or cooperatives might struggle with uptake. Various models may need to be tested before implementation and anyone in particular must not be adopted without proper evaluation.

Thirdly, big questions about the transparency of data trusts remain. As these institutions will be the focal point of data exchange, ensuring their independence and accountability will be crucial. Methods such as auditing, continuous review, and reporting mechanisms will need to be baked into the regulation to ensure the accountability of data trusts.[214]

In light of this discussion on data trusts, we recommend that the government must consider the growing body of academic literature on such regulatory tools before writing them into law. Data trusts could be developed for playing a much larger role in the data

---

[211] Smitha Krishna Prasad, 'Information Fiduciaries and India's Data Protection Law' (2019), Data Catalyst <https://datacatalyst.org/reports/its-a-matter-of-trust-exploring-data-fiduciaries-in-india/>; Rishab Bailey and Trishee Goyal, 'Fiduciary relationships as a means to protect privacy: Examining the use of the fiduciary concept in the draft Personal Data Protection Bill, 2018' (2019) (Working Paper 04), Data Governance Network <https://datagovernance.org/report/fiduciary-relationships-as-a-means-to-protect-privacy>.

[212] ibid.

[213] ibid.

[214] Kieron O'Hara, 'Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship' (Whitepaper) (2019) <https://www.researchgate.net/publication/331275252_Data_Trusts_Ethics_Architecture_and_Governance_for_Trustworthy_Data_Stewardship>.

sharing economy, rather than being mere facilitators of data exchange. They should be evaluated as institutions which have the capabilities of shifting back control in the hands of data principals and assist them in informed decision making.

As data trusts are still a new concept in data governance theory, we recommend that the government keep in mind challenges arising from accountability, funding, and serving beneficiary interest while formalising these structures. Controlled pilots at smaller scale must be run before the formal implementation of one of the many models of data trusts as suggested.

## 5. Data Trusts, Trustees, and Trust Law in India

The NPD Report is silent on the legal structure of data trusts. It states that they are institutional structures comprising specific rules and protocols for containing and sharing a given set of data. Similarly, data trustees are conceptualised as the closest and most appropriate representative bodies for the community concerned, including government entities.

The NPD Report does not specify the legal relationship between data trustees and communities. However, it states that rights over community non-personal data should vest with the data trustee, with the community being the beneficial owner. The data trustee will utilise community non-personal data in the best interest of the community. This loosely borrows the logic of legal trusts as it exists in India.

The legal relationships between data trusts, data trustees, and data custodians also remains unspecified by the NPD Report.

### *Law of Trusts in India*

Trusts in India are broadly of two kinds - private trusts and public trusts. The creation and management of private trusts is governed by the Indian Trusts Act, 1882 ("**Indian Trusts Act**") and public trusts are created under state-specific legislation enacted for that specific purpose. Some examples of laws governing public trusts are - the Bombay Public Trusts

Act, 1950, the Charitable and Religious Trusts Act, 1920, the Religious Endowments Act, 1863, and the Charitable Endowments Act, 1890.[215]

The main reason for the creation of a trust is the transfer of movable or immovable property by an asset owner (a settlor or author of trust), to a specified individual/entity (trustee) for the benefit (economic or social) of either a defined set of people or the public at large (beneficiaries).[216] The trustee is the legal owner of the property and is bound in equity to hold it for the beneficiary.[217]

The key difference between private and public trusts is of ascertained beneficiaries. In case of private trusts, the beneficiaries are a defined set of individual(s) whereas for public trusts, the beneficiaries are uncertain and a fluctuating body of persons.[218]

As per the Indian Trusts Act,[219] the essential ingredients for a private trust are: a) the author or settlor of the trust, who agrees to set aside property for the benefit of the beneficiaries; b) the trustee(s), to whom the ownership is transferred and they manage the trust property on behalf of the beneficiaries; c) the beneficiaries; d) the trust property; and e) legal instrument for the trust for example, a trust deed.[220]

The biggest challenge[221] applying Indian trusts law to data trusts and data trustees in the proposed non-personal data governance framework is whether non-personal data can

[215] Medha Srivastava and Akshit Kapoor (Atlas Law Partners), 'India: Trusts' (*Mondaq*, Dec. 19, 2019) <https://www.mondaq.com/india/trusts/876980/trusts>.

[216] As generally, the beneficiaries are not capable of acting in their own interest, such as minors.

[217] Rishab Bailey and Trishee Goyal, 'Fiduciary relationships as a means to protect privacy: Examining the use of the fiduciary concept in the draft Personal Data Protection Bill, 2018' (2019) (Working Paper 04), Data Governance Network <https://datagovernance.org/report/fiduciary-relationships-as-a-means-to-protect-privacy>.

