The Dialogue™

INFORM ENGAGE IDEATE

SUBMISSION ON THE REPORT BY THE COMMITTEE OF EXPERTS ON NON-PERSONAL DATA GOVERNANCE FRAMEWORK

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

Authored by: Karthik Venkatesh[1], Ayush Tripathi[2] and Harsh Bajpai[3]

Editor: Kazim Rizvi[4]

Cover Illustration: Abhinav Kashyap

[1] Research Coordinator, The Dialogue,
[2] Policy Research Associate, The Dialogue
[3] PhD Scholar, Durham University, UK
[4] Founding Director, The Dialogue

TABLE OF CONTENTS

[I] Executive Summary

The Dialogue appreciates the efforts made by the committee of experts on coming up with the non-personal data governance framework. This report is an unprecedented effort towards discussing non-personal data (NPD) in the country. The Report emphasizes on leveraging data in a data economy for social, economic, sovereign purposes. It rightly identifies India as the largest user of smartphones in the world and consequently the "leading consumer market in the world, and data markets for the foreseeable future." Noting that the digital economy is producing greater opportunities to market participants to innovate, the Report talks about imbalances in the market. While there is merit in the argument of using data for the growth of the digital economy, there needs to be clear processes in place, and regulatory checks and balances to allow for the free flow of NPD. The Report builds a case for the regulation of data by referring to its economic and social value and refers to the harms it could cause to the subjects, in the absence of regulation. It also speaks of the concept of collective privacy of vulnerable groups and the harms that could befall them due to exposure or handling of data.

However, given the fact that there is very little jurisprudence developed in the world on this form of data, anything that India comes up with will set a benchmark for the world. Therefore, it becomes important that India sets a right example in this sphere and this report is a welcome step towards this direction, however, addressing certain gaps will ensure a better framework.

First and foremost, there is a need for a voluntary data sharing mechanism rather than mandatory sharing of data. It needs to be kept in mind that processing of non-personal data and the insights generated from it gives a competitive advantage to the companies. Making all the data publicly available will discourage the companies to even collect such forms of data. Indian data businesses themselves will hesitate in processing NPD if ultimately they have to give it away. Further, a data sharing framework should be based on the principles of necessity, proportionality, accountability and transparency. Without these principles in place, no data custodian will be willing to give their data.

Secondly, the definitions under non-personal data must be relooked. The concept of community data must be formulated cautiously as there seems to be significant overlap with private and public

data. Given the responsibilities of data trustees, there is a need for clarity on the eligibility criteria as well as the rights and liabilities of the trustees. The report does not go into the depth of roles and responsibilities of data trust which is supposed to play one of the most crucial parts in the ecosystem. It is important that this trust runs independent and has robust transparency and accountability measures.

Thirdly, the report also talks about ownership of data, however it does not identify the Intellectual Property Rights issue involved with respect to the data sets. The concept of beneficial ownership has been introduced where communities can exercise their rights through a data trustee. The concept will create multiple communities exercising their rights through different data trustees which would increase the complexity. Communities exercising their rights over NPD would also inhibit the ability of the enterprises, including Small and Medium Enterprises and start-ups, to use the data they have invested to collect and curate.

Fourthly, the creation of a separate regulator for NPD seems unnecessary. It will result in regulatory overlap, creating legal uncertainty and adversely impacting the business and investment climate of India. The intersectoral clashes are bound to happen resulting in jurisdictional issues and litigation. Rather than a new regulator, a separate wing for NPD should be made in the Data Protection Authority created under the Personal Data Protection bill which would only overview the compliances and standards of anonymisation. The present submission also incorporates a framework for anonymisation standards which recommends a privacy by design principle for anonymised data sets and valid consent norms.

[II] Key Asks

- Definitions of the key roles must be articulated with clarity to demarcate how a data trustee would be different from custodian. A future regulation must flesh out how the conflicts of interest would be resolved and the oversight mechanisms must be established.
- Inclusion of "Private NPD" within the scope of mandatory data sharing could be disruptive to existing business models and weaken IP protections accorded. This would create second order issues including reduced investment and competition.
- Ownership, the corresponding rights and duties must be established to ensure accountability of authorities within the framework.
- Government is one of the largest custodians of data. To address issues of information asymmetry and alleviate fears of surveillance, a surveillance law reform is essential.
- Considering the widespread overlaps with the PDP framework and other frameworks of law such as the competition and intellectual property, it would be effective to mandate MoUs between regulators and pre-empt the clashes.
- A relaxed regulatory regime that allows data sharing on a voluntary basis could be envisaged as jurisprudence on data regulation is still nascent.
- Accountability of regulators must be to the Parliament and not to particular ministries, to reduce conflict of interest and to ensure independence.
- A code of anonymisation should be developed to mitigate the risk of re-identification and lack of standard anonymisation tools.

[III] List of Abbreviations

AI/ML	Artificial Intelligence/ Machine Learning
API	Application Programming Interface
CCI	Competition Commission of India
CoE	Committee of Experts
DPA	Data Protection Authority
IP	Intellectual Property
IRDAI	Insurance Regulatory and Development Authority
NDSAP	National Data Sharing and Accessibility Policy
NPD	Non Personal Data
NPDA	Non Personal Data Authority
PDP	Personal Data Protection Bill, 2019
RTI	Right to Information
SPD	Sensitive Personal Data
TRAI	Telecom Regulatory Authority of India

[IV] Governance of Non Personal Data: Key Principles

A. Privacy respecting data governance regime: Interplay between PDP and NPD frameworks

Data about one person allows for inferences about other people, creating a collective dimension where decisions about data-sharing affect not only the individual(s) involved in the data generation.[5] The report deals with the concept of "collective privacy", which aims to enhance the modes in which persons can take control of various data flows that pertain to them as a collective. However, setting up of institutional infrastructures such as data trusts need further deliberation, along with the key roles that the various entities will play. Moreover, technical standards such as anonymisation, which is aimed at protecting privacy of individuals must be framed in consonance with the obligations under the Personal Data Protection Bill, 2019.

When it comes to non-personal data, privacy concerns arise when either a person is identified while processing the data or when a data is deanonymised. The data sharing protocols between the government and the private entities or between two private entities must conform to the principles of privacy. It is important that the tenets of the Puttaswamy judgement[6] are followed, when the matter relates to persons who are identified from combining various data points/ data is deanonymised.

Similarly, private entities also have to keep in mind that anonymised data comes with the risk of re-identifications. There is evolving scholarship that claims that no anonymisation technique is absolute[7]. The development of cryptography has increased the re-identification risks. It is important that anonymised data must not be made subject to deanonymization that could harm the privacy of individuals whose data has been anonymised. In order to mitigate that, a code of anonymisation should be developed. This code of anonymisation should entail the techniques that

[5] Mantelero, A., 'Personal Data For Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection', Computer Law & Security Review, 32(2), 2016, pp. 238-255
[6] Justice K.S. Puttaswamy v Union of India, (2017) 10 SCC 1
[7] Bourdillon, S. & Knight A., Anonymous Data v. Personal Data - False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data, 34 Wis. Int'l L.J. 284 (2016-2017)

should be followed by the companies, consent mechanism and data audits which will help in making the non-personal data framework a privacy respecting regime.

A privacy respecting regime would help India garner foreign investments and collaborate with other liberal countries. Europe, for example, currently has the highest respect for privacy in terms of personal as well as non-personal data. India has to improve its adequacy standards in order to do business with these countries. Given the boost to the data market, it would mean that a strong privacy regime will prove as an incentive for the countries and companies to invest.

B. Voluntary model of sharing

The main goal is to incentivize sharing, to keep data related markets open for competition and to foster innovation. Therefore, it would seem that a structure of relatively flexible rules aiming more at equal treatment of all involved stakeholders rather than mandatory access rules seems appropriate.[8] Rules or guidelines from authorities such as the proposed regulator may be helpful to establish best practices and avoid behavior by the stakeholders that may be detrimental to the development of the industry and new business models. In this sense, it is important to create a culture of cross sector information sharing based on a voluntary model. In order to check the issues regarding monopolies, the answer to that lies in Competition Law reform. A contractual model of sharing of data between interested parties can be looked at for ensuring that the companies/businesses retain their right to exercise their choices in the digital economy.

