Symbiosis Law School, Nagpur Multidisciplinary Law Review
Volume II Issue II (2023), ISSN: 2583-1984 (Online), Pg. Nos. 196-202

National Data Governance Policy: A Prudent Plan Or A Premature Policy?

Jibisa Janvi Behera[*]
Zoya Farah Hussain[**]

ABSTRACT:

The trail of legislations through India Accessibility and Use Policy 2022 to National Data Governance Policy 2022, has given a bird-eye view of the explicit and implicit consequences of it. The rollback of the first draft on being accused of commercialization of data made the stage for the release of a new draft policy, open to suggestions. Considering the lack of appropriate data regulations, this article attempts to analyse the feasibility of the proposed legislation which falls back on the "guarantee of privacy" of "anonymization" and exudes a lack of discussion with the real stakeholders. A few solutions have been suggested for the identified problems which would not just encapsulate the practicality of the data market of India but would also analyse the technological stance of the country in the present scenario. A comparative evaluation is also done with data governance policies of other countries and an overall exposure to such technical measures has put forward the analysis of the said legislation.

***Keywords**: Commercialization of Data, Guarantee of Privacy, Anonymisation, Data Governance Policies, & Real Stakeholders.*

[*] Student, B.A. LLB. National Law University Odisha, Email- 22ba044@gmail.com.
[**] Student, BB.A. LLB. National Law University Odisha, Email- 22bba056@gmail.com.

NATIONAL DATA GOVERNANCE POLICY: A PRUDENT PLAN OR A PREMATURE POLICY?

Introduction

The “India Data Accessibility and Use Policy"[1] has been surfacing in the headlines for various reasons. The heart of the said legislative instrument lies with data or more precisely, “non-personal data". Its intended goal was to use public datasets made available by various government agencies and ministries to help citizens with product creation, service delivery, and governance. On May 26, 2022, the draft legislation was released for public comment. However, it was quickly replaced by the National Data Governance Policy, which permits the government to sell anonymized public datasets for surveys, and commercial value addition through analysis and investigation in the startup world, companies, etc. A superficial look would build a promising picture of the data market of India but a substantial indulgence in its technicalities brings forth certain inadequacies which are worth addressing. This article is divided into 2 parts. The first section would elaborate on the provisions of the draft, give a detailed analysis of the provisions, and would attempt to chalk out the loopholes and suggestions in response to those and the last section would provide the conclusion arrived at after the analysis.

Building the Baseline

One of the initial proposals of the draft envisages the formation of “The India Datasets platform, which will be composed of non-personal and anonymized datasets from Central government institutions that have gathered data from Indian nationals or those in India will be designed and managed by the India Data Management Office (IDMO)”[2]. A non-personal dataset is any collection of data that excludes information that may be used to identify a specific individual. Essentially, this means that using such data to identify any real person is difficult. Any non-personal data exchange by any entity may only occur on platforms that IDMO has chosen and approved to maintain security and trust. Once finalised, the policy will apply to all central government departments, along with all non-personal datasets, and will include any

[1] India Data Accessibility and Use Policy, 2022, Acts of Parliament, 2022 (India).

[2] National Data Governance Framework Policy (Draft), 2022, § 5.1, Acts of Parliament, 2022 (India).

relevant norms and regulations controlling start-ups' and academics' access to such datasets. The state governments will be "encouraged" to implement the policy's guidelines.[3]

The catch with this perfectly sounding independent body working for stability and privacy of the data obtained is that there has been no mention of any structure or pathway through which standards would be set for monitoring of data and opens a room for arbitrary treatment of data as there is no mention of a supervising body for it. Further, Section 5.2[4] of the Policy provides for the IDMO to hold at least 2 semi-annual consultations with representation of government and industry. However, the extent of participation of stakeholders and the way consultations will be structured is yet to be determined. The absence of established standards for anonymisation of data, specific qualifications for terming certain data to be "accessible" for the public and other such crucial functions of the said body has been provided for without it entailing any technological backing or administrative support.

Section 6.18[5] of the draft mentions the charge of user fees for the maintenance of IDMO. That, again, comes with loose strings when it remains silent on the mechanism, need and evaluation method of "charging" such fees which puts a question on one of the major objectives of the draft to increase the accessibility of data for public welfare.

Therefore, in absence of proper laws relating to storage, segregation and dissemination of data, IDMO should be made robust with a proper structure of operation with technical experts on table. A regulating body should be set up along with it which would make the process of denying someone access to data- public and would monitor the operations of IDMO. While IDMO should initially start functioning with some areas of the nation as a sample run after which its effect should be expanded proportionately.