[218] Nishith Desai Associates, 'Use of Trusts in Wealth Management and Succession Planning' (2017) <http://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research_Papers/Use_of_Trusts_in_Wealth_Management_and_Succession_Planning.pdf>.

[219] Indian Trusts Act 1882, s 6.

[220] Medha Srivastava and Akshit Kapoor (Atlas Law Partners) (n 215).

[221] Kieron O'Hara (n 214).

itself be considered property or not.[222] The transfer of property to a trustee is at the core of a trust relationship.[223] As non-personal data by definition is unidentifiable (unless it is de-anonymised or re-identified)[224] assigning property rights to such data will go against the basic premise of the proposed governance framework. Even the Report while ascribing 'ownership' of non-personal data, states that such ownership might only translate to certain economic or other statutory rights over non-personal data. More importantly, as the report also recognises, data is non-rivalrous i.e. it can be replicated easily and stored with multiple holders. Ascribing traditional property ownership rights to such data is practically unfeasible.

A transfer of economic or other related rights to a data trustee, as proposed by the Report cannot be considered as a transfer or property as required under the Indian Trusts Act. Additionally, in the model as recommended by the Report, the holders of the data rights in question (i.e. a community) might be considered both, the authors/ settlors of the trust and the beneficiaries.

***Fiduciary Relationship***

Though a traditional, legal trust form might not be feasible for data trusts and data trustees to adopt in a non-personal data governance framework, they can still borrow the concept of a fiduciary relationship from the law on trusts. The Report already envisages placing fiduciary duties on data custodians, discussed in greater detail in the next section of this document. Placing fiduciary duties in the nature of traditional legal trusts on data trusts, and data trustees might help provide a distinct framework of duties that recognise the best interests of their 'beneficiaries' (in this case the communities).

[222] s 8 of the Indian Trusts Act specifies that the subject-matter of a trust must be property transferable to the beneficiary; Open Data Institute, 'Data trusts: lessons from three pilots' (2019) <https://theodi.org/?post_type=article&p=7888>.

[223] Rishab Bailey and Trishee Goyal (n 217).

[224] At which stage, provisions of the Personal Data Protection Bill, 2019 will get triggered.

In the law of trusts, the most essential duty is that of the trustee who acts in 'good faith' in the sole interest of the beneficiaries.[225] Section 15 of the Indian Trusts Act imposes a 'duty of care' on trustees. A trustee must deal with the trust-property, as carefully as they would deal with their own property. Ordinarily, a trustee is also not permitted to charge for their services or seek remuneration for them.[226] Multiple provisions of the Indian Trusts Act,[227] put a barr on trustees for acting in self-interest or self-dealing too.[228]

Trustees are bound by law to fulfill the purpose of the trust and to obey the directions of the author of the trust (unless contrary to the benefit of the beneficiaries).[229] Section 49 of the Indian Trusts Act also enables courts to intervene, when a trustee does not exercise their duties reasonably and in good faith. Ultimately, a trustee is liable for breach of trust if they abrogate their duties. As per the Indian Trust Act, a trustee is liable to make good the loss sustained by the beneficiaries in cases of breach of trust.[230]

To illustrate the duty of care of a trustee, in *Fatima Fauzia vs. Syed Ul-Mulk*[231] the High Court of Andhra Pradesh held:

> *The very concept of trust indicates an implied condition that the trustee would use all reasonable diligence and act with due care and attention in the execution of the trust.* When the trustee proposes to sell the trust property, he must be not only reasonable, honest, prudent, cautious, but also diligent to obtain the best price for the trust property, the trustee is, therefore, bound to sell the estate to the best possible advantage to the beneficiaries. (emphasis added)

[225] Rishab Bailey and Trishee Goyal (n 145).

[226] Indian Trusts Act 1882, s 50.

[227] Rishab Bailey and Trishee Goyal (n 145).

[228] Indian Trusts Act 1882, ss 52, 53, and 88.

[229] Indian Trusts Act 1882, s 11.

[230] Indian Trusts Act 1882, s 23.

[231] AIR 1979 AP 229.

These duties and responsibilities are imposed upon the trustees to avoid any abuse of property which is vested with them for the benefit of the beneficiaries. Similar obligations must be imposed on data trusts and data trustees as proposed by the Report, so that these stakeholders always act in the best interest of their beneficiaries (communities or the public at large). In the context of the non-personal data governance framework, it will also be essential to properly lay down what is meant by beneficial interest, as data is not akin to property.

We recommend that the government may consider adopting the core principles of trust law in India, which place various duties and responsibilities on trustees, to data trustees in the non-personal data governance framework. Duties on trustees such as - no profit or self-dealing, to always act in the best interest of beneficiaries, to strictly abide by the subject matter of the trust, and a general duty of care in all its dealings, will help cement the relationship between data trustees and beneficiaries (in this case communities).

