aapti institute

To
Ministry of Electronics and Information Technology (MeitY)
Government of India,
Electronics Niketan, 6,
CGO Complex, Lodhi Road,
New Delhi - 110003.

31st January 2021

We thank the Ministry of Electronics and Information Technology (MeitY) for the opportunity to provide feedback on the second draft of the Report by the Committee of Experts on Non-Personal Data Governance Framework (hereafter, "the Report" or "report"). The process of consultation is a welcome one, and we hope that this spirit of transparency and due process is continued in the framing of future reports, policies by the Ministry.

We appreciate the Committee for reflecting feedback into this new draft of the Report. We have been particularly pleased to see the purpose of NPD data sharing limited to "public good". However, we believe that this draft is still quite far away from a meaningful, implementable policy document - many of the ideas posited in the document, including the definition of NPD and public good, lack clarity and will lead to confusion and incorrect implementation. This is not to say that the conversation on NPD is not important but in India where a personal data protection framework still does not exist, and sector specific data related regulations are being contemplated - a bit premature.

At Aapti Institute, we have been working on the idea of data stewardship, unlocking the value of data while safeguarding rights. Our detailed submission below builds off our work on data stewardship, and draws from international best practices. We hope that this draft will go through several transparent iterations.

We look forward to engaging further on this issue,

Aapti Institute

aapti institute

Comment on the Revised Report by the Committee of Experts on Non-Personal Data Governance Framework

1. Communities as conceived by the Committee under Section 7.2 is ambiguous and risks causing more harm than good to their interests

The Committee explores economic rights over non-personal data by making a case for community data rights under Section 7.2 of the Report. Accordingly, 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." A community could be virtual or delimited by common geographic, livelihood, economic or social interests. Such communities inherit the right to be compensated for economic value derived from their data as also the right to eliminate or minimise any harm arising from data to the community.

While it is commendable that the Committee recognises that communities ought to be firmly in control of their data, the Report leaves much to be desired by way of a clear articulation of the definition of 'communities' and their relationship with constituent data principals. Not only is the Report's conception of "community" too broad, but also too vague. It is possible that a group of individual data principals sharing a common interest, say members of a Shah Rukh Khan fan club page on Facebook or subscribers of a milk delivery service by Bigbasket Daily in a particular locality, might not be aware that they are indeed a "community". How such communities would create a Section 8 company, Society or Trust and choose a Data Trustee to represent and protect their interests is challenging to comprehend, particularly when the individual data principals involved in common social or economic interactions do not view themselves as a community.[1]

Subsequently, it becomes imperative to ensure that any attempts directed towards instituting community structures must be grounded in individual privacy and consent.[2] To this end, the Report is unclear about the terms of association between data principals and a larger community. Consent, particularly community consent, is another dimension that the Committee has not considered in report. In its current form, individuals must merely consent to the anonymisation of their data and are given an option to withdraw the same, prior to anonymisation. [3] An effort must be made to

---

[1] https://www.medianama.com/2021/01/223-nama-issues-with-definition-of-communities-public-good-and-unabated-sovereign-access-to-non-personal-data/

[2] https://www.orfonline.org/expert-speak/data-development-revisiting-non-personal-data-governance-framework/

[3] Paragraph 5.4(iii) of the Report.

involve communities and/or their representatives at every step of the data value chain[4] - from collection to processing through to data exchange and use by third parties - to ensure that their data is used for social benefit, as purported by the Committee itself. A Trustee, as described in Section 7.7 of the Report, could solve this issue of community consent but comes with its own set of challenges which are explicated below.

Communities are allowed to raise complaints with a regulatory authority, here the Non-personal Data Authority, only through an intermediary - the Trustee - appointed by the Section 8 company, to assert their right to minimise harms arising from NPD about the community. However, the Committee's conception of the Trustee and how such a Trustee will actually uphold the collective interests of the community is not sufficiently explored in the Report. The possibility that the Data Trustee could succumb to regulatory capture, thereby failing the fiduciary duties of care and loyalty stipulated of them,[5] is not addressed by the Committee.[6] In all, there is no accountability mechanism by which a community can ensure that the Trustee acts in their best interests and their data is actually used for "public good". This opaque, ambiguous characterisation of a Trustee, a lack of transparency of their powers and responsibilities engenders a relationship of subordination between the community and Trustee,[7] opening the former to exploitation by the latter.

