CENTRE FOR COMMUNICATION GOVERNANCE AT NATIONAL LAW UNIVERSITY DELHI

COMMENTS TO MEITY ON THE DRAFT NATIONAL DATA GOVERNANCE FRAMEWORK POLICY[1]

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

[1] Authored by Joanne D'Cunha and Bilal Mohamed. Reviewed and edited by Jhalak M. Kakkar and Shashank Mohan.

Table of Contents

The Centre for Communication Governance is an academic research centre within the National Law University Delhi and is dedicated to working on information technology law and policy in India. It seeks to embed good governance and constitutional principles in this domain through rigorous academic research, policy input and capacity building.

We are grateful to the Ministry of Electronics and Information Technology ('MeitY') for inviting public comments and suggestions on the National Data Governance Framework Policy ('NDG Policy').

The NDG Policy published by the MeitY is a welcome step insofar as it seeks to ensure privacy respecting sharing of non personal data and harnessing it for better data-led governance. It is further encouraging to see that the Policy has taken into consideration stakeholder recommendations by moving away the monetisation of non personal data proposed in the Draft India Data Accessibility and Use Policy.

However, we wish to highlight certain aspects of the Policy that merit further scrutiny.

In the first part of the comments, we highlight the need for a comprehensive data protection legislation prior to the implementation of a data governance framework, to safeguard citizens from potential privacy risks. In Part II (section 2-3) of this document, we indicate that it may be crucial for the NDG Policy to be clearer in its objectives, scope, and key terminology, in order to ensure it is effectively operationalised. In Part III (section 4) we highlight concerns around re-identification of anonymised data and the individual and collective privacy harms that may arise from the lack of a pre-existing data protection regulation. In Part IV (section 5-6) our comments touch upon data accessibility and the NDG Policy's proposed institutional framework vis-a-vis the IDMO and the functional overlaps with regulatory structures proposed in other data governance frameworks. Before concluding, in Part V (section 7) we reiterate the need for a data protection legislation by comparing approaches to data governance frameworks in other jurisdictions.

These comments set out our concerns with respect to some of the Policy's clauses.

1. Need for a comprehensive data protection legislation

The NDG Policy aims to use non personal government data for 'effective digital government, public good and innovation'. To achieve this, it envisions the creation of an India Datasets program to house such non personal and anonymised government data and additionally, such data provided by private entities.

We would like to highlight that in the absence of a comprehensive data protection legislation, the collection, processing and use of non personal data as laid out under the NDG Policy could pose privacy and data protection risks for Indians.

With anonymised data, there is a risk of re-identification of an individual's personal data. Our submission delves into further detail on re-identification in section 4. However, when personal data has been re-identified, it is crucial that it receives the protection of a data protection legislation. Affording such protection is important as with the possibility of re-identification of anonymised data, violation of individual privacy, and potential discrimination and exclusion may occur. However, even without re-identification, as non personal and anonymised data tends to be largely aggregated, harms can arise. There is increasing evidence that privacy harms, discrimination and exclusion can be more collective in nature, i.e., impacting groups of individuals at a time.[2]

Without a data protection legislation in place, it may be critical to safeguard against individual and possibly collective harms from systematic data sharing involving public and private entities. To achieve this, it may be necessary for data protection principles such as lawfulness, consent, and purpose limitation, as laid down in the Data Protection Bill 2021 ('DPB 2021'), to be codified within the NDG Policy.[3] This is to ensure that in

[2] 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.

[3] Report of the Joint Committee on the Personal Data Protection Bill, 2019, s 57(2)(d) available at https://prsindia.org/files/bills_acts/bills_parliament/2019/Joint_Committee_on_the_Personal_Data_Protection_Bill_2019.pdf.

the absence of a data protection legislation, principles are also established within the Policy to afford safeguards for re-identified data. The submissions made in our comments elaborate these concerns to further demonstrate the need for such legislation before establishing data dependent governance structures and standards across various levels of government.