As non-personal data is not property, such duties and responsibilities will need to be encoded in law by a separate legislation (probably in the form of the legislation on non-personal data governance as proposed and recommended by the committee of experts) with the subject matter of the trust-relationship being protection of rights and interests distinct from property..

We also recommend that such models and novel mechanisms be tested or piloted, to assess their usefulness, before formal adoption.

## 6. Data Custodians and Data Businesses

This section of the comments examines the two different sets of entities that undertake collection and processing of NPD as identified by the Report – data custodians and data businesses.

*Data Custodians*: The Report identifies entities that undertake the collection, storage, processing, use, etc. of data as a 'data custodian' and states that a data custodian will undertake such activities in a manner that is in the best interest of the data principal[232].

*Data Businesses*: A data business is any organisation, public or private, that is deriving new or additional economic value from data, by collecting, storing, processing, and managing data. The Report suggests that this should be a horizontal classification, applicable across sectors. Registration and compliance with requirements applicable to data businesses, including data sharing requirements will be mandatory after a certain threshold. This threshold is to be determined by the NPDA upon consultation with sectoral regulators.

The Report suggests several different obligations and frameworks to govern the actions of such data custodians and data businesses in the context of the collection, use and processing of NPD. The structuring of this framework suggests that the definition and role of a 'data custodian' is linked to the collection and processing of NPD that has been derived from personal data.

Data businesses on the other hand are entities that deal more broadly with NPD, and the threshold for regulating these entities is based on their function and the volume of NPD they process, rather than the type or origin of the NPD itself, or even the public or private nature of the organisation.

There will presumably be significant overlap between the two categories i.e. data custodians, and data businesses. In fact, depending on the way thresholds are determined and applied in the context of data businesses, it is possible that data custodians will be a sub-set of data businesses. The different legal and regulatory frameworks that would govern data custodians and data businesses are discussed below:

[232] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.7.

### *(i) Fiduciary Duty (applicable to data custodians)*

The Committee states that data custodians will have a fiduciary duty towards the data principal or the community that collectively exercises rights over such data. The appropriate data trustee will communicate to the data custodian, what the best interests of a community of data principals are, on behalf of such community. However, the report also suggests that irrespective of such communication, data custodians have a duty to act in the best interests of the data subject community / data principal.

Fiduciary relationships are relationships where one party provides a specific service to the other, in a manner that requires the recipient of the service to trust / depend on the service provider.[233] They are typically characterised by a specialised service provided by the fiduciary, which requires entrustment of property or power in such fiduciary, and a resultant risk to the beneficiary / service recipient.[234]

In India, fiduciary relationships have been recognised under law[235] and by courts. Where such relationships have been recognised in legislation, the law has not always explained the contours of such relationships, and it has been left to the courts to define the characteristics of fiduciary relationships and the duties that arise from them[236]. The specific duties of a fiduciary depend upon the nature of service provided, and the actual fiduciary relationship in a given case. However,

[233] Smitha Krishna Prasad, 'Information Fiduciaries and India's Data Protection Law' (2019), Data Catalyst <https://datacatalyst.org/reports/its-a-matter-of-trust-exploring-data-fiduciaries-in-india/>.

[234] Tamar T Frankel, *Fiduciary Law* (Oxford University Press 2011) p 4. For a more detailed discussion of fiduciary law in relation to information/data fiduciaries in India *see* Smitha Krishna Prasad, 'Information Fiduciaries and India's Data Protection Law' (2019), Data Catalyst <https://datacatalyst.org/reports/its-a-matter-of-trust-exploring-data-fiduciaries-in-india/>.

[235] Indian Trusts Act, 1882, Guardians and Wards Act, 1890, Benami Transactions (Prohibition) Act, 1988, Right to Information Act, 2005, SEBI (Alternative Investment Fund) Regulations, 2012 and SEBI (Investment Advisers) Regulations, 2013.

[236] Central Board of Secondary Education & Ors *v* Aditya Bandopadhyay & Ors (2011) 8 SCC 497; Reserve Bank of India & Ors *v* Jayantilal N Mistry (2016) 3 SCC 525.

some broad duties that are required of all fiduciaries include: loyalty, care, confidence, candour and good faith.[237]

The idea of fiduciary duties in this Report has been borrowed from the concept of a data fiduciary under the PDPB 2019, which in turn borrows from the concept of information fiduciaries proposed by Jack Balkin in 2014[238]. However, research suggests that the impact of the concept of a data fiduciary under the PDP Bill is limited.[239] Some of the issues of concern that are relevant in the context of imposing fiduciary duties on data custodians as well are discussed below:

(a) *No Profit Rule*: One of the core elements of most definitions of 'fiduciary duty' under Indian law is the idea that a fiduciary cannot gain any benefit or advantage from any activity or relationship that relates to its fiduciary character[240]. This restriction directly conflicts with the business model adopted by most technology companies that run on data and advertising-based business models.