A useful rubric is to reimagine communities as "bottom-up data trusts" which would foreground collective efforts at data governance, anchored firmly in perspectives of human rights and dignity.[8] This not only places a community in control of their data but also moves away from the current top-down regulatory approaches to data governance advocated by the Committee. This approach disrupts the 'one-size-fits-all' homogenised conception of communities as monolithic entities bound by the same interests and attempts to account for the variation in abilities and interests of the people constituting a community.[9]

2. Data Principal v. Community: Lack of vision in organising principles for Communities and Data Principals; lack of procedure for protection or rights

What is the relationship between a principal and community? What are their terms of association? How is a "community" constituted? How does a principal exit a community?

---

[4] http://www.jenitennison.com/2020/01/17/community-consent.html
[5] Paragraph 7.7(ii) of the Report.
[6] https://thedataeconomylab.com/2020/10/27/aapti-submission-on-the-report-of-the-committee-on-non-personal-data/
[7] https://www.financialexpress.com/opinion/the-definitions-of-community-and-community-data-are-too-broad/2054456/
[8] https://academic.oup.com/idpl/article/9/4/236/5579842
[9] https://thedataeconomylab.com/2020/10/27/aapti-submission-on-the-report-of-the-committee-on-non-personal-data/

Can individual principals raise grievances against a community? What is their relationship with a data custodian and a data trustee?

The NPD Policy has stated that the creation of regulatory structure including Data Trustees, Data Custodians and Data Processors, the Non-personal Data Authority (NPDA) and High Value Datasets (HVDs) will help Data Communities exercise rights over datasets that are relevant to them.

This objective has not been completely achieved in the subsequent provisions of the Report, given the lack of specificity of procedures available to a Community (itself lacking specificity in form) to engage with Data Trustees and Custodians.

Paragraph 7.7 states that Data Trustees are "obligated to establish grievance redressal mechanisms so that the community can raise grievances". It is unclear as to why the burden of coming up with the procedure of instituting grievance redressal mechanisms has been laid completely at the feet of the Data Trustee, without principles on how it ought to be designed. At the level of an initial analysis, the process should comply with tenets of principles of natural justice and sound administrative law. Excessive delegation is a violation of these principles,[10] and were this policy to progress to any kind of draft legislation, it would require greater elaboration on the form of recourse available to Data Communities in order to engage with the activities of Data Trustees.

Similarly, it is also unclear as to how Data Trustees may engage with the actions of Data Custodians in order to prevent or mandate actions with regard to their respective Data Communities, or with general regard to Data Principals as a whole. Given the level of detail the Report has gone into on a number of other subjects, this is an oversight in the drafting of the Report.

<u>Grievance redressal processes under the Report need to be fleshed out in terms of the specific responsibilities of regulatory bodies in this regard – including those of Data Trustees, Data Custodians, and Data Businesses. The body to be approached by a Data Principal or Community to articulate grievances, and the form of submission of these grievances, should be made explicit in the Report.</u>

3. **<u>The Report requires a clearer articulation of the harms from which it seeks to protect Data Principals and Communities</u>**

[10] *Hamdard Dawakhana v. UoI*, (AIR 1960 SC 554).

The Report states that Data Custodians and Trustees have a right to act to minimize harms, and they are in fact obligated to carry out this task by virtue of their duty of care towards the relevant Data Communities. What forms these harms take, however, has not been specified.

This lack of detail leaves stakeholders in the dark as to what possible claims or causes of action they may have. Definition of harms under the Report cannot be left to the imagination of stakeholders.

The one harm that has been clearly articulated in the Report is reidentification and de-anonymization of NPD. That remains the sole responsibility of the Data Custodian. Again, the form of grievance mechanism needs to be specified in the Report. It cannot be left to the individual decisions of each Data Custodian. Uniformity of grievance redressal processes is essential to a transparent system where Data Principals and Communities are able to participate in order to protect their interests.

The Report must lay out the processes involved and form of submissions to be made as part of the grievance redressal requirements prescribed in the Report under Paragraphs 7.2(ii), 7.7(ii) and 7.4(iv). The procedures are left to the individual preferences of organisations involved. Clarity of procedure and uniformity of process is necessary for transparency in protection of harms against Data Communities and Principals. the Report in its current form ignores Constitutional principles of procedural justice.