The introduction of a non personal data governance framework should not precede that of a data protection law. As the DPB 2021 is still pending legislative approval, the Government could reconsider the implementation of this Policy. The NDG Policy and its overall objectives would substantially benefit from the safeguards and processes laid down by the DPB 2021. Further, considering that the DPB 2021 also incorporates non personal data within its ambit, there might be overlap in governance. Therefore, introducing other data governance frameworks once the DPB 2021 is passed, could aid in addressing the above concerns. Our inputs, hereon, reiterate foundational privacy concerns that are currently being debated around the use of non personal data.

2. A need for clarity on the objectives and purpose of the NDG Policy

2.1 Objectives

The NDG Policy envisions the management and sharing of non personal government data through its Datasets program. To achieve this, the Policy proposes the establishment of the India Data Management Office ('IDMO'), which shall, among other things, be responsible for developing rules, standards, and guidelines for data collection and management processes and systems. The objectives and the applicability of the NDG Policy focus specifically on data collected and managed by the government. However, various parts of the Policy, such as paragraphs 5.4 and 6.3, encourage private players to share data collected by them without explicitly engaging with how the Policy would be applicable to them.

The manner in which the NDG Policy will interact with the proposed non Personal Data Governance Framework ('the NPD Framework') is not entirely clear. For instance, it is currently unknown whether the NDG Policy intends to replace the proposed NPD Framework or if the Framework will be implemented separately or simultaneously. However, since the NDG Policy intends to include data provided by private companies, it is unclear whether these datasets will fall within the ambit of the NDG Policy and the government's Datasets Program or whether the proposed NPD Framework and its institutional structures (i.e., data trusts/data custodians) will be applicable.

Careful consideration must be given to the position of this Policy with regard to other potential legislation such as the DPB 2021 and the proposed NPD Framework prior to conceptualising the contours of such a data governance framework.

2.2 Purpose

With regard to the purpose of the NDG Policy, a significant point to note is that the Policy advocates for the potential of non personal data in helping catalyse the transformation of AI, analytics, and the start-up ecosystem. The NDG Policy does not, however, evidence these benefits. For instance, it may be helpful for the NDG Policy to clearly indicate how it envisions the organising and sharing of government data or creating a repository of datasets, will practically further AI capabilities. Although directly enabling access to data would be helpful to AI and data led research and start-ups, a primary concern here is whether this would be useful if there is a lack of adequate computing abilities. This is an important challenge considering the perception that increased collection, processing, and sharing of data is more beneficial to deploy AI more effectively.

Such an approach may warrant reconsideration as scholars point out that recent advances in AI suggest that developing better computing powers should receive greater focus instead.[4] Computer scientists have been working to train AI with better reasoning capabilities that mimic human decision making, instead of simply increasing data fed to

[4] H. James Wilson, Paul R Daugherty and Chase Davenport, 'The Future Of AI Will Be About Less Data, Not More' [2019] Harvard Business Review <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more> accessed 10 June 2022.

AI systems.[5] With these developments, it might be worth re-assessing whether by simply creating a quantum of datasets we will be able to 'transform AI'. Better efforts might be placed on how to address the increasing amount of computational power required to train AI more efficiently.

3. Definitions and linkages with other proposed frameworks

The NDG Policy suffers from ambiguity around certain key terms and phrases such as non personal data, anonymisation, data usage rights, Open Data Portal, Chief Data Officers (CDOs), datasets ecosystem, and ownership of data. Concepts such as ethical and fair use, trust and safety, and data security have been merely mentioned but the NDG Policy is silent about broad principles or processes for achieving them. It simply states that the IDMO will set out the rules and protocols necessary.

In this section, we highlight concerns specifically with the definition of non personal data while anonymisation is dealt with in sub-section 4.3. With regards to non personal data, it may be necessary for the NDG Policy to define the term to indicate whether it will align with the definition of non personal data in the DPB 2021.

Clarity on the manner in which it will interact with the proposed NPD Framework would be useful to understand for the following two reasons. Firstly, whether non personal data includes anonymised data or if they are distinct categories. Secondly, if non personal data includes anonymised data and whether it is envisaged that the definition of non personal data will also be categorised into public data[6], community data[7], and private data as recommended in the proposed NPD Framework.[8]

Similarly, as the NDG Policy is silent on its interaction with the proposed NPD Framework, it might be useful to understand if the concept of sensitivity under the proposed framework will be incorporated to categorise datasets. The proposed NPD

[5] ibid.
[6] such as 'anonymised land records data, vehicle registration data etc.'
[7] such as 'datasets collected by municipal corporations and public electric utilities.
[8] such as inferred or derived data/insights, involving application of algorithms, proprietary knowledge'.