C. Role of Data in data intensive business models

One of the key elements for the success of some of the major companies around the world is how they analyse and derive data. This helps them target their consumers and accordingly modify their business to suit their consumers. When it comes to data, the sheer volume is not the sole indicator of the economic value. Most of the data is unstructured and exhaust data and as long as this data

[8] Richter, H., Slowinski, P.R. The Data Sharing Economy: On the Emergence of New Intermediaries. IIC 50, 4–29 (2019). Available at: https://doi.org/10.1007/s40319-018-00777-7

is inaccessible for analysis, its potential remains unrealised.[9] It is the potential of a company to turn this data into a useful insight that creates the economic value of such data. However, these insights are also what gives companies a competitive advantage in the market and creates a right on these data.

Data collected is never neutral and it is highly dependent on the context in which it is collected.[10]Data is ultimately just an abstraction that is shaped by the context in which it is created and for what it's created for. Not all datasets can be used in multiple settings, because the impact of blindly adopting without questioning the ends for which such data will be used can create unintended consequences. Especially with the advent of AI systems with training datasets being a vital resource for scaling this, it's important to understand the norms these data sources codify before commoditising data for advancing the digital economy. That being said, data and insights are not natural resources[11] that are available for everyone to exploit. It involves investment in terms of collection, processing and maintenance costs. Tech companies invest heavily in data science to make use of collected data. That being said, it is important to acknowledge the role and incentivise the companies to continue generating such data without creating a free ride market for all. In the information age, the driving force for any business is insights and intelligence. In an attempt to create value out, we must ensure we do not set back the development and innovation.

This data is as valuable for a small company as much as it is for a tech giant. It needs to be kept in mind that in an effort to compel tech giants to give their data, we do not end up harming our own start-up ecosystem. Moreover, the value that can be assigned to datasets is not clear and is yet to be formalised.

[9]The Economic value of Data: Discussion Paper, HM Treasury UK. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/731349/20180730_HMT_Discussion_Paper_-_The_Economic_Value_of_Data.pdf

[10] Wu, A. X., & Taneja, H. (2020). Platform enclosure of human behavior and its measurement: Using Behavioral trace data against platform episteme.New Media & Society.

[11] Kazim Rizvi., Karthik Venkatesh. (2020). Why the non personal data governance framework needs a rethink. Financial Express. Available at: https://www.financialexpress.com/opinion/why-non-personal-data-governance-framework-needs-a-rethink/2069892/

D. Intellectual Property Regime

In the form proposed today, there are considerable conflicts that can be anticipated with the existing IPR regime. Datasets enjoy copyright protection under the current national laws and also international conventions that India is a signatory to. Having said that, there are widespread concerns that a mandatory sharing regime and democratizing data might create market disruption. Many companies that hold these insights and data for a competitive edge might lose out and this could create an adverse economic impact.

Moreover, the IP regime already contains provisions for sharing of protected information through licenses and other mechanisms. It might be prudent to explore enabling sharing data and key resources through existing frameworks and provide incentives for sharing of such data. When public interest is involved, the access to information must still follow the course of law that is already in place. Moreover, conflicting frameworks paves way to confusion and disputes in the market. IP rights protect the interests of the companies with regard to key aspects of their business. It allows the companies to innovate without fearing having to part with protected elements in their business model. Besides copyrighted databases, insights can be part of confidential information and trade secrets. By allowing a data-for-all approach, this trade secret/confidential information could get into the hands of the competitors in the market forcing businesses to part away with deeply researched insights which will stifle innovation in the digital economy.

E. Surveillance concerns

Most products or services that we avail rapidly and constantly generate data that forms the basis for future development. In such a context, the policy approach and design must set up a mechanism that can guide the market and prevent the citizens from unwarranted harm. However, the report seems to raise alarm bells with regard to unwarranted surveillance. The need for a surveillance law cannot be understated.

There is a fundamental shift from the PDP bill which starts with the protection of individual rights and serves as a protection framework. The NPD report needs further engagement with the individual rights and community rights, as the state is granted increased access.

The report makes a swift dive into government access to data and industry access to data. This is premised on the assumption of marginal good to communities and the public as a whole. As it stands today, the intelligence agencies are not created by an Act of Parliament. This takes away an effective oversight mechanism through accountability to the Parliament. This creates considerable fears and apprehension on the usage of data, purpose of such use etc. The enforcement of law and development of law in this regard is extremely minimalistic and is far away from global standards and best practices. The focus should be on the fallacy of dissociating NPD and surveillance. Certain intelligence operations around the world start with de-anonymising data by combining various data points. The resultant data, which could be extremely sensitive and personal could be misused.

The Government is one of the largest custodians of data. This in turn leads to information asymmetry between the state and its people. In addition, by allowing backdoors for government access, it accentuates the fears of surveillance. This power dynamic leads to a trust deficit in the absence of restrictions on the state exercise of power and lack of accountability mechanisms.

Similar arguments were brought to fore during the discussions surrounding the PDP bill, particularly on Clause 35 of the bill that grants widespread exemptions for the Government. It was opined that, in both cases effective checks and balances on the government control over individual and private lives is a crucial facet enforcing citizen trust in the system. The report leaves a blank slate on this front, and the need for a surveillance law is more eminent than ever.

[V] Chapter wise response

A. Definition of Non Personal Data And Key Roles

This section of the report lays down the definition of Non Personal Data (**"NPD"**) and creates roles in the ecosystem. The committee recommends that the NPD be categorised into three categories and four key roles should be created in the ecosystem. The committee also recommends the concept of sensitive non-personal data and mandates that consent must be given by the Data Principal for the use of anonymised data sets.

1. Definition of NPD

The committee states that any data that is not related to an identified or identifiable natural person, or is Personal Data ("**PD**") that has been anonymised is Non-Personal Data. This data, the committee recommends, should be further classified into three categories i.e public NPD, private NPD and community NPD.

The report defines Public NPD collected or generated by the government or any agency of the government or data collected or generated in the course of executing all publicly funded works. This definition also carves out an exception with respect to the NPD collected or generated by the government which acquires the status of confidential treatment under any law.

It is important that status for any data collected by the institution under the Public NPD domain should be at par with the Article 12 of the Constitution of India and be subject to writ jurisdiction of the court. Further, it is necessary that any NPD so collected by the institutions mentioned in the definition should come under the ambit of RTI Act. The exception carved out in the definition should have checks and balances in case of an executive order which has a force of law. The executive order should be well reasoned backed by a law and must be put in the public domain. Further, the report needs to elaborate on the publicly funded works. The report does not clarify the threshold for a work to be called as publicly funded. This could create confusion in case of public private partnerships.

The report further goes on to define community NPD as the data whose source is the "community". It defines a community as any group of people that are bound by common interests and purposes, and involved in social and/or economic interactions. The definition includes geographic community, a community by life, livelihood, economic interactions or other social interests and objectives, and/or an entirely virtual community.

It is worth mentioning here that the concept of community NPD has not been recognized anywhere in the world. Further, the way in which the term community has been defined, it poses chances for significant overlap with the public and private data. For instance, the Report states that Community NPD includes data collected by municipal bodies, which would also fall within the definition of Public NPD. Similarly, many NPD data sets relating to groups of customers including from virtual communities (such as social media users) form part of Community NPD. Currently, the definitions are separated by thin lines and the fundamental base for their segregation is itself confusing, which could later prove to be an issue. Further, it also mentions that raw/factual data without processing may be characterised as the community data. **The report does delve into the fact that all raw data should also be considered private data**. In the case of *BanxCorp v. Costco Wholesale Corp*[12], a US court said that *"when confronted with raw data that have been converted into a final value through the use of a formula, courts should put significant weight on the degree of consensus and objectivity that attaches to the formula to determine whether the final value is fundamentally a 'fact.'"*. Therefore, if the raw data attaches significant value to the final result, it must be protected under copyright.