Rethinking "Anonymisation"

The major reliance of the sharing of data is on "anonymisation". But the problem with that is that it does not guarantee individual privacy! So, this information, which is purportedly

[3] Sarvesh Mathi, *National Data Governance Policy: What is it and What are some concerns around it,* MEDIANAMA (Feb 2, 2023), https://www.medianama.com/2023/02/223-national-data-governance-policy-what-is-it-concerns-around-it/.

[4] National Data Governance Framework Policy (Draft), 2022, § 5.2, Acts of Parliament, 2022 (India).

[5] National Data Governance Framework Policy (Draft), 2022, § 6.18, Acts of Parliament, 2022 (India).

anonymous, might be helpful to obtain and analyse.[6] However, if it coexists with the processing of personal data, it puts people's privacy at danger. Organisations and governments "scrub" databases of personal data to exploit this information. They are allegedly left with a tonne of "non-personal data" that has been "anonymized" since it can no longer be used to identify a specific person or compromise that person's privacy. The datasets are then made accessible to the general population. Information security measures are simply reversible.[7] This can result in the "re-identification" or "deanonymization" of a dataset, disclosing the identity of a person or group of individuals while invading their privacy and subjecting them to a variety of damages. It is a very simple exercise for hostile actors who have received the appropriate training. These actors may purchase this data from the dark web with ease using cryptocurrencies like Bitcoin, which also provide them some kind of anonymity. For instance, the “BBC recently discovered that 80GB of NATO's secret security data was being sold online for 15 Bitcoin (about £273,000)”. Given the lack of a data protection regulation, Indian people have no recourse to defend themselves if anything similar occurs. Researchers looking at US census data from 2006 discovered that “combining just three demographic indicators—gender, zip code, and birth date”—could identify 63% of the population studied.

In its now-retracted instructions, the government offered a wide range of anonymization procedures that departments may adopt, possibly to lessen the privacy dangers associated with deanonymization. Strong anonymization methods can make it more difficult for malevolent actors to carry out their plans, but they may not always shield datasets against deanonymization. By considering de-identified data as personal information, using proper data protection practises including data minimization, limiting usage, access, and sharing, and implementing security measures like encryption, organisations can lessen the privacy concerns associated with re-identification. Therefore, simply because there is a chance of deanonymization, it doesn't imply we should ignore anonymization. It indicates that we must work to enact rules and standards that are more stringent. There's a purpose why India's pioneers gave census data sharing a high level of privacy. We now need to incorporate that same idea of privacy into how we handle data. Now, academics argue that we should have the

[6] . Samson Yoseph Esayas, *The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the ‘all or nothing’ approach*, Vol 6, ELJT.2, 4 (2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746831.

[7] AIHIK SUR, “*REGULATING NON-PERSONAL DATA: HOW THE FREE FLOW OF DATA MAKES ANONYMISATION HARDER”,* MEDIANAMA (FEB 28,2022), HTTPS://WWW.MEDIANAMA.COM/2022/02/223-NON-PERSONAL-DATA-ANONYMISATION/.

right to reasonable inference. The concept does away with the boundary between personal and non-personal data, saying that the user must be safeguarded whenever a data processor infers something about them based on their data and there is a chance that this might do them damage. However, for it to be put into practice, there must be clear definitions of personal and non-personal data, how they should be treated, and why such conditioning is necessary.

The Stakeholder Question

The draft Non-Personal Data Framework directed that all public and private entities should share data at an aggregate level for public purposes. Startups or companies will have to give data but it is undefined as to what kind of data they would give and even if they do, how would it be beneficial for the economy or the startup culture discussion is left unanswered and it seems it is completely based on an assumption without any material or statistical backing.[8] Data being one of the greatest assets of a business which helps it survive in the competitive market would be one of the least things it would be interested in sharing unless a strong policy of protection against data abuse and clear reason for public welfare is announced and that espouses infringement of IP of that organization. The data purported to aid in "public welfare" would demand curated data and a lot of start-ups lack established structures of data collection so collaborating with the government for the sharing of data would open room for high-cost compliance on their part which would compel unplanned costs for the businesses. There are certain small enterprises that fall within the category of a data business. There is no justification in such circumstances for them to dedicate more resources and financial support to constructing compliances. Making all of this "mandatory" for all businesses raises concerns because they have no reason to think that the data they have been collecting will be used for the intended purpose, and even if they do, there are no provisions in place for them to opt-out or refuse to participate in the program at all. If data requesters are required to utilise the information for the public good, does this mean they cannot use it for personal gain as well? Is there a system in place for holding them accountable if they don't use it for the welfare of the public? Some of these framework inadequacies give rise to worries.