Framework states that non personal data could be sensitive in nature under specific conditions and may also inherit sensitivity if it was derived from sensitive personal data. Borrowing from the DPB 2021 requirement to store critical data in India, the proposed NPD Framework suggests similar storage conditions for non personal sensitive data. However, the proposed NPD Framework does not include other special obligations regarding processing and access to sensitive data, unlike the DPB 2021. If the NDG Policy intends to categorise certain non personal datasets as sensitive, it might be worth considering whether introducing risk-based obligations when sharing sensitive data can further protect against privacy harms.[9] For example, anonymised health data may be more sensitive in nature than road traffic data and could require different obligations.

It may be crucial to establish and harmonise the definitions of terms in the NDG Policy with the DPB 2021 and possibly the proposed NPD Framework to avoid any confusion in interpretation and challenges with potential implementation.

4. Privacy harms

4.1 Concerns around re-identification of data

The NDG Policy in paragraph 2 envisions the collection, processing and use of non personal and anonymised data in various ways for the overall purpose of furthering data-led governance. In paragraph 6.4, the NDG Policy tasks the IDMO with prescribing anonymisation standards for all entities to adhere to. Previous submissions by this Centre[10] and various other stakeholders have referred to recent literature demonstrating inadequacies in anonymisation techniques, and the challenges of the re-identification of anonymised data.[11] Concerns around re-identification of non personal and anonymised

[9] Obligations that are determined by the likelihood of re-identification and degree of sensitivity of the data/datasets involved.

[10] Centre for Communication Governance at National Law University, Delhi, 'Submission of Comments on Report by the Committee of Experts on Non-Personal Data Governance Framework' (2020) <https://ccgdelhi.org/wp-content/uploads/2020/09/CCG-NLU-Comments-to-MeitY-on-the-Report-by-the-Committee-of-Experts-on-Non-Personal-Data-Governance-Framework.pdf>.

[11] Paul Ohm, 'Broken Promises Of Privacy: Responding To The Surprising Failure Of Anonymization.' (2022) 57 UCLA Law Review <https://www.uclalawreview.org/pdf/57-6-3.pdf>.

data and related privacy risks are exacerbated by the lack of a comprehensive data protection legislation in India.

The NDG Policy also does not discuss the challenges around mixed data-sets and the blurring lines between personal and non personal data. As the EU's Guidance on the Regulation on a framework for the free flow of non personal data in the European Union notes, most datasets are likely to contain both personal and non personal data, and is especially the case in data originating from individuals.[12] Even within non personal data, the Policy does not distinguish between non personal data originating from individuals and non human non personal data. This categorisation is useful as there are risks of re-identification for the former and may not be as heightened for the latter.

Given these challenges, the continual risk of re-identification should be protected against by possibly setting strong privacy respecting standards for anonymisation that keep up with technological developments.[13]

4.2 Individual and collective harms arising out of re-identification

The concerns of re-identification do not simply relate to the identification of personal data that was once anonymised. They may extend beyond personal data concerns to where the combination of multiple datasets provide greater context on an individual or even groups of individuals. Various factors may contribute to creating such context, where the technological capabilities such as advanced data analytics or the ability to access and combine numerous datasets may play a role in the degree of the risk of re-identification.[14] As studies have shown, the risks of reidentification of anonymised data are proportional to the number of data points available. So, while the NDG Policy envisages all datasets to be anonymised, the increased availability of datasets from different sources can weaken the anonymisation. When re-identification occurs, privacy

[12] European Commission, 'Guidance On The Regulation On A Framework For The Free Flow Of Non-Personal Data In The European Union' (European Commission 2019).