The report also mentions that the NPD collected or generated by the government / agency of the government will amount to public NPD except data that is explicitly treated as confidential under any law. However, confidentiality obligations in the commercial context are largely driven by contractual obligations. Without a clarification that information that is contractually confidential is an exception to public NPD, any information provided to any government department or agency in the course of collaboration with the government by a private entity or individual is also at risk

[12] 978 F Supp (2d) 280 (SDNY 2013)

of being categorized as public NPD. This could disincentivize skilled private players from bidding for government projects.

The report further recommends that in order to minimise the risk of re-identification, the consent of data principles must be obtained for anonymization and usage of the anonymized data at the time of obtaining their consent for processing their personal data. The consequence of this recommendation would be that the notice given under the PDP Bill will also have to include a specific provision to obtain consent for anonymization and the use of such anonymized data, even though in certain cases where the NPD is not based on the personally identifiable data, obtaining consent would be irrelevant.

Further, the Report defines private NPD as NPD "*collected or produced by persons or entities other than the governments, the source or subject of which relates to assets and processes that are privately-owned by such persons or entities, and includes those aspects of derived and observed data that result from private effort, insights involving application of algorithms or proprietary knowledge, and data included in a global dataset and which is collected in foreign jurisdictions"*.

This would mean that all Indian data custodians, including those collecting NPD of foreign individuals, and all foreign data custodians collecting NPD of Indians would be subject to these regulations. Therefore, the presence of Indian NPD in a dataset will invoke applicability of the NPD regulations, even for datasets created by foreign entities which largely contain NPD of foreign individuals. This would also act as a strong deterrent for several service providers that operate at a global level, or even smaller players from offering their services to customers/ users based in India, since even any incidental collection of NPD of users in India is likely to subject the entire dataset(s) of such organisations to the proposed requirements and obligations.

Additional point of concern is also that The PDP Bill contemplates that the Data Protection Authority (DPA) would stipulate standards of irreversibility for personal data to be considered anonymized. Therefore, if the NPD regulations also stipulate standards for anonymization, **there is a possibility of conflict / overlap in the event if these regulations are not harmonized.** This may lead to uncertainty for businesses and increased compliance costs.

Recommendations: (a) The committee should revisit the definition of community data and prevent any overlap of it with private data.

(b) The anonymisation standards and consent mechanism should be dealt with by DPA under the PDP Bill.

2. Sensitive Non Personal Data

The committee has also introduced a new concept of Sensitive NPD similar to the lines of sensitive personal data ("SPD") which finds its place in the PDP Bill, 2019. The committee recognises the following perspectives from which NPD can be considered sensitive:

1) It relates to national security or strategic interests;
2) It bears risk of collective harm to a group
3) It is business sensitive or confidential information;
4) It is anonymised data, that bears a risk of re-identification

The Report mentions that NPD inherits its sensitivity from either the underlying sensitive personal information, or based on collective harm it imposes. The report further mentions that with regard to the anonymised data, it derives the sensitivity of NPD from Clause 36 of the PDP Bill which provides the exhaustive list of SPD. However, unlike the PDP Bill, this Report **does not provide any information with regard to the additional measures that should be taken to protect Sensitive NPD.**

It is also pertinent to note here that this principle is only going to help in cases where the data can clearly be categorised as SPD. We believe that classification of sensitive NPD should not solely rely on the presence of SPD because in some instances there is a possibility that certain NPD can be sensitive even though there are no underlying SPD datasets. The **provision in its current form creates ambiguity and uncertainty, and no mention is provided on the nature of obligations on bodies when it comes to handling these classes of NPD**. There needs to be clearer definitions of the treatment of NPD derived from "critical personal data" as defined under the PDP. Moreover, clearer articulation of rights and obligations related to this is vital.

Recommendations

(a) Committee should remove the concept sensitive non-personal data from the framework.

(b) In arguendo, if the committee is willing to keep it, the nature of obligations and additional measures must be defined in order to protect such data.

Further, in this case scenario of sensitive personal data, sharing mechanisms should also be robust and appropriate framework must be developed to reduce malicious use.

3. Roles In The NPD Ecosystem

The committee report identifies three key stakeholders in the NPD Ecosystem i.e. Data Principal, Data Custodian and Data Trustee. When public and private NPD is concerned, data Principal is the person to which that NPD relates and as far as Community NPD is concerned, data principal would be the community that is the source of such data. The report also recommends the constitution of Data Trusts which will act as the institutional structures in the ecosystem.

a. Data Custodian

Data Custodian is somewhat similar to Data Fiduciary in the PDP Bill and undertakes collection, storage, processing and use of NPD. They have a duty of care to the individual or community from which NPD has been collected. It is pertinent to note here that the standard for ensuring the 'best interest' of the Data Principal is presently unclear, and the Committee has recommended that it should be detailed in the NPD Legislation. Though they have indicated that such standards may relate to anonymisation techniques and graduated data sharing.

Data custodians have a duty of care, in respect of community NPD, to the concerned community. Such duty of care will be specified in the proposed NPD law and will include anonymisation requirements, protocols for data sharing etc. The Report provides that 'duty of care' must be operationalised through the best interest standard in order to prevent harm to communities and individuals from processing NPD. Businesses may make data sharing requests to the data custodian. If the data custodian refuses to share the data, the request can be made to the NPDA and if the NPDA determines that the request is genuine and beneficial it can require the data custodian to share the raw/factual data. Equating the obligations of a data fiduciary under the PDP Bill which relates to protection of personally identifiable data to that of obligations of entities that

process NPD is not justified since the nature of data being processed is different. NPD may be used for business development or market strategies which do not directly benefit the data principal and may therefore hinder the ability to process such data effectively. Further, the standard for ensuring the "best interest" of the data principal is presently unclear, and the Committee has recommended that this be detailed in the NPD regulation.

b. Data Trustee

Data Trustee is the person through which a community exercises its data rights and who takes action to protect the community against any collective harm arising from the use of Community NPD. **However, there is no clarity on how a data trustee would be identified and the eligibility criteria for such an entity,** if any. Due to this, there are high chances of multiple trustees exercising rights over the same NPD and in order to mitigate it, the committee has introduced the concept of Policy Switch to manage regulations over NPD wherein the trustees can issue regulatory guidelines. However, it fails to take into consideration that this could lead to multiplicity of regulations given that there could be different Data Trustees for the same NPD.

There is a need for deliberation on the concept of data trustees and data custodians. Clearer prescription on who can be eligible to be a trustee needs to be achieved along with the rights and liabilities needed to be arrived at. It is also important to understand the role of the government in this ecosystem. The Report points towards the government taking on the role of both the trustee and the custodian. Since they are both expected to do completely different functions, and the trustee enables the “collective bargaining function” of data principal communities, there may be a conflict of interest. Moreover, accountability mechanisms and transparency requirements for these players in the NPD ecosystem must be prescribed.

c. Data Trusts

The report put forward the establishment of data trust which will be governed by certain rules and protocols. The aim of this trust is to store data from various sources, custodians etc. **The role of data trust is not defined in the report.** It only mentions that all the data collected through either mandatory or voluntary purposes will be stored here and can be accessed. The report fails to mention as to who will access such data and what would be the eligibility criteria for accessing

such data. **In order to ensure transparency, rather than public authorities controlling the trust, it must be operationalised in an independent manner in order to ensure maximum privacy and security of the data.** The report has to clarify the sharing mechanism for the data trust.