(b) *Duty of Confidentiality*: In this context, the Supreme Court has stated that a fiduciary must use any information entrusted to them by the beneficiary in confidence and must not use it for his own advantage, or for the benefit of another person, or disclose the relevant information to any third party.[241] In the case of this Report, the Committee suggests that data custodians should be

[237] Tamar T Frankel, *Fiduciary Law* (Oxford University Press 2011) p 106; Andrew S. Gold and Paul B. Miller, *Philosophical Foundations of Fiduciary Law* (Oxford University Press 2014).

[238] Jack M. Balkin, 'Information Fiduciaries and the First Amendment' (2016), U.C. Davis Law Review 1183, p 49 <https://lawreview.law.ucdavis.edu/issues/49/4/Lecture/49-4_Balkin.pdf>; *see also* Lina M. Khan and David E. Pozen, 'A Skeptical View of Information Fiduciaries' (2019) <https://harvardlawreview.org/wp-content/uploads/2019/12/497-541_Online.pdf>.

[239] Smitha Krishna Prasad, 'Information Fiduciaries and India's Data Protection Law' (2019), Data Catalyst <https://datacatalyst.org/reports/its-a-matter-of-trust-exploring-data-fiduciaries-in-india/>; Rishab Bailey and Trishee Goyal, 'Fiduciary relationships as a means to protect privacy: Examining the use of the fiduciary concept in the draft Personal Data Protection Bill, 2018' (2019) (Working Paper 04), Data Governance Network <https://datagovernance.org/report/fiduciary-relationships-as-a-means-to-protect-privacy>.

[240] *Reserve Bank of India v Jayantilal N. Mistry* (2016) 3 SCC 525; and Indian Trusts Act, 1882, s 88.

[241] *Central Board of Secondary Education & Ors v Aditya Bandopadhyay & Ors* (2011) 8 SCC 497; *Reserve Bank of India & Ors v Jayantilal N Mistry* (2016) 3 SCC 525.

required to share NPD that is derived from the personal data of individual data principals, or data related to communities with the government and other third parties. This poses a direct conflict with the data custodian's fiduciary duties where such underlying data has been provided to the data custodian in confidence, and there is a risk of re-identification or harm, even post anonymisation of such data.

In addition to the above, depending upon the original purpose for collection and processing of the underlying data or the NPD itself, there maybe potential for a conflict of interest between the data custodian, the data principal / community, and third parties that the data custodian is required to share data with (who in effect may also act as the data custodian's customers).

A fiduciary relationship implies that the principal beneficiary's interest must be centred, and the fiduciary must ensure that there is no conflict of interest. These are issues that can be determined and resolved on a case to case basis, and the framework for regulating NPD will need to be flexible enough to accommodate such situations.

Further, as with the role of data fiduciaries in the context of the PDP Bill, there are concerns regarding the broader legal framework under which fiduciary duties of data custodians may be defined and developed. Fiduciary law has developed on the basis of the duties that arise from a specific set of circumstances, that typically involve a specialised service, requiring certain power over the beneficiary and an element of trust. If all data custodians, i.e. all entities that undertake the collection, storage, processing, use, etc. of data are considered fiduciaries, questions arise regarding whether the fiduciary duties will be enumerated under the regulatory framework itself, and what kind of duties to act in the best interest of the data principal / community exist beyond such a framework.

***(ii) Duty of Care (applicable to Data Custodians)***

In addition, the Report also suggests that data custodians have a duty of care towards the concerned community. The Committee seems to look at the duty of care as a broader set of obligations that will be specified by means of regulatory guidelines, practices, rules, legislation (as opposed to the fiduciary duty which will require action based on communication from the data trustee). The requirements that the Report itself discusses, in terms of anonymisation standards, and protocols for safe data sharing will form a part of such duty of care.

A duty of care is considered to be one component of a broader fiduciary duty.[242] In common law, there also exists a separate duty of care that typically arises in the context of the tort of negligence.[243] An important principle of liability in tort law, the duty and degree of care that is owed to another can vary depending upon the facts of the case.[244]

However, to the extent that a standalone duty of care is suggested under the Report, similar concerns as discussed in the context of fiduciary duties above, exist. Specifically, the Report suggests that this duty of care will be enumerated by means of regulatory guidelines, practices, rules, legislation. In this case, more clarity is needed on how this duty of care will be measured against the best interest of the data principal / community, and implemented on a case by case basis.

***(iii) Soft Obligations (applicable to Data Custodians)***

[242] Tamar T Frankel, *Fiduciary Law* (Oxford University Press 2011).