[13] The approach to the development of these standards have been discussed in section 4 below

[14] Nadezhda Purtova, 'The Law Of Everything. Broad Concept Of Personal Data And Future Of EU Data Protection Law' (2018) 10 Law, Innovation and Technology.

related harms, discrimination, and exclusion concerns may arise from unbridled data flows. To account for this, the NDG Policy could necessitate data protection principles such as purpose specification, data minimisation, accountability, amongst other recognised principles, to guide data sharing.[15] More importantly, without these safeguards and obligations in place, the resulting privacy harms would squarely stand in violation of crucial elements of privacy laid down in *Puttaswamy*.[16]

As also recognised by the proposed NPD framework, some datasets can be more sensitive/critical in nature. Implementation of a policy that envisions the sharing of such datasets, without requiring higher levels of security and privacy, could lead to greater privacy risks. These concerns are particularly exacerbated when we consider that certain government departments such as the National Health Authority, the Income Tax Department and the Central Board of Indirect Taxes and Customs, already possess sensitive personal data such as health records, financial records and biometric information. One of the key objectives of the NDG Policy is to harness and standardise data sharing between government entities for a 'data-led governance' that "transform(s) government services and their delivery to citizens". However, the linkage of data points using datasets from different departments can also enable the creation of detailed profiles of citizens and violate their right to privacy.

The proposed sharing of different datasets amongst the various ministries and authorised data requesters must also preempt for harms from the use of community or population level insights from data. As recent scholarship highlights, in the current digital economy, an individual's data might not just impact their own privacy but impact the privacy of others as well.[17] Essentially, given how aggregated data generates insights, consensual sharing of data by some members of a group that have common traits (such as race, ethnicity, or even interests) can have a bearing on other members of the group who were not directly involved in the relationship.

[15] such as collection limitation, openness, security safeguards, etc; 'The OECD Privacy Framework' (2013), www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

[16] *Justice K.S Puttaswamy (Retd.) vs. Union of India* (2017) 10 SCC 1**.**

[17] Salome Viljoen, 'A Relational Theory For Data Governance' (2020) 131 The Yale Law Journal <https://www.yalelawjournal.org/feature/a-relational-theory-of-data-governance>.

While the NDG Policy intends to 'accelerate digital governance', the Policy must ensure that datasets are assessed for their robustness, accuracy, and the potential risks and harms of relying on them. This is particularly important as data-driven governance can disproportionately impact historically underrepresented communities. For instance, the formulation of policies on urbanisation that rely on non personal data could exclude the poorest sections of the society as they may not be adequately represented in the data collected.

4.3 Anonymisation standards

Unlike the proposed NPD Framework, this NDG Policy does not acknowledge the risk of re-identification which can be observed in the little detail it provides around anonymisation standards. A more specific concern arises from the lack of clarity provided by the Policy regarding the manner in which these anonymisation standards will interact with those to be set out by the DPB 2021 or even the proposed NPD Framework. Similar concerns around the interactions between these institutional structures and their overlaps have been highlighted in section 5.

To ensure that the standards set are of the highest security and take into account all privacy risks, the EU's Article 29 Data Protection Working Party in its 'Opinion on the Concept of Personal Data' warns that risks must be considered keeping in mind the 'state of the art technology at the time of processing', but also any possibilities of future development of technology during that period, must also be given due attention.[18] Given these challenges, we believe that the proposed Data Protection Authority under the DPB 2021 is best placed to effectively consult with all stakeholders and set relevant standards.[19]

[18] See, Art. 29 WP, 'Opinion 4/2007 on the Concept of Personal Data'

[19] In the current framing of draft law and policy in India, the DPAI might be better placed due to the statutory nature.

5. IDMO's regulatory structure

We welcome the Ministry's clarification on the roles and responsibilities of the IDMO and the inclusion of a redressal mechanism to address citizens' grievances.[20] However, there may be concerns on how the proposed redressal mechanism in its current form will be able to achieve its objectives.

Firstly, it may be important to determine a particular threshold of independence for the IDMO as it has oversight over access and sharing of non personal data that may include data from the government and private entities. The lack of clarity around the IDMO's composition - especially with regard to whether it will include technical experts and representation from civil society and industry - raises concerns regarding its independence. Inclusion of procedural safeguards and diverse representation from stakeholders within the NDG Policy will go a long way in ensuring independence. Relatedly, the Policy is silent on the minimum qualifications and technical expertise necessary for the Chief Digital Officers of the Digital Management Units of government departments and ministries.