It is critical that data trust is independent. While no framework has been provided on the functioning of data trusts, it is important that four key elements are observed in the data trusts[13]:

1. **Privacy and Protection:** The data trust must adopt mechanisms to ensure the privacy and protection of the datasets which are contributed to the trust. This trust must have the highest security standards in order to prevent any cyber attack.
2. The data trust **must protect the commercial confidentiality** of the data providers.
3. Some aspects of data including curation, insights etc are protected under the relevant IP laws. In such a case, the data trust will need to secure appropriate licences from rights owners and to ensure that the terms of those licences are complied with.

It is also important that data trust observe a balance between the rights and interests across both participants and wider stakeholders and generate maximum trust to ensure the proper conduct of the activities of this trust. This trust should provide adequate representation from all the stakeholders, a mechanism for agreeing changes to the rules and protocols of the data sets and develop an independent oversight on the function of the trust. It must be kept in mind that the overriding aim of the governance structure is to achieve trust.

Recommendations:

(a) The duty of care and best interest comes when there is a chance of re-identification or the identification of a natural person while mixing several datasets. It is important that the committee must define the standards to maintain best interest of the data principals.

(b) The Data Trustee has a very important role in the ecosystem. However, there is no clarity on how a data trustee would be identified and the eligibility criteria for such an entity. The

[13] Masons P., et.al., (2019) Data trust: Legal and Governance Considerations, Retrieved From: https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf

committee must define the rights and liabilities of data trustees. Further, the committee must revisit the conflict of interest that arises when the government plays the role of data custodian as well as data trustee.

(c) The concept of data trust must be revisited. There is a little to no information about the functioning, security and management of data trust which will store the data. Further, mandatory obligation to provide data to the trust must not be recommended. The trust must function in accordance with the standards of privacy and security as well as it should have the representation of the ecosystem. The operation of the trust should be independent to the government and if the committee wants to go ahead with forming data trust, it must be transparent and accountable.

B. Ownership of Data

Committee aims to articulate a legal basis for establishing rights over NPD. The term ownership implies a set of economic and other statutory rights. Since data is an intangible asset, the committee recognizes that many actors may have simultaneous overlapping rights and privileges. The Committee has introduced the principles of 'best interest' and 'beneficial interest' that will aid in tackling this overlap in ownership of data with respect to Community NPD and Private NPD. In this context, it is pertinent to unpack the three main threads that comes out from such a framing of policies.

Committee has not laid out the decision making processes the Data Trustee will follow while acting in the "best interest of the community". It is left to be seen how the aims and objects of the data trust would be drafted, and how this would be operationalized. In cases of a dispute based on individual interest clashing with community interest, there needs to be clearer articulation on the balance the trustee will strike while acting in their capacity. It is important to note how decisions on "beneficial interest" would be arrived at. It could entail public consultations, and community participation- the modalities of which are to be finalized based on the Indian context.

1. Data Commons

Every community from which data is taken from was to be considered as data commons, where the communities can exercise certain rights and privileges over. However, that articulation of the contours of the rights that communities will accrue will take a long time to take shape. From a legal standpoint, unless there is absolute clarity over data ownership, it will be extremely difficult to allocate any rights and/or duties. Given the fact that community is an amorphous entity, it is harder to ascertain specific boundaries. For example, a community under the framework, does not take into account the various subsets under each community; many of which overlap with one another. Additionally, it's important to determine the criteria based on which such definitions are employed. In case of a geographical criteria, the consequence of someone moving from one place to another has not been iterated.

It is important to analyse and demarcate community rights, and data principal rights over data. At what point does one's community right trump the individual right, if at all it does? Privacy is an

individual right, community rights over data is a new set of rights that the framers are aiming to establish.

3. Data Sovereignty

Given the protectionist approach towards data governance in recent policy proposals, it is causing considerable concerns. It raises concerns regarding unwarranted state surveillance through data. In the absence of a strong surveillance framework in India, the narrative of data as a national resource and unfettered access could lead to considerable risks of mass surveillance.

The Report states that Private NPD may "include such data in a global dataset that pertains to non-Indians and which is collected in foreign jurisdictions (other than India)". The extension of any rule to the data of foreigners which is collected in a different jurisdiction would be a serious overreach of the application of India law. Extending regulatory powers to a domestic regulator over all data held by private companies tantamounts to regulatory overreach and must be avoided.

4. Asserting Intellectual Property rights over data

There is a push for mandatory sharing of raw/factual anonymised data, which could be in some cases protected under certain IP laws. As it stands today, datasets fall under "literary work" under the Copyright Act, 1957. Datasets tend to satisfy the tests of 'modicum of creativity' and the sweat of the brow' theories.[14] With the key roles in place, it is unclear if there will be a clash with ownership of intellectual property rights over datasets. Having said this, it is possible that there will be considerable litigation when it comes to claims over data, especially when mandatory sharing is envisaged. It is prudent to revisit clauses that specify sharing of raw data as they might be protected under copyright law if they pass the test of "modicum of creativity".

It is pertinent to note that there is considerable investment and resources spent on maintaining datasets, collecting data in the format in which it is to be stored etc. Organizations incur costs in terms of data collection, data preparation, data storage, data security, and ensuring data quality. Infrastructure would also be required to facilitate secure transfer of the data. Besides the type of

[14] Shyam Lal Paharia v. Gaya Prasad Gupta Rasal, 1971 A.I.R. 58 (All) 192, 195, 199;

data being shared, compensation should be linked to frequency of access, the effort and investment incurred by the Data Business to make the data available, and also the opportunity costs to account for a fair return on investment.

Moreover, if data sharing is made mandatory, including private proprietary data- India could be violating its obligations under international agreements and treaties such as TRIPS. While the report mentions the possibility of compensation that could be provided to the entity sharing such data, establishing a data market is a scary precedent. If at all, a well regulated market can be created for this purpose, it is left to be seen how valuation of data would be done. It was pointed out that the value of datasets for businesses is extremely contextual.

5. Beneficial Ownership

In case of community data, allocation of primary economic and other statutory rights over the data is to be operationalised through the concept of "beneficial ownership/interest". This is done with the aim of protecting the community's interest in mind and creates many implementational and conceptual roadblocks. It comes down to how communities would be demarcated, how the overlaps would be taken care of. For example, a person could be part of two communities, in such a case who would have access to the data? It is critical to conceptually clarify the basis for community ownership and the exact grounds for which the right extends. Here, the data trustees play a huge role towards furthering data rights, since they are holding "beneficial ownership" of the community data. The principle of 'beneficial interest' necessarily requires decisions to be taken on the basis of the needs of the community as a whole, and therefore may be used for purposes that certain members of the community may not agree with. The contours of rights of the members of the community, along with the obligation of the data trustee need to be chalked out for any future regulation.

Recommendations

(a) We believe that a strong framework of responsibilities of custodians, and trustees must be arrived at to ensure accountability and transparency, it must include clear provisions regarding disputes, its resolution, remedies in case of breach, reporting of breaches of duty

by the trustee acting on behalf of the community, and oversight mechanisms that would ensure smooth functioning.

(b) IPR protection to these data sets must be given in order to enhance competitiveness and protect the rights of the companies. It needs to be kept in mind that TRIPS obligations have to be observed. Denying IP protection would hamper investments as well as hamper the Indian data ecosystem.

(c) The rights of the members of the community, along with the obligation of the data trustee need to be chalked out for any future regulation.

C. Data Businesses

The Report defines a new horizontal category namely," data business". This would include any commercial, government, or non-government entity that processes or manages data beyond a certain data-related threshold. The NPDA would determine the "data related threshold" which may vary from time to time. This would apply to government and other non-government organizations that collect, process or otherwise manage data as well.

The meta-data about data being collected, stored, and processed by Data Businesses will be stored digitally in metadata directories in India and would be open access. While allowing open access could create opportunities for businesses, security of the metadata and prescriptions on security protocols are yet to be finalised. Moreover, the economic impact on companies to maintain an open access repository of all meta data directories would create excessive compliance cost, without sufficient incentives in place. Indian citizens and India-based organizations will have open access to the metadata of the data collected by different Data Businesses and Government.