Hence, solely opening data will not be enough. The policy should enable IP creation, which will provide an impetus for data distribution. There is a need for some defining, which should involve what kind of businesses would be the stakeholders in this. What would be the data

[8] Stakeholder Consultation by The Dialogue, *Data Governance Policies and their Impact on Startups,* The Dialogue (Aug 28, 2020), https://thedialogue.co/wp-content/uploads/2020/09/August-28-Startup-Event-Report-.pdf.

compliance mechanism for all of them or would it be taken from a case-to-case basis? What would be the incentive of sharing the data? What would guarantee proper use of data and availability of a supervisory body in case of non-compliance with the projected objective? Such inherent loopholes in the draft have been noticed which clearly indicate a lack of understanding of reality and discussion with the actual stakeholders in this arena.

Section 6.3[9] mentions about "India Datasets Program in which IDMO will enable and build the India Datasets program, which will consist of non-personal and anonymized datasets from the Government entities that have collected data from Indian citizens or those in India. Private entities will be encouraged to share such data." Non-personal data are those that do not directly relate to a living individual (for example, weather, supply chains, census data, tax receipts, company-generated data, land records, vehicle registrations, traffic challans, etc.). There is no information that can be used to identify a specific person. However, there are a number of issues that were also covered in the prior report on the "Non-Personal Data Governance Framework," including the fact that even anonymized personal information can cause harm because the original data is still available and may be used for a number of illegal activities, the risk of re-identification for certain non-personal data, such as caste and tribe information, and even information pertaining to national security or strategic planning.

Duplication of Work?

Non-personal data, such as weather reports, surveys, market research, data from various organisations, demographic data, crash reports, data gathered from IP addresses, or other data, are already stored and effectively distributed by their respective departments; therefore, establishing an additional data repository in the form of IDMO would complicate the process of data usage and would also violate the IP protection of some fields carrying the same data.[10] In addition, there is a false assumption that there is a clear-cut distinction between personal and non-personal data. This is because there are specific categories of non-personal data that, due to multiple layering, are of that nature, so demystifying those levels can turn it back into personal data, and this would require case-by-case handling of the data rather than a strait jacket

[9] National Data Governance Framework Policy (Draft), 2022, § 6.3, Acts of Parliament, 2022 (India).

[10] Lokesh Choudhary, *what to expect in the Draft National Data Governance Policy 2022,* Analytics India Magazine (Sept. 13, 2022), https://analyticsindiamag.com/what-to-expect-in-the-draft-national-data-governance-framework-policy-2022/.

formula. However, on the plus side, it would be a good choice for fields or ministries that are having trouble deciding who should be the handling authority for vague categories of data.

To increase transparency and make government and private data that are raw more accessible to the public and businesses, this data policy proposes anonymizing this data and making it available, particularly for research and development purposes. However, there are many important data produced by different departments- health, electronic, road and waterways, IT, agriculture, finance, etc which are valuable in their own areas, which help in establishing their own ways of implementation in divergent fields, but the question arises as to which data is more valuable and less valuable, there is neither proper licensing mechanism nor an appropriate data valuation tool to categorize the importance of each of the data, in this process, private companies can take advantage to demand the data at a lower cost based on the political party funding as the data which might be important to government, might not be so for the private companies and vice versa.

An adversary can easily identify the record or small sets of records if they have some understanding of a subscriber, which may also raise privacy concerns for people and their personal information. The likelihood of re-identification of anonymized data from data sets that provide coarse credit card metadata is high

Conclusion

The Ministry of Electronic and Information Technology is currently drafting the “India Data Accessibility and Use Policy” with the goal of releasing the policy draft for stakeholder consultation. They are seeking input from all sectors, including research, academia, startups, and industry, on how the policy can institutionalize a data-sharing framework for the next decade. In terms of future accessibility and transparency, the ministry has made progress. India's ambitions to become a $5 trillion digital economy depend on its ability to harness the value of data, but the draft is not without its complications; the policy needs revision, ideally administered by a panel of experts and through meaningful public discussion, to achieve the stated goals. Furthermore, it is essential that such a policy wait until strong data protection legislation has been established before being released.