[243] James C. Plunkett, 'The Historical Foundations of the Duty of Care' (2015) Monash UL Rev. 41, p 716.

[244] Usha Ramanathan, 'Tort Law in India 2001', (2001), International Environmental Law Research Centre, Annual Survey of Indian Law 2001, pp. 615-658 <http://www.ielrc.org/content/a0110.pdf>.

The Report also envisages 'soft obligations' on data custodians, such as transparency and reporting mechanisms to be recommended by the data trustees to the regulatory authority.

The Report currently provides little information to explain what 'soft obligations' will be, beyond the above examples. There should be clear criteria for distinguishing between mandatory and non-mandatory requirements applicable to the different stakeholders identified in the Report, including data custodians. For example, requirements relating to transparency / openness and accountability, are considered essential in the context of personal data protection[245]. Depending upon the nature of actual action to be taken to ensure transparency or the nature of reporting in question, such requirements may need to be mandated by regulation, rather than be treated as 'soft obligations'.

Further, there is limited clarity in terms of the role of different stakeholders in this regard. The Report suggests that data trustees will recommend soft obligations (that the data custodians must undertake) to the NPDA. The process and duties of each stakeholder in this context needs to be elaborated upon, keeping in mind the concerns expressed elsewhere in these comments regarding the data trustee model, and the NPDA.

In addition to the above, the Report states that the appropriate NPD framework legislation, will contain principles and guidelines with regard to the role and incentives for data custodians – referring to ideas such as data privileges, compensation, and well-regulated data markets. The Report does not discuss these ideas or the role of data custodians itself in detail. However, it does acknowledge that the same data sharing obligations cannot be applicable to all businesses, and might negatively impact smaller businesses. To address this, the Report states that the framework will aim to protect and promote the

[245] Planning Commission, Government of India (erstwhile), *Report of the Group of Experts on Privacy* (2012) ("AP Shah Committee Report").

interests of small Indian companies and start-ups by means of incorporating thresholds for data sharing, and graduated sharing obligations.

The Report provides more detailed information on the framework that will be applicable to data businesses, including requirements for registration, data disclosure and compliance. The Report suggests that this framework will be rigorous, yet light-touch, in order to minimize costs of compliance.

Given the different categories of obligations that data custodians and data businesses will need to follow, and the issues highlighted in the context of other stakeholders identified in the Report, this framework highlights two primary concerns. *First*, there is a need for clear demarcation between the duties and obligations of each stakeholder group, the levels of compliance required in a given situation, and the accountability measures in place to ensure compliance. The development of jurisprudence on mandatory obligations, soft obligations, and fiduciary duties will be dependent on clarity in the initial framework in this regard.

*Second*, these obligations and duties need to be explained in a clear manner so that the data principals and communities whose underlying data is being processed as NPD, are able to exercise their rights, and obtain remedies where harms have occurred.

We recommend that an impact assessment that centres the rights of the data principals and communities (including the right to privacy and other relevant rights) is undertaken before implementation of such a framework.

## III. Intersection with the Personal Data Protection Bill, 2019[246]

### A. Data Anonymisation and the Risk of Reidentification

Clause 91 of the PDPB 2019 permits the Central Government, in consultation with the Data Protection Authority of India ("**DPA**"), to direct any data fiduciary or data processor

[246] Submissions in this section have been borrowed from CCG's comments on the Personal Data Protection Bill, 2019, as submitted to the Joint Parliamentary Committee, which is reviewing the bill before it is

to provide anonymised/non-personal data. Since the proposed non-personal data governance framework recommends mandatory cross-sharing of data between multiple stakeholders, and suggests a new law for enforcing such mandate, ideally clause 91 of the PDPB 2019 should be removed or modified before the law is passed.

The unequivocal recognition of the risk of re-identification of anonymised data is a positive step taken by the Report. The Report also acknowledges that even if data remains anonymised, patterns may emerge from it which might compromise the privacy of groups/communities. This is a move in the right direction as compared to the PDPB 2019, which considers the process of anonymisation as irreversible.[247] One of the tasks assigned to the proposed NPDA is to certify rules and technology frameworks for the anonymisation of data. The Report states that the NPDA must work together with the DPA for mitigating against the risks of re-identification of anonymised data.

Clause 50 of the PDPB 2019, empowers the DPA to issue codes of practise for the methods of de-identification and anonymisation of personal data. The Report does not explain how this mandate of the DPA will intersect with the NPDA's function to assign rules for anonymisation. These two requirements must be harmonised to ensure there is no conflict in identifying standards and practices for anonymisation of personal data.

The Report also erroneously states that clause 82(1) of the PDPB 2019 provides for offences related to the re-identification of anonymised data. Clause 82(1)(a) only punishes the re-identification of 'de-identified' personal data, not 'anonymised' personal data. The PDPB 2019, creates a distinction between de-identification and anonymisation. It considers anonymisation as an irreversible process[248] and de-identification a simpler process of removing or masking identifiers from personal data.[249]

considered by Parliament. The comments are protected by parliamentary privilege and thus we have not published them.