In the digital age, with most businesses being online, the majority of them would-be classified as data businesses. The " data related threshold" that is set out as a criterion for the registration of a data business needs further deliberation. It has not been defined in the report, and is left for the authority to determine. This would play a massive role in effective implementation. It is important to ensure that such a mechanism does not disincentive Data Businesses from processing NPD, as this would defeat the Committee's aim of leveraging NPD collected by data-intensive businesses for innovation, and social and economic development.

This is first of its kind move and has not been seen in any jurisdictions. It raises concern over the compliance burden that will be increased. Given the fact that the majority of the start-ups are in the data businesses, this move would put additional burden than what this framework already imposes. Mandatory data disclosures would disincentive the startups from even processing such data.

Apart from the compliance burden, there is no clarification on where the metadata collected from the companies will be stored and who will have access to such data. It is important to secure the

privacy and security of the data disclosures made under the report. Again, access to such data raises concerns for the private companies to which that data belongs to. India has to create the necessary infrastructure to protect the data obtained and given the sheer magnitude of the data that would come from this mandatory disclosure of the majority of companies, the standards for protection of these meta-deta directories must be very high.

Recommendations

(a) Majority of the businesses in India and the world are data intensive. Every company processes certain data to either improve their company or the services they offer to their consumers. Most of the businesses would become a data business in some form or the other. Therefore, the basis of this whole chapter is on the threshold that is to be determined from time to time. A low threshold would mean that start-up companies would have to comply with the obligations under the data business that would impose additional compliance burden on them. However, keeping a high threshold will become a tool to extract data from larger companies thereby affecting investment opportunities and the adequacy standards. This concept of data business seems misplaced in the current framework and should be removed.

D. Data Sharing

As mentioned earlier, the Report introduces the concept of "Data Businesses", i.e organizations that are collecting, storing, processing, or otherwise managing data and meet thresholds of data specified by NPDA from time to time. The Committee further recommends mandatory sharing of sufficiently anonymized raw/factual NPD (private, public and community) at no cost. Data Businesses are also met with additional compliance of submitting the meta-data about the data user and community from which data is collected, along with other details such as classification, closest scheme, volume etc. This imposes additional burden on the companies that are already faced with a resource crunch. Take the case of a startup, which invests a substantial chunk of its finance in growing and scaling. A barrage of regulations that are vague and with excessive compliance, lead to unintended roadblocks.

The report also provides that Indian citizens and organizations may access such metadata. Subsequently, they may request for underlying data for the following - sovereign purposes such as national security, legal purpose, core public interest such as research and innovation, better delivery of public services or economic purposes such as encouraging competition. Key points to be noted here are that the underlying assumption that raw data is to be shared without any consideration monetarily. Many data intensive businesses, including startups, base their business models on the basis of data collection, curation, analysis. This data forms the basis of the subsequent insights that are developed after analysing the data, resulting in greater value creation for the business and provides the much needed competitive edge. It is a legitimate worry that a regime that would allow for sharing of raw data mandatorily would lead to greater disruption in the market, where custodians lose the agency to enter into such agreements after assessing the value the dataset would play in their business model.

1. Open Public Data

The proposed legislation for regulating NPD aims to

(i) create a framework to generate economic, social and public value from the use of NPD;

(ii) incentivize innovation and creation of new products and services in India and encourage start-ups;

(iii) make community, public and private data available for social, public, and economic value creation; and

(iv) address privacy concerns from processing NPD and to examine the concept of collective privacy.

The need for a legislation is not clear, as the first three objectives can be adequately addressed through a policy framework. We can find the National Data Sharing and Accessibility Policy("**NDSAP**") model as a reference point. In 2012, the Government notified this policy and was designed with the intention of enabling open sharing and leveraging of vast quantities of data generated by the various governmental departments and agencies. This was followed by a Government Data platform under the NDSAP. The existing framework allows for data sharing and one can leverage public NPD for value creation. The government, being the largest repository of data, can utilise this to meet the third objective proposed by the Committee.

As will be illustrated in the section below, there are concerns regarding mandating share of raw data collected by private companies, especially surrounding competition in the market and valuation of data. Having said that, to come up with a model on how to price data, a pilot scheme that centers around public data can be applied first. Consequently, there is the possibility of extending the framework for private NPD as well, with very clearly defined data sharing purposes and processes, which can be built on the learnings from the pilot.

2. Data sharing purposes

Under the report, the data sharing purposes are classified under three heads- sovereign, core public purposes, economic purposes. In all three cases the purposes for which data requests can be made are broadly defined.

This gives scope to a large range of requests that could qualify for mandated data sharing. Assuming but not conceding that data sharing is vital in some sectors and areas and mandatory methods are the only means of achieving them, it is important to reduce scope of misuse of wide prescriptions that could be detrimental to business interests and innovation. Many businesses collect and invest substantial resources in the process of collection, curation and maintenance of

these datasets. Having said that, it is crucial for a more nuanced approach to mandate data sharing. Every data sharing request must be determined on clear mechanisms, which balances private interests and larger public interest.

3. Data Sharing Mechanism

The Report does not touch upon accountability principles that are necessary pursuant to data sharing, which may further de-incentivise companies from investing or carrying out business in India. Bad actors can deploy methods to re-identify individuals and breach privacy and use it for irresponsible activities without accountability. The Report considers certain checks and balances to sharing, they seem to be largely from the perspective of data safety and do not consider the rights of the Data Businesses. In the market, the companies must be provided the agency to identify and disseminate data that can be used by all, in such a manner that it does not affect its own business interest.

a Data Marketplaces

Before establishing data marketplaces, there needs to be an attempt made to do a thorough analysis of the freely available data, working in silos. To standardise already existing data and make it usable for the industry to unlock benefits will be a valuable use case before stepping into including private NPD within the ambit of the regulation. This will also provide an indication regarding the pricing models, how sectors operate and the data sharing practices that already exist, among other things. These considerations can be useful when deciding on pricing models for the various kinds of data.

A voluntary sharing model, where even private companies can volunteer to make data available for the public, in a manner that is usable and which does not affect the commercial interests of the business, must be followed.

This also throws into question how this marketplace of data will be regulated, what are the checks and balances in place to prevent monopolies to thrive. We also need to deliberate on how the resources/lack thereof would influence the dynamics in the marketplace. The framework also places the NPDA in a position of authority to determine the validity of the request from a

social/public/economic perspective. Although the report hints towards the advisory role of the NPDA in data sharing across industry, it is not clear how much autonomy would rest with the data custodian.

b Pricing of data

The pricing of data depends on a variety of factors. As the report indicates there are various stages of value creation as far as data is concerned, business intelligence being the most valuable.

If a standard model for data pricing existed – one that considered many aspects of value such as the age of the data, the reliability of the sample, and other factors – sellers would be able to price optimally in the market and buyers could make appropriate comparisons across data service providers to get a fair price. [15] For long term benefit and easier collaboration between commercial entities and researchers, a method to assess data quality and pricing could be useful.

An important consideration to be kept in mind is the difficulty in estimating the value of data. Many data sets do not have public prices associated with them, and even when it is available, there is no method to the valuation and are often subjective.

4. Mandatory Data Sharing

Mandating compulsory sharing of NPD for private organisations is in fact creating compulsory licensing provisions for copyright through a new legislation. India already has a comprehensive Copyright Act, 1957. Any incremental laws on compulsory licensing etc, should be continued to be covered under Copyright Act, 1957. Therefore this regulation poses a challenge of demarcation between NPD that cannot be shared and non-copyright NPD that can be used as a public resource. It is vital from an ease of doing business point of view that the international community also recognise this principle and adopt it holistically. Therefore, the IPR laws are a more appropriate legislation for deliberation and decision on provisions relating to compulsory licensing of copyrightable datasets. Private NPD should not be brought within the ambit of mandatory sharing.