4. **The Report fails to make a convincing case for the NPDA outlined in Section 7.10 and avoids any discussion on the possibility of sectoral overlaps that could manifest as enhanced regulatory burden for data-driven businesses.**

The Committee recommends the creation of an independent regulatory body, called the Non-personal Data Authority (NPDA), to govern NPD. The NPDA is endowed with a set of enabling functions - from unlocking the economic value of data for India and its communities to maintenance of metadata registries. On the other hand, its enforcing functions include - establishing rights over Indian NPD; adjudicating on sharing of HVDs in cases where a Data Custodian refuses to share relevant information; and preventing misuse of data by addressing privacy concerns around reidentification of NPD.

Despite its attempt at articulating the nature and roles of the NPDA, the Committee has failed to make a convincing case for the institution of an independent regulatory

authority.[11] For one, a variety of existing mechanisms could achieve the same ends as that of the NPDA. An MoU between two consenting parties, say a hyperlocal delivery platform and a govt. agency, for release of data relating to self-employed gig workers in order to ascertain the number of informal workers in a city, could result in a similar disclosure of data, without the regulation of an overarching entity such as the NPDA.

It is also possible that a public authority already possesses requisite data and the same could be accessed via a simple RTI application. This is particularly true of various heads mentioned under the proposed framework for High-value Datasets.[12] The government has already amassed critical data relating to financial inclusion,[13] education[14], agriculture[15] and skill development[16] and efforts should be made to share this data as public good HVDs for the benefit of the larger Indian community.

Further, the Committee has failed to consider the ramifications of regulatory overlaps as a result of the creation of an NPDA. Firstly, the jurisdiction of DPA vis. NPDA is ambivalent inasmuch the presence of confounding variables such as mixed datasets and reidentification of NPD creates challenges for outlining clear adjudicatory mandates for the two authorities. In this context, the regulatory parameter for NPDA is undefined and requires more deliberation on the principles that compel the creation of the NPDA. As also, the Committee proposes that HVDs be excluded from copyright protection. The rationale offered is that formulation of HVDs does not entail any creativity or skill, but mere compilation of existing "fields of data". However, the existence of certain "fields of data" might be a product of innovation and skill, producing competing claims for "trade secrets" protection simultaneously. In such a case, the mandate of the NPDA to compel disclosure by Data Custodians will necessarily be in conflict with the existing IPR regime. [17]

Lastly, the NPDA's scope impinges on the defined powers of the Competition Commision of India (CCI) which could also compel disclosure of data for public welfare purposes under the 'essential facilities' doctrine.[18] This doctrine could be applied in cases of denial of market access by a dominant enterprise, as could be the concern with large Data Businesses.[19]

---

[11] https://www.medianama.com/2021/01/223-non-personal-data-authority/
[12] See Paragraph 7.6 of the Report.
[13] https://financialservices.gov.in/financial-inclusion
[14] http://mospi.nic.in/statistical-year-book-india/2017/198
[15] http://mospi.nic.in/agriculture-statistics
[16] https://niti.gov.in/verticals/skill-development-and-employment
[17] https://hasgeek.com/PrivacyMode/npd-week/sub/regulatory-overlap-issues-in-the-npd-committee-rep-3HQVNRki8tyQdkhGS6x9fn?utm_campaign=webshare
[18] See Section 4, The Competition Act, 2002
[19] https://nujssitc.wordpress.com/2018/04/07/position-of-essential-facilities-doctrine-in-india/#_ednref1

The compounded regulatory burden on businesses is an important consideration which ought to be factored into the Committee's proposal to create an NPDA. In its current form, the Report suggests that sectoral regulators can delineate protocols for data sharing, in addition to those imposed by the sector-agnostic horizontal NPDA. This could result in a situation of regulatory excess where data-driven businesses will have to comply with multiple legislations: the PDP Bill in cases of personal data, the NPD framework for anonymised data and a third set of compliance obligations outlined by sectoral regulators. Efforts must be made to harmonise compliance requirements across regulatory bodies such as the DPA, NDPA and other sector-specific authorities to ease the consequent regulatory burden on businesses.

5. **Dichotomy between the functions of a non-profit under Paragraph 7.2(ii) and a Trustee under Paragraph 7.7 of the Report, leading to inadequate protection against harms from processing of NPD**

The Report appoints a Data Trustee to exercise a duty of care towards Data Principals and Data Communities.