[247] The Personal Data Protection Bill 2019, cl 3(2).

[248] ibid.

[249] The Personal Data Protection Bill 2019, cl 3(16).

As the reidentification of anonymised data is a perpetual risk, both the PDPB 2019 and the proposed NPD governance framework must provide for protection against it, in the form of sanctions and penalties. The two proposed authorities are indeed established with separate mandates, they must indeed work together to devise mitigation strategies against the risk of re-identification.

We recommend that between the proposed DPA and NPDA, only one authority must lay down the standards and rules for data anonymisation. However, both regimes of data protection and regulation must provide adequate protection against the dynamic risk of re-identification of anonymised data.

### B. Sensitivity of Non-Personal Data

The Report borrows the concept of sensitivity of data from the PDPB 2019 and adapts it to the non-personal data governance framework. It states the non-personal data could be sensitive due to: a) national security or strategic interests, b) risk of collective harm, c) sensitive business or confidential information, or d) anonymised data which bears a risk of re-identification.

In addition to the above stated categories, the Report states that non-personal data may inherit the sensitivity of its underlying personal data (if derived from it). In effect, if sensitive personal data is anonymised into non-personal data, such data may remain to be sensitive even in anonymised form. Similarly, non-personal data derived from critical personal data will inherit its sensitivity.

The reason for categorizing data as sensitive personal data under the PDPB 2019 is to introduce higher standards for its processing as compared to general personal data. For example, the PDPB 2019 places a higher burden on data fiduciaries to obtain consent from data principals for the processing of sensitive personal data.[250] Additionally, the

[250] The consent for sensitive personal data must be explicitly obtained, after informing the data principal of anh processing which may cause significant harm, and giving a choice to the data principal for granular level of control over giving consent for processing sensitive personal data, the Personal Data Protection Bill 2019, cl 11(3).

PDPB 2019 requires data fiduciaries to keep mirror copies of all sensitive personal data and continue to store critical personal data only in India (data localisation requirements).[251] The Committee recommends that such data mirroring requirements may continue for sensitive and critical non-personal data as well. The Report does not specify any other encumbrance or obligation on the processing of sensitive non-personal data.

As with the data mirroring requirements from the PDPB 2019, there is no clear rationale provided in the NPD Report explaining why sensitive personal data which can be transferred abroad subject to certain conditions, must also be stored in India. Similarly, the government enjoys wide discretion in notifying the categories of critical personal data and hence critical non-personal data too. These powers must be limited by specific criteria for recognition of such categories, and requirements to provide reasoned explanations in writing for such recognition.

There is a need for clarity on the rationale behind borrowing the concept of sensitivity of data from the PDPB 2019 before formalisation of a regulatory framerwork. If such a concept is introduced for non-personal data, then the additional obligations, if any, for the processing of such data should also be clarified. Similarly, the reasoning behind data mirroring and data localisation requirements instituted for non-personal data should also be clarified and discussed in greater detail.

### C. Consent for Anonymised Data

In addition to setting appropriate standards for data anonymisation, the Report recommends that data principals provide their consent for anonymisation and processing of such anonymised data to mitigate against risks of re-identification.

The report states that by doing this, any harms arising out of re-identification of anonymised data may be acted upon by the data principal themselves. The guiding principle being that personal data that has been anonymised, continues to be the non-

[251] The Personal Data Protection Bill 2019, cl 33.

personal data of the data principal. However, this model of consent may not work for protecting data principals from the risks of re-identification of anonymised data and the subsequent loss of privacy.

Consent models are known to be broken[252] on the internet when it comes to the collection and processing of personal data.[253] In the era of big data analytics, it is extremely difficult for individuals to be aware of every data processing activity that might include their data.[254] There is a power and information asymmetry between data processors and individuals from whom they collect such data, whether at an individual or a group level. When people accept to be bound by privacy policies, they rarely understand what these contracts entail and at best such agreements are one-sided in nature.[255] Social and network effects complicate the meaning of informed consent, leaving users/data principals with little room for negotiation on binding contracts.[256]

*Tisné*[257] summarises this conundrum well:

> *The era of machine learning effectively renders individual denial of consent meaningless.* Even if I refuse to use Facebook or Twitter or Amazon - the

[252] Even the Srikrishna Committee which drafted the first version of India's Personal Data Protection Bill, cautioned against the broken nature of online consent models, Committee of Experts under the Chairmanship of Justice B.N. Srikrishna, 'A Free and Fair Digital Economy, Protecting Privacy , Empowering Indians' ("Justice Srikrishna Committee Report") <https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf> ch 3.