Recommendations:

[15] A pricing model for data markets, available at: https://core.ac.uk/download/pdf/158298935.pdf

- **Evolving a framework for data sharing purposes:**

A principled approach must be evolved to further qualify the data requests, and the purposes for which data can be shared.

- **Proportionality**: The scope of any request (i.e. frequency, quantity, granularity) for NPD should be related to the purpose for which the data is being requested for. This balancing of interests can be documented in an impact assessment (see "Transparency and Trust" below).

- **Accountability**: Data recipients must be accountable for the data which is received by them. Accountability could come in the form of:
  - ➢ The data recipient must ensure an adequate disclosure mechanism to make sure the data sharing purpose and data use towards that is strictly followed. It might be prudent to study the sector and understand various technical and governance safeguards that can be applied there. This will ensure accountability and build overall trust in the data sharing requirements. In addition, this can lead to an increase in the level of awareness of the value in data sharing and create a strong culture around it.

  - ➢ It is important to note that the data is used for ethical purposes and does not contribute to deepening existing biases.

  - ➢ A main concern when it comes to certain sectors such as health, finance etc- anonymised datasets are valuable, although they come with underlying risks of re-identification and privacy harms. Having said that, a sectoral approach to specifying minimum security standards is crucial when we design systems.

- **Transparency and trust:** In a digital economy, when data sharing and information sharing is the core of unlocking economic and social good, it is important to secure the foundations on values of transparency and trust amongst the various stakeholders. Transparency and

trust in data sharing arrangements can be promoted through *impact assessments*. Any data sharing (whether with a private entity or public authority) needs to be subject to an impact assessment which considers both the benefits and risks involved in the data sharing.

While coming up with mechanisms of data sharing in various sectors, a number of factors must be taken into consideration. A risk- benefit analysis of the proposed data sharing must be conducted. This would also take into consideration the possible harm to data subjects/community in cases where misuse could be anticipated. Another aspect to be kept in mind is to assess if there is a market failure that restricts the availability of data and check if there are other sources to obtain the data.

Developing frameworks and processes which takes into account the interests of both Data Business and the data recipient would lead to accountability and transparency in the decision-making process.

- **Standardization of Formats**: Requiring data to be collected and shared in standard formats only could impact the business operations of start ups requiring them to adhere to formats which may not be conducive to the business.

The Report suggests in many parts that Indian companies and Indian start-ups will be the recipients of this data. This lack of reciprocity has the potential to create the asymmetries in access to data and could eventually lead to a distortion of markets and the lessening of competition.

E. Non Personal Data Regulatory Authority

1. Challenges with establishing a new regulator:

The rationale to allow for a separate regulatory authority rests on the fact there is a market failure that needs intervention. It is said that there is not enough data available for the startups to thrive and grow, even when India boasts of the third largest startup ecosystem in the world.[16]

When setting up a regulatory structure, it is crucial to study the underlying mandates and problem statements the regulation is trying to address to ensure utmost clarity in the role it would play in the ecosystem. In the past, the setting up of many regulators were unclear, thus leading to more problems than solutions.

Creating a regulator involves considerable effort and care must be taken to prevent regulatory arbitrage. A company, which has the liberty to choose between various frameworks that govern similar subject matter, a lack of harmonisation in the goals and obligations under these frameworks allows them to pick the framework that is more convenient for them to comply with.

2. Greater Accountability and Transparency Standards

Regulation of sharing NPD is a relatively novel concept, it would be wiser to enforce a far more relaxed regime that allows data sharing on a voluntary basis. Where the data requests are analysed and are then accepted or rejected by the NPDA. The companies must be given a choice to refuse the request for data sharing. Once the implications of such regulation can be effectively analysed, a more well-rounded policy can be introduced. Frameworks and standards that bridge the trust deficit are crucial for the future of data governance. An impartial regulator with all stakeholders can go to for their grievances is very important to achieve this end.

In this context, for instance, there is the need for democratic processes and accountability mechanisms to be implemented, even by statutory entities and independent regulators. There needs to be considerable thought put into the design and processes of regulation and it's not sufficient to

[16] Guide to Start up India Ecosystem, Retrieved from: https://www.startupindia.gov.in/content/sih/en/international/go-to-market-guide/indian-startup-ecosystem.html

establish a toothless yet "independent" regulator. The report provides insufficient detail about the regulatory setup and accountability grievance mechanisms that need to be articulated. Accountability of regulators must be to the Parliament and not to particular ministries, to reduce conflict of interest and to ensure independence.

The legal regimes created by the NPDA regulating NPD and the DPA regulating personal data are conflicting in nature. It is impossible to create a binary distinction between two overlapping concepts such as these. Therefore, enforcement is going to be ambiguous. Considering the widespread overlaps with the PDP framework, it would be effective to mandate MoUs between regulators and pre-empt the clashes. Additionally, such a regulatory body must be formed in the most transparent manner possible in order to minimise misuse of power. For example, start-ups and small businesses, the intended beneficiaries of this framework might end up facing the overlapping requirements of compliance that could affect the growth much more than the larger organisations in the ecosystem.

All through the report, we find barely any mention of individual autonomy. **Additionally, any redressal mechanism in place must take into consideration the low rates of digital literacy in India. The lack of inclusivity within this sphere will only end up increasing the digital divide in the country.**

3. Regulatory Overlaps

In the report, the government proposes setting up a NPDA. This authority would work with the DPA, the CCI and other authorities to regulate the data, safety and privacy framework of the country. With the report's recommendation for the creation of a NPD authority, the possibility of the kind of conflicts that have been seen in other sectors between different authorities could arise in the data sector as well.

While there is bound to be significant **overlaps in powers,** functions care must be taken to ensure that the regulators involved work together and inter regulatory synergy is built. From past experience, we have seen that there have been overlaps when it comes to the Reserve Bank of India (RBI), Insurance Regulatory and Development Authority of India (IRDAI). In the past there have been instances of overlap of the application of different legislations in a particular situation where both the legislations had a 'non- obstante clause' in it. The courts in India have applied

various principles in such cases, inter alia, including specific legislation prevails over general legislation[17], newer legislation prevails over older legislation[18].

An example of this conflict can be observed in the case of the Competition authority and the Telecom authority with regards to complaints made by Reliance Jio in 2016. The complaint was made about being denied access by the other telecom operators. The main point that arose in the case was whether the CCI was competent to take up the complaint, especially as the Telecom Regulatory Authority of India (TRAI) had taken it up. The Bombay HC in it judgement in *Vodafone India Limited v the Competition Commission of India*[19] made the observation that as TRAI had already begun looking up the case, the CCI did not have the authority to also oversee the case.

Another example of inefficiency as a result of too many regulators is the work related to the rejuvenation of the Yamuna. A Hindu report from 2016 showed that the process of Yamuna rejuvenation was hurt by the involvement of multiple agencies that caused delays in decision-making, along with different agencies making different plans with no harmonization between these plans.[20] Such problems could also arise within the data sector in the country, especially as there is still some debate over the procedures through which personal data becomes non-personal data, and whether personal data can ever be truly anonymized. Thus, it is imperative that the government consider the potential for regulatory conflicts in this sector before it goes ahead with setting up the NPD authority.

4. Regulatory Model

In case of dynamic and fast changing areas such as data governance and protection, **it is prudent to take a soft touch approach to regulation.** Involving industry bodies, civil society must be taken into the scope of regulation, to come up with a regulatory model that is agile and dynamic.

[17] Allahabad Bank v. Canara Bank, [2000] 2 SCR 110
[18] KSL and Industries Ltd. v Arihant Threads Ltd., (2008) 9 SCC 763; Bank of India v Ketan Parekh, AIR 2008 SC 2361
[19] 2017 SCC OnLine Bom 8524 : (2018) 143 CLA 429
[20] Goswami S., Too many Cooks Spoil the Broth, The Hindu, Mar 28, 2106. https://www.thehindu.com/news/cities/Delhi/too-many-cooks-spoil-the-broth/article8403792.ece

It is also to be noted that there needs to be a sectoral approach in coming up with codes of conduct/rules for that sector. The stakeholders involved are widely different, so implementing a law will be harder if a one-fit-all approach is followed. The various models that could be adopted can be found in *Annexure-1*.