Secondly, the Policy is also silent on the IDMO's interaction with other regulators and bodies such as the proposed Data Protection Authority, National Health Authority, and the Competition Commission of India. In the absence of clear provisions outlining the IDMO's structure and functioning, there may be overlaps in regulatory oversight between the various institutions.

Lastly, given the aggregated nature of these anonymised datasets, delineating an individual's data usage rights can be challenging.[21] The Policy does not lay out how individuals or communities will be notified of breaches and the avenues of recourse available to them. It also does not specify the type of grievances that can be raised and the process of redressal in the event any stakeholder is dissatisfied with the outcome of such a process.

[20] Paragraph 6.14 of the Draft National Data Governance Policy

[21] Paragraph 6.10 of the Policy stipulates that "IDMO may ensure that data usage rights along with permissioned purposes to be with the Data Principal". These rights however are not defined in the Policy.

6. User charges and access

Under paragraph 6.18 of the current proposal, IDMO is allowed to prescribe user charges/fees for its maintenance or services. This is a welcome move away from the commercialisation and monetisation of data mooted in the Draft India Data Accessibility and Use Policy. However, the NDG Policy does not provide a framework or define principles that the IDMO must rely upon when prescribing such user charges.[22] In developing the framework for these charges, the Policy should consider issues of access when it comes to user charges to avoid barriers for researchers seeking to rely on such datasets. Excessive user charges can also disproportionately skew access to datasets towards dominant players by overincentivising monetisation of datasets.

Paragraph 2.2 of the NDG Policy highlights the creation of datasets "to enable and catalyze vibrant AI and Data led research and Start-up ecosystem" as one of its objectives, and paragraph 6.8 requires the IDMO to notify rules on the provision of data to requesting entities and assess the nature of requests for data beyond the Open Data Portal.[23] However, the Policy does not sufficiently clarify the criteria that must be adopted by the IDMO in processing these requests. Similarly, while paragraph 6.13 of the Policy requires the IDMO to define the principles for ethical and fair use of data shared beyond the government ecosystem it falls short in stipulating penalties in the event of misuse or breach.

7. Data governance approaches in other jurisdictions

It is worth highlighting that various jurisdictions globally have developed or are in the process of developing data governance frameworks to promote the flow of data within and across different sectors.

The European Union, for instance, post the enactment of the GDPR, released the European Union Data Strategy to create a single digital market in Europe by facilitating

[22] Paragraph 6.18 of the Draft National Data Governance Policy

[23] While the Open Data Portal is mentioned in paragraph 6.8 of the Policy, it is not defined in the document.

access and reuse of data within the economy.[24] The Data Governance Act is one such legislative proposal under the EU Data Strategy that seeks to enable the safe reuse of specific categories of public-sector data.[25] In the United Kingdom, public sector data sharing was primarily facilitated through the Digital Economy Act 2017 and is now being further strengthened through the National Data Strategy. In Singapore, the Personal Data Protection Commission ('PDPC'), constituted under the Singapore Personal Data Protection Act 2012, has taken active steps to enable and encourage data sharing across sectors. In 2019, the PDPC released the *Trusted Data Sharing Framework* to provide guidance to the private sector on data sharing.[26] On the other hand, in Ghana, the National Information Technology Agency (NITA), in partnership with the Web Foundation (WF), launched the Ghana Open Data Initiative in 2012 to make government data available to the public for re-use. While these jurisdictions have taken approaches and intent to enable data sharing, all of these frameworks in these jurisdictions have been foregrounded by dedicated data protection legislation.

8. Conclusion

We would like to thank the MeitY once again for the opportunity to submit comments on the National Data Governance Framework Policy. While it is vital to harness the potential of non personal data in improving governance and delivery of services, such a governance framework must be built upon strong legislative, institutional, and technical foundations. We urge the Ministry to consider the issues and analysis articulated in this policy submission before enacting this Policy.

[24] European Commission, 'A European Strategy for Data' (European Commission 2020).

[25] COM/2020/767 Final Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)

[26] Personal Data Protection Commission, 'Trusted Data Sharing Framework' (Infocomm Media Development Authority and Personal Data Protection Commission 2019).