The role of the Data Trustee is to protect the interest of its user groups and communities. The Data Trustee firstly plays a pivotal role in the creation of the relevant HVD. It then takes on a role of safeguarding interests of Data Communities. Paragraph 7.7(iv) and 7.4(v) make a couple of things clear in this regard: the Data Trustee, in its interactions with the Data Custodian, both requests as well as supplies data – the latter in cases where the Data Custodian plays the role of a data requester. However, it is the Data Trustee that is put expressly in charge of each HVD.[20]

However, the Community relevant to the HVD *cannot approach the Data Trustee to express their concerns* on its usage through a grievance redressal mechanism. In order to do this, they are required to approach a different non-profit, as described under Paragraph 7.2(ii). Any engagement on harm experienced by Data Communities is required to go through this alternate regulatory route, as opposed to the better-established framework of the Data Trustee under the Report.

Given that articulation of economic interests of the Community is in the hands of the Data Trustee, it is unclear as to why the prevention of harm has been delegated to a different entity. Both these processes aim to protect interests of the same group – the Community impacted by the HVD. The Report adds to regulatory burden on Data

[20] Paragraph 7.7(iv) of the Report.

Communities when different platforms are required to be approached for different concerns. Indeed, the division of concerns between these two kinds of organisations has also not been clearly laid out. What is clear is that *the organization protecting Data Principals and Communities against harm does not have a duty of care towards them.* This is in and of itself is an egregious oversight in the regulatory approach taken by the Report.

We recommend that the responsibility of protection against harm be given at least the same strength of grievance redressal process and accountability standards as that of protecting the interests of Data Communities in granting access to HVDs. Potential harm should be an explicit consideration in the Data Trustee's decision in granting access for HVDs to any data requester.

6. **Vague relationship between the Community and Data Trustee open to exploitation:**

Under Paragraph 7.7(ii), the Data Trustee has a 'duty of care' to the user groups relevant to the NPD that it handles. The rights protected here are not of individual Data Principals, but rather those of the community relevant to the HVD, as described in Paragraph 7.2 of the Report.

However, how exactly the Community ensures that the Data Trustee is acting in its interest is a question that is not addressed by the Report. Aggrieved parties are not even afforded a grievance redressal mechanism against the Data Trustee for harms – there is no direct accountability measure defined by the Report against the Trustee. It does state, in Paragraph 7.7(ii), that the Trustee is obligated to "establish a grievance redressal mechanism", with no explanation of the process or principles to be adhered to in such a procedure, nor the responsibilities of the bodies and authorities involved. There is also no clear appellate procedure that has been prescribed as part of this process.

The grievance redressal process available to Communities under the Report has been ignored completely. There is no elaboration on the principles or procedure that will be involved in the decision making process of Data Trustees. There is also no mention of the role of the Community in this process. The only obligation on Data Trustees is to design their own procedure, without any requirement of principles to be followed in such a procedure. For this Policy to move towards legal force, it should fill these gaps so that it may pass basic Constitutional tests against vagueness and granting excessive discretionary powers.

7. **<u>Anonymisation and data sanitization with respect to "mixed datasets" are loopholes in the framework that could be used to circumvent mandatory data sharing obligations under Section 7.4 (iii).</u>**

The mandatory data sharing obligations sought to be imposed on Data Custodians and Data Businesses by the Revised Report introduces complications of regulatory arbitrage and enhanced compliance burden on private entities. This is highlighted by the clauses on anonymisation and governance of mixed datasets - wherein Data Custodians/Businesses could simply alter their terms of service and choose not to anonymise data or claim that they operate with mixed datasets (containing both PII and NPD) - which could be used to circumvent the NPD framework itself as explained below.

The Committee, under Section 5.1, delves into the interface between regulation for non-personal data and the proposed Personal Data Bill, 2019. In an attempt to avoid any confusion arising from overlaps between the two frameworks, the Committee suggests that all personal data to which certain "transformative techniques" are applied and consequently render it impossible to identify the underlying data principal, should be considered as non-personal data.

Further, the Committee stipulates certain consent mechanisms be instituted for collection and processing of non-personal data. Under Section 5.4, the Committee requires data collectors to notify data principals that their data will be anonymised and offer an to principals to "opt out of anonymization", if their data is yet to be anonymized.