[253] Rishab Bailey and others, 'Disclosures in privacy policies: Does "notice and consent" work?' (2018) (Working Paper), NIPFP <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3328289>.

[254] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 1.

[255] Kieron O'Hara, 'Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship' (Whitepaper) (2019) <https://www.researchgate.net/publication/331275252_Data_Trusts_Ethics_Architecture_and_Governance_for_Trustworthy_Data_Stewardship>.

[256] Alessandro Mantelero, 'The future of consumer data protection in the E.U. Re-thinking the "notice and consent" paradigm in the new era of predictive analytics' (2014), Computer Law and Security Review <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2529245>.

[257] Martin Tisné and Marietje Schaake, 'The Data Delusion: Protecting Individual Data Isn't Enough When the Harm Is Collective' (2020) Stanford, Cyber Policy Center, Freeman Spogli Institute <https://cyber.fsi.stanford.edu/publication/data-delusion>.

> fact that everyone around me has joined means there are just as many data points about me to target. (emphasis added)

These challenges get further magnified when it comes to non-personal data. Often, even data processors themselves might not know the exact purpose for collecting certain data whether at the individual or the group level.[258]

The responsibility and burden of protecting data principals from the risks of re-identification of anonymised data should be on data custodians, data businesses and the proposed NDPA. There should be a mix of policy interventions to respond to such risks in the form of - statutory obligations on data custodians, penal consequences for non-adherence of such obligations, review and monitoring of data processing activities in the form of audits, and data protection impact assessments.

It is essential for the non-personal data governance framework to comprehensively define what it means by privacy rights and harms in the context of non-personal data, to bring in transparency and build the mechanics for accountability.

### IV. Additional Areas to be Addressed

There are various other recommendations made by the Report which require further explanation to build a more robust data governance framework which is also in harmony with the PDPB 2019.

[258] Lanah Kammourieh and others, 'Group Privacy in the Age of Big Data' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 3.

## A. Protection from Collective Harms[259]

The NPD Report explicitly states that no anonymisation technique provides perfect irreversibility.[260] It is well accepted that even after data is aggregated, anonymised or encrypted, reversing de-identification is increasingly easy in the big data era.[261]

In addition to recognising the risks of re-identification of anonymised data, the Report does well to acknowledge that collective privacy harms of exploitation and discrimination may result even with the processing of anonymised data.[262] Big data provides a proverbial 'god's eye view' of people's lives and this stems from behavioural data that may be de-identified and subjected to a range of aggregation or blurring techniques in terms of individual identity, but still reflects behavioural patterns of users.[263] Processing of anonymous data or metadata can be as problematic as processing personal data as metadata can just as easily provide intrusive insights into the private communications and lives of people.[264]

It might be useful to illustrate the possibilities of collective privacy harms, by using anonymised data, with a couple of examples:

a) A psychological experiment conducted using Facebook's platform in 2014, showed that such social media companies using big data could influence people's mood on a mass scale. Researchers demonstrated that they could depress or elevate the mood of a massive group of subjects (in this case, two groups of 155,000)

[259] Submissions in this section have been borrowed from CCG's comments on the Personal Data Protection Bill, 2019, as submitted to the Joint Parliamentary Committee, which is reviewing the bill before it is considered by Parliament. The comments are presently protected by parliamentary privilege and are barred from public disclosure.

[260] Ministry of Electronics and Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2020) para 4.5. iv.

[261] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, 'Conclusion: What Do We Know About Group Privacy?' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 12.

[262] *Report by the Committee of Experts on Non-Personal Data Governance Framework* (n 261).

[263] Taylor and others (n 255).

[264] Taylor and others (n 261).

simultaneously by manipulating their news feeds on the social network, noting that doing so had the potential to affect public health and an unknown number of offline behaviours. The anonymisation of users in this case did nothing to protect them from unethical research practices.[265]

b) In another example from 2013, data processors who got access to New York City's taxi trip data (including trip dates and times) were able to infer with a degree of accuracy whether a taxi driver was a devout Muslim or not (taxi license and medallion numbers had been anonymised). Data processors linked pauses in taxi trips with adherence to regularly timed prayer timings to arrive at their conclusion. Such findings and classifications may result in heightened surveillance or discrimination for such groups or communities as a whole.[266]

Such profiling and behavioural manipulation could manifest itself into unforeseen harms such as loss of autonomy and choice and right to self-determination over everyday life decisions.