Recommendations:

1) Adopt a soft touch approach to regulation and work with the industry and stakeholders to fine tune aspects regarding regulatory mechanism
2) Consider a co-regulatory model that ensures that the equal participation from the state and the industry and other stakeholders to ensure that a bottom up approach is followed.
3) Consider establishing a wing that comprises technical experts which will be tasked with establishing technical standards that allow for a privacy respecting and secure data sharing regime.
4) A sectoral approach to be followed, wherever possible. In circumstances where there are sectoral regulators, MoUs could be signed regarding powers and means of cooperation. This will be critical as a horizontal regulator will have to deal with subject matter that is already under the purview of a sectoral regulator.

[VI]. Anonymisation Framework

A legal framework around anonymization is not totally absent in India and can be explicitly or implicitly borrowed from sector-specific laws. For example, Collection of Statistics Act (COS), 2008 prohibits sharing/disclosing of data without suppressing the identification of the data respondent. This is synonymous to the requirement raised by the CoE in the report that anonymization of personal and non-personal data is mandatory as it could provide collective insights that could open the way for harms against communities. However, there is a need for an Anonymization Framework due to several limitations in the sector specific laws. The same COS act allows disclosure of the data for statistical or research purposes, once the name and address of the informant is removed. This removal is not enough, due to studies showing fast retrieval of personal information by simply aggregating multiple databases. Keeping this in mind, we provide a legal framework which should be considered while anonymizing any NPD and also recommend the Government of India to come up with sector specific codes around anonymized data.

A. Valid Consent Norms

Page 17 of the CoE report talks specifically about the consent norms around Anonymized data and considers seeking consent from the data principal in the same manner as stipulated in the PDP Bill. However, the four pillars of consent in the PDP Bill i.e. informed, specific, clear and capacity to withdraw are inadequate in the age of big data analytics. Thus, there is a need for a cohesive consent framework.

It is possible to shift towards 'graduated consent', in which data principals can give consent to anonymization for each type of data throughout their contract with the service provider, rather than just having a binary choice. Further, another idea of consent can be time limited so that data is no longer used after the time limit has expired, is worth exploring.[21] This is in addition to the four principles stated in the preceding paragraphs. For example, there are electronic marketing calls or bank calls etc. which generally do not ask for the caller's permission to record the call, rather, it is

[21] Royal Academy of Engineering, "*Connecting data: driving productivity and innovation. Royal Academy of Engineering*", 16 November 2015. Available at http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity.

mandatory by the recipient organisation that the calls would be recorded for security and training purposes. Thus, a clause needs to be added in the consent framework that, "the data fiduciary i.e. the recipient of the call has previously notified the caller that he consents for the time being to such communications being sent by, or at the instigation of, the caller".

The phrase "time being" is important here, and how long consent can be relied on depends on context, as also highlighted by Helen Nissenbaum in her "Theory of Contextual Integrity". It might be that we give our consent to re-identify or aggregate datasets as the present privacy policies of a particular product and the terms of the relationship are such that it requires our data. However, on the launch of new features in that same product, it cannot be a case of ongoing consent. This is where graduated consent plays its role. Further, the report does not talk about maintaining records of consent which records the date of consent, what information was provided to the person consenting etc. These records of consent are necessary as a proof to compliance and auditing purposes.

Recommendations:

a) Different consent models should be designed and deployed, for e.g. Graduated consent.

b) Apart from the four principles of consent in the PDP Bill, a fifth principle of time being consent should be thought about and implemented accordingly.

B. Privacy By Design Principles And Anonymisation

Privacy by design is synonymous to anonymized or pseudonymised data so that it is no longer possible to identify an individual from the data itself or from that data in combination with other data. Though, the CoE recommends the data fiduciaries to certain techniques stipulated under the Primer on Anonymization, however, no specific duties are characterized in the context of anonymized data.

Anonymization always deals with the tension between utility and privacy. If certain privacy respecting anonymization techniques are used, but decrease the utility of a particular dataset then a less damaging anonymization should be used. Thus, anonymization should try to balance between Utility and Privacy which the current framework around both Personal Data and NPD lacks.

We strive to recommend some of the duties which guardian data fiduciaries, data fiduciaries and significant data fiduciaries can adhere to and also balance between utility & privacy:

Recommendations:

a) Fiduciaries need to be able to show the means of assessing the risk of re-identification, and the proportionate solution adopted by them.

b) There should be an expert body providing expert advice on anonymization techniques from time to time to keep in line with the best practices globally. CoE can refer to the UK Anonymization Network (UKAN)[22] which is a consortium of University of Manchester and Southampton, the Open Data Institute and the ONS.

c) To maintain utility as well as privacy, CoE can refer to the work of *J. Domingo Ferre and V. Torra*,[23] which examine three procedures to measure the impact of anonymisation: Direct comparison of the categorical data, to obtain average distance between them, Computation of contingency tables of both original and anonymized datasets and the distance between the two, and finally, probabilistic measure which measure the uncertainty on the values of the original dataset given the values of the anonymized dataset.

C. Disclosure Risk Of Anonymized Data

In the context of anonymization, privacy can be compromised by means of two types of disclosure: identity disclosure and attribute disclosure. Though the CoE highlights different types of data involving attributes (like derived data or e-commerce attributes), it fails in distinguishing between Identity and attribute disclosure, as also explained in the preceding paragraphs. However, it leads to a narrow framework on assessing the disclosure risk.

For identity disclosure, disclosure risk is based on the number of reidentifications of a particular anonymized dataset. '*Number of re-identifications*' is defined as the number of records that an intruder may be able to identify, for which data integration algorithms or record linkage algorithms

[22] UK Anonymisation Network website http://ukanon.net/.

[23] J. Domingo-Ferre and V. Torra, "Disclosure control methods and information loss for microdata.," in Confidentiality, Disclosure and Data Access.: North-Holland, 2001, http://vneumann.etse.urv.cat/webCrises/publications/bcpi/cliatpasa01Disclosure.pdf

are used.[24] The report mentions k-anonymity (meant for identity disclosure) as an anonymization method which is a Boolean approach - an approach which focuses on maximizing data utility given privacy constraints. However, there are other model frameworks and advancements of k-anonymity or t-closeness methods, which should be considered by CoE as they are more utility friendly.

Recommendations:

a) Distinguishing between Identity and Attribute disclosure under Anonymized data will aid to develop and bifurcate the anonymization techniques too.

b) In addition to the already proposed anonymization methods, CoE can look to certain advancements in the field of anonymity tools and move towards k-concealment[25], n-confusion[26], (k,t) - confusion[27].

D. Broader Recommendations Under Anonymization

a) The report categorizes anonymized data into a single stream, rather the manner of anonymization would differ between static, dynamic and streaming data. Dynamic data is the case when a database changes with respect to time and data has to be published regularly. For dynamic datasets, the k-anonymous vector space model developed by Guillermo, Abril and Torra[28] can be used as it enables addition of more data continuously while maintaining the k-anonymity property.

[24] W.E. Yancey, W.E. Winkler, and R.H. Creecy, "*Disclosure Risk Assessment in Perturbative Microdata Protection,*" in Inference Control in Statistical Databases, 2002, pp. 135-152. See also, W. Winkler, "*Re-identification Methods for Masked Microdata*," in Privacy in Statistical Databases, 2004, pp. 216- 230.

[25] T. Tassa, A. Mazza, and A. Gionis, "*k-Concealment: An Alternative Model of k-Type Anonymity*," Transactions on Data Privacy, vol. 5, no. 1, pp. 189-222, 2012. See also, A. Gionis, A. Mazza, and T. Tassa, "*k-Anonymization Revisited*," in IEEE 24th International Conference on Data Engineering, ICDE 2008., Cancun, 2008.