The Committee also notes that certain "High-Value Datasets" be created by a Trustee[21] who will collect requisite data from Data Custodians. In turn, Data Custodians are obligated to share relevant NPD when such requests are made.[22] The Committee, in this revised report, has retained provisions that necessarily mandates coercive data sharing by private entities. This is problematic because the <u>creation of structured datasets, including raw and factual datasets, is a resource-intensive activity which could potentially disincentivise Data Custodians and Data Businesses from anonymising data at all.</u> In fact, it is possible that <u>mandatory data sharing could engender practices of regulatory arbitrage - a phenomenon by which Data Custodians could circumvent mandatory data sharing obligations by explicitly choosing not to anonymise data.</u> Regulatory arbitrage is all but common in the tech industry where

[21] Paragraph 7.7(iv) of the Report.
[22] Paragraph 7.4(iii) of the Report.

companies like Facebook have modified its terms of service to escape stringent privacy guidelines under the GDPR.[23]

In the Indian context, Data Custodians could avoid the mandatory disclosure obligations imposed on them by either choosing not to anonymise user data or claim that such data is a "mixed dataset"[24] which is "inextricably linked" to personal data of users, thereby bypassing the mandatory data sharing obligations. Consequently, the Data Custodians would not fall under the purview of NPD framework, but that of the Data Protection Authority mentioned in the PDP Bill. This glaring loophole in the NPD framework is only strengthened by a lack of coherence in the definition of a "mixed dataset", specifically the scope of what constitutes as being "inextricably linked" to personal data. Whether such a link is determined by economic or privacy considerations is unclear although their association to personal data might be tenuous, at best. The Committee has failed to clarify about the degree of sensitivity attributed to mixed datasets, whether they're critical, sensitive or personal in nature, which would determine the related data localisation requirements pertaining to different categories of data. Thus, a situation of arbitrage emerges in which Data Custodians and Data Businesses could alter their terms of service and subsequently introduce flexibility in regulatory compliance required of such Custodians and Businesses.

Lastly, concomitant concerns arise when one considers the enhanced regulatory burden placed on small and medium enterprises as a result of mandatory data sharing obligations stipulated by the NPD framework. As stated, structuring raw or factual datasets is a resource-intensive activity which could divert investments away from innovation.[25] A mere processing charge[26] paid to Data Custodians could hardly compensate for the time and resources lost to regulatory compliance foisted by this framework. Small businesses would be better disposed to employ their limited resources in data analytics which could yield better outcomes for their operations and ensure profitability than merely obtaining raw datasets through HVDs that lack the insights which data analytics promises to provide. Similarly, mandatory data sharing is associated with increased risks to privacy and security where Data Custodians may be forced to share data with Trustee/other requestors who do not follow the best practices of maintaining data security.

Alternatives to Mandatory Data-sharing: Creating a voluntary sharing ecosystem

Data sharing cannot occur just on the basis of mandates because the Report fails to articulate a principle jurisdictional basis for recommending mandatory data sharing,

[23] https://www.bbc.com/news/technology-43822184
[24] See Paragraph 5.1(v) of the Report.
[25] https://www.bsa.org/files/policy-filings/09102020indiabsanpd.pdf
[26] See Paragraph 8.5(vi) of the Report.

without accounting for the technical infrastructure that ought to be instituted for such a move. Additionally, there is a need to create an ecosystem that encourages and incentivises the sharing of data. This ecosystem must create trust in the process of sharing, and clearly express the value of sharing across stakeholders. It should further consist of technical infrastructure (like Estonia's X-Road) to enable sharing, and capacities within organisations, and beyond to address additional requirements of sharing data.

Accordingly, a plausible alternative to mandatory data sharing obligations imposed by the NPD framework is the creation of "data marketplaces". Data marketplaces would provide an avenue for data providers/suppliers to offer curated datasets at a price where data consumers/requestors can purchase or subscribe to such datasets.[27] Such marketplaces could be hosted by third parties that promise to ensure data security and a coherent platform for secure data exchange between data suppliers and consumers. Similarly, the proposed EU Data Governance Act, 2020 makes provisions for "data altruism". Data altruism refers to a mechanism of data sharing in which data subjects (data principals in India) consent to processing of their personal data, or permissions of other data holders (custodians, processors) are obtained to authorise the use of their non-personal data without reward, for purposes of public welfare, such as scientific research or enhanced delivery of public services.[28] Data altruism promises to unlock the value of data for "public good" without the coercive implications of mandatory data sharing imposed by the NPD framework. Incentivising voluntary data sharing by Data Custodians through a system of economic and public incentives is a more sustainable approach to create a coherent data ecosystem in which all actors - Custodians, Businesses, Trustees and the State - stand to harness the potential of data for the digital economy.

[27] https://www.ikigailaw.com/data-marketplaces-a-schema-for-voluntary-data-sharing/
[28] https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2103