A challenge of building a data governance framework which provides protection against such harms is that the more collective the harm is the lesser it is visible. As compared to individual discrimination arising from privacy violations (as data processing gets more complex, privacy harms at the individual level may also go unnoticed), discrimination at the group level is likley to be less visible and hard to seek redress for.[267]

The Report proposes that groups/communities be protected from such collective privacy harms by exercising their data rights via data trustees and enforce such rights through the NPDA. The Report does not lay down what these data rights enjoyed by communities/groups are (it does mention economic rights and informational rights, but

[265] Linnet Taylor, 'Safety in Numbers? Group Privacy and Big Data Analyticsin the Developing World' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Philosophical Studies Series, Vol 126, Springer, Oxford 2017) ch 2.

[266] Lanah Kammourieh and others (n 258).

[267] Martin Tisné, *The Data Delusion: Protecting Individual Data Isn't Enough When the Harm Is Collective*' (2020) Stanford Cyber Policy Center <https://cyber.fsi.stanford.edu/publication/data-delusion>.

does not comprehensively discuss what these informational rights might entail). Nor does it lay down in detail the different kinds of collective privacy harms which could affect these communities/groups.

For defining collective privacy harms, some assistance could be taken from the PDPB 2019. Clause 3 (20) of the PDPB 2019 defines harms and includes measures like loss of reputation or humiliation, discriminatory treatment, and unreasonable and unexpected observation and surveillance. Though this definition is not perfect and it does not include privacy harms i.e. direct harm occurring from the loss of privacy by breach or misuse of personal data, it is a good starting point for the government to consider while creating the non-personal data governance framework. Collective privacy harms must include, *inter alia*, instances of intrusive profiling, discriminatory activities, behavioural manipulation, and a complete loss of choice and autonomy over community non-personal data.

As *Tisné* observes,[268] people suffer data related harms in three main ways: a) individual harms, typically protected by personal data protection laws, b) inferred harms, wherein people are profiled, put in similar groups and then targeted for specific purposes such as online ads, and c) optimised harms, these are harms suffered as a result of how machine learning systems are optimised. It is essential that any non-personal data governance framework comprehensively address the latter two types of privacy harms, as personal data laws might not be adequately equipped to deal with such violations.

As individuals do not have requisite knowledge for enforcing their data rights, especially when it comes to group/collective privacy, the burden to establish the loss of collective privacy should not be imposed on them. Measures such as algorithmic transparency of automated systems, public accountability, such as conducting algorithmic impact assessments, and imposition of sanctions on violators, may be explored to respond to the threat of loss of collective privacy.[269]

[268] ibid.

[269] ibid.

In line with these observations, we recommend that the non-personal data governance framework incorporate comprehensive definitions of collective privacy rights and collective privacy harms. The current and future levels of data analytics must be kept in mind while defining these terms.

We also recommend that the burden to enforce data rights must not solely lie with communities/community data principals. Regulation should be aimed at data custodians and data businesses, to ensure that processing activities adopted by them adhere to adequate privacy and security standards.

### B. Inferred Data

The Report categorises inferred/derived data as private non-personal data. However, the Report also states that only 'raw/factual data' pertaining to a community may need to be mandatorily shared by private organisations. In the case of data which is a result of processing value-add, the Report suggests sharing under FRAND[270] terms, or through a data market. The Report does not explain how this will intersect with the definition of 'personal data' as per the PDPB 2019.[271] The PDPB 2019 includes inferences drawn from the collection of personal data for the purpose of profiling, under the ambit of personal data. It also includes data about or relating to a natural person, who is indirectly identifiable, under personal data. Currently, there seems to be an overlap between the definition of personal data and non-personal data, which needs to be resolved for further clarity.

### C. Mixed Datasets

The Report states that mixed datasets (i.e. a combination of personal and non-personal data) represent the majority of datasets in the data economy. It also observes in this context that the EU Regulation on the free flow of non-personal data, applies to non-personal data of mixed datasets, but does not apply to those datasets which are

[270] Fair, reasonable and non-discriminatory terms.

[271] The Personal Data Protection Bill 2019, cl 3(28).

inextricably linked.[272] However, the Report does not explain how the non-personal data governance framework will apply to mixed datasets and whether it will be applicable to datasets which are inextricably linked.

### D. Grievance Redressal

The Report makes it clear that the proposed regulator in the form of an NPDA, will have the dual role of enabling and enforcing. However, as discussed in section II.C.3 of this document, the Report does not specify a mechanism for grievance redressal by the NPDA for data principals.[273] For collective privacy harms like profiling and the risks of privacy violations from the re-identification of anonymised data, data principals must have access to a robust grievance redressal mechanism.

These points of conflict must be addressed and discussed in greater detail, prior to formalisation of a regulatory framework for NPD in order to harmonise the governance of personal data and non-personal data.

[272] Council Regulation (EC) 2018/1807 on a framework for the free flow of non-personal data (Non-Personal Data Regulation) [2018] OJ L 303/59, art 2(2).

[273] The Personal Data Protection Bill 2019 provides for a grievance redressal mechanism to data principals, cl 53.