[26] K. Stokes and O. Farràs, "*Linear spaces and transversal designs: k-anonymous combinatorial configurations for anonymous database search notes*," Designs, Codes and Cryptography, vol. 71, no. 3, pp. 503-524, 2014.

[27] K. Stokes and V. Torra, "*n-Confusion: a generalization of k-anonymity*," in 2012 Joint EDBT/ICDT Workshops, 2012, pp. 211-215.

[28] G. Navarro-Arribas, D. Abril, and V. Torra, "*Dynamic Anonymous Index for Confidential Data*," in 8th International Workshop on Data Privacy Management and Autonomous Spontaneous Security, 2013, pp. 362-368.

b) In the case of streaming data, primarily the challenge is that data is feeded in portions and in an unstructured sequence. For example, as the authors have described in a report,[29] that during winters people commute less by bike due to snowfall. However by knowing a person comes to work by bike and a set of GPS traces, it may not be possible to identify the person in summer, but possible in winter. Thus, static privacy rules are not the way ahead and the new adaptive privacy preservation techniques are required. So, apart from differential privacy, certain perturbative approaches can be considered.[30]

[29] G. Krempl et al., "*Open challenges for data stream mining research*," ACM SIGKDD Explorations Newsletter - Special issue on big data, vol. 16, no. 1, pp. 1-10, 2014.
[30] F. Li, J. Sun, S. Papadimitriou, G.A. Mihaila, and I. Stanoi, "*Hiding in the Crowd: Privacy Preservation on Evolving Streams through Correlation Tracking*," in IEEE 23rd International Conference on Data Engineering, 2007, 2007, pp. 686-695

Annexure 1: Models of regulation

Initiatives for co-regulation and self-regulation have taken many forms, such as codes of conduct, declarations, charters, agreements, rules, standards and labels. Their links with regulations are also extremely diverse, ranging from legally compatible autonomy, to requirements laid down in the wake of framework legislation (co-regulation) or even the local definition of rules then supported or even made mandatory by the legislator.

1) **Voluntary Codes of conduct**: In general, however, codes of conduct are a self-regulatory model with a low degree of government involvement. They are based on agreements among industry members, which may be embodied in contracts, to adhere to the code. The contracts may provide for sanctions for breach of the code and may provide mechanisms, such as mediation or independent third-party arbitration, to deal with disputes or non-compliance. There may even be provisions for independent auditors to aid in code enforcement.[31]

   Internationally, codes have been encouraged. The United Nations Consumer Protection Guidelines issued in April, 1995 endorses codes of conduct as a means of achieving fair trading outcomes. In the United States under the Insider Trading and Securities Fraud Enforcement Act,[32] brokers and dealers must establish and enforce written policies and procedures to deter insider trading. Voluntary codes are an integral part of Australian fair trading legislation. The Australian Competition and Consumer Commission (formerly the Trade Practices Commission) encourages industry codes that do not inhibit competition.

2) **Statutory Self-Regulation**: The self-regulatory organization is often established by statute, which delegates a range of regulatory powers to the SRO. At one extreme, these can include a complete range of regulatory powers, including rulemaking, monitoring, enforcement and sanctions. This full form of delegation is generally limited to the regulated occupations. The actions of these bodies have a sufficiently public character that they are subject to judicial review. Occupational regulation ranges from licensure to certification to

[31] Margot Priest, 'The Privatization of Regulation: Five Models of Self-Regulation', 29(2) Ottawa Law Review 233, 242 (1998)
[32] 15 U.S.C.A., §78o(f) (1997).

registration. The strictest regime is licensure; entry is controlled and a member of the occupational group cannot practice the profession without a professional licence

3) **Firm- Defined Regulation**: In the third model, firm-defined regulation, the private sector regulated firm takes on regulatory responsibilities such as developing rules, enforcing rules, and even imposing sanctions. The rules are tailored to the firm. The regulatory structure is based on two elements: the public enforcement of privately written rules, and the publicly mandated and publicly monitored private enforcement of those rules. Under the enforced self-regulation model, the government and the firm negotiate to produce a set of rules that are particular to the firm; these custom tailored rules must not result in any less protection than would the otherwise applicable public rules. The U.S. Mine Safety and Health Act of 1977 and regulations allow mine operators to submit their own plans for ventilation, dust control and roof supports for the agency's approval. The U.S. Environmental Protection Agency requires companies involved in the production, distribution or storage of oil to prepare a Spill Prevention Control or Countermeasure Plan, and the U.S. Clean Water Act authorizes civil penalties of up to $5000 per day for deviations from these privately written rules.
4) **Supervised Self-Regulation:** Membership is usually mandatory for participation in the activity being regulated. There are a number of points at which the supervisory body can exert control, including approval of certain decisions, investigations, appeals, reports and imposition of legally binding rules. There are also a number of points where more subtle influence can be brought to bear on the SRO by the supervising agency, such as informal consultation, exchanges of personnel, and formal or informal exchanges of information. The other regulatory functions, such as monitoring and enforcement, are also usually shared, although the oversight body generally performs an appellate function in dealing with enforcement and sanctions. The oversight body may be more or less active in its supervision, and it may concentrate on certain parts of its mandate, such as approval of rules. Supervised self-regulation involves a strong role for the SRO, but the parallel supervisory activity of the agency means that less emphasis is placed on the initiative and powers of the SRO in comparison to the statutory SROs, which may have a relatively high degree of autonomy and deference to their decisions. The primary examples of supervised self-regulation are found in the securities industry. Although the structure of securities self-

regulation has not proven to be the widely adopted model it was originally intended to be, it is an example of a mature, well-established self-regulatory structure that highlights both the advantages and pitfalls of statutory self-regulation. In the United States, the Securities and Exchange Commission supervises the various U.S. stock exchanges and the National Association of Securities Dealers; the Commodity Futures Trading Commission supervises the various commodities exchanges. In India, SEBI can be a fitting example.

5) **Regulatory Self- management**: Rulemaking and policy making remain the responsibility of the government, which also retains a residual enforcement and sanctioning capacity. Implementation of the regulatory program through the application of rules and monitoring of compliance is carried out by an industry self-management organization, which is a nonprofit corporation formed to fulfill the self-management responsibilities. The arrangements between the government and the self-management organization are contractual (and authorized by statute), but the organization does not have the same broad range of responsibilities and powers found in the SROs operating in statutory self-regulatory and supervised self-regulatory regimes.

Mandatory self-regulation, a variant of regulatory self-management, represents a hybrid of direct government regulation and voluntary self-regulation. Rule communication, enforcement and sanctions are carried out privately, at the firm level; the government is not directly involved in these functions. Government involvement comes about through the legal requirement that firms self-regulate and, generally, the establishment of the rules of the regulatory regime. Mandatory self-regulation operates as both a direct complement to and a part of a government system of regulation. The U.S. Occupational Health and Safety Administration Cooperative Compliance Program that was established in 1979 in California was a forerunner of many of these programs. Mandatory self-regulation also changes the role for government regulators; it does not necessarily diminish or reduce the importance of the role, however. There is general agreement that a regulatory presence is required, particularly to deal with non-compliers. The regulator may also require a higher than usual degree of expertise since the focus of the regulation is on risk management and industry processes.

About The Dialogue:

The Dialogue is an emerging public-policy think-tank with a vision to drive a progressive narrative in India's policy discourse. Founded in 2017, we believe in facilitating well-researched policy debates at various levels to help develop a more informed citizenry, on areas around technology and development issues.

Our aim is to enable a more coherent policy discourse in India backed by evidence and layered with the passion to transform India's growth, to help inform on public-policies, analyse the impact of governance and subsequently, develop robust solutions to tackle our challenges and capitalise on our opportunities. To achieve our objectives, we deploy a multi-stakeholder approach and work with Government, academia, civil-society, industry and other important stakeholders.