Legal Issues
in
Government's Report on Non-Personal Data Governance Framework
&
Recommendations made by ALG in its *Firm Submissions* in response thereto

Why So Much Hype Around Data?

- 5 Exabytes of data created between the dawn of civilization through 2003, but that much information is now created every 2 days
- 90 Zettabytes of data expected to be generated in 2020 and 175 Zettabytes in 2025
- Business value derived from global Artificial Intelligence (AI) in 2020 – USD 2.65 trillion
- AI deployed organizations grew from 4% to 14% between 2018 and 2019
- Data analytics talent requirement to grow from 5,10,000 (2018) to 8,00,000 (2021)

References: Page Nos. 5-7 of '*Report by the Committee of Experts on Non-personal Data Governance Framework*'

Basic Concepts of Non-Personal Data

❖ What is Non-Personal Data?

- Data which is not ‘Personal Data’ as per Section 3(28) of the Personal Data Protection Bill, 2019.
- Data without any personally identifiable information (which includes Personal Data which has undergone the process of anonymisation).

❖ Categorization of Non-Personal Data

- Public non-personal data
- Community non-personal data
- Private non-personal data

Ownership of Non-Personal Data

- **Public non-personal data** – Data principal will be the government
- **Private non-personal data** (Data which is collected by the private organisation) –
  - Raw data to be disclosed at no remuneration
  - Value-add over the raw data would lead to addition of monetary value
  - Committee recommends - Algorithms/proprietary knowledge may be considered for data sharing (Point of major concern)
- **Community Non-personal data** –
  - Many overlapping interest
  - Report recommends ownership on the notion of 'beneficial ownership/interest'
  - Data principal will be the trustee of that community

Sensitivity of Non-Personal Data

- **What is the need to bring in the concept of sensitivity?**
  - Risk of de-anonymisation of anonymised data – Numerous instances worldwide
  - Threat to national security, collective harm to a group etc.
  - For personal data, the PDP Bill provides for three categories:
    - General Data
    - Sensitive Data
    - Critical Data

The Report recommends application of same criteria while analysing sensitivity of non-personal data.

Consent for Anonymised Data

- Guiding principle – Anonymized personal data should continue to be treated as non-personal data of the data principal
- PDP Bill (for personal data) – Consent is necessary for collection and processing of personal data
- Recommendation of the committee – Data principal should also provide consent for anonymisation and usage of this anonymized data while giving consent for collection and usage of his/her personal data
- Committee also recommends that appropriate standards of anonymisation be defined to prevent/minimize the risk of re-identification

Key roles in Non-Personal Data Ecosystem

- Data principal
  - For public non-personal data
  - For private non-personal data
  - For community non-personal data
- Data custodian
- Data trustees
- Data Trusts

Data Business

- Horizontal classification (Example – Healthcare industry)
- Organizations that meet a certain threshold of data collection, processal, storage etc.
- Lightweight process for registration as data business
- While registering, a data businesses need to disclose what they do and what data they collect, process and use, in which manner, and for what purposes
- Provide open access to meta-data and regulated access to the underlying data
- Meta-data collected from them will be available in open access meta-data directories
- Requests for underlying data can be made to data businesses

Data Sharing Purposes

- **Sovereign purposes**
  - National security, law enforcement, legal and regulatory purposes
- **Core Public Interest Purposes**
  - Community uses/benefits, better delivery of public services, policy development, etc.
  - Research purposes
- **Economic purposes**
  - Start-ups and businesses to have access to meta-data to spur innovation
  - Data requests by the governments
  - Leverage data as training data for AI/ML systems

Report recommends health sector as a pilot use-case for Non-Personal Data Governance Framework

Non-Personal Data Authority

- Specialized knowledge of data governance, technology, latest research and innovation
- Objective of unlocking value in non-personal data
- Will work in consultation with DPA, CCI and other sector regulators
- Sector regulators can build additional data regulations (if required)

❖ Two roles to be played by the NPDA:

- (i) Enabling – Ensure that data is shared for sovereign, social and economic welfare, and regulatory and competition purposes
- (ii) Enforcing – Ensure that stakeholders follow all rules and regulations, provide data when requests are made, undertake evaluations to minimise risk of re-identification of anonymised personal data, etc.

ALG's Comments & Recommendations

1. Data-sharing purposes encompass a broad range of activities which could be of concern from the perspective of dilution of IP rights.

- The Report provides for requests being made for the underlying data for various purposes including sovereign, core public interest and economic purposes
- These purposes have been very broadly interpreted in the report to include a variety of activities
- Mandating data sharing for a broad range of activities could possibly pose a risk in the form of dilution of value in data, thus, undermining the investment by the organization towards collection and processing of the data

ALG's Recommendation – "*We recommend that the scope of the purposes for data sharing be reconsidered keeping in mind the situations wherein data sharing is absolutely necessary (for the larger public interest).*"

2. There should be clear provisions for protection of mixed datasets

- The Report refers to the concept of a 'mixed dataset' which it explains as – *"A mixed dataset, which represent a majority of datasets used in the data economy, consists of both personal and Non-Personal Data."*
- The Report also reads – *"In the European Union context, the Non-Personal Data Regulation applies to the Non-Personal Data of mixed datasets; if the Non-Personal Data part and the personal data parts are 'inextricably linked', General Data Protection Regulation apply to the whole mixed dataset."*
- On one hand, while the Report has recognized the framework of protection of mixed datasets in the European Union, it has not elaborated on the applicable framework in India

ALG's Recommendation - *"There does not seem to be clarity to the effect whether 'mixed datasets' shall be protected by the Non-Personal Data Protection framework which is being envisaged by the Report or through the legislation arising out of the Personal Data Protection Bill, 2019. It is recommended that there should be clear cut demarcation of laws applicable to mixed datasets."*

3. Requirement of penal consequences of unauthorized usage of data

- The Report very briefly and generally discussed the liability arising out of breach
- The Report reads as - *"One reason for standards driven approach is that organisations, that comply thoroughly with the laid-down standards via annual lightweight self reported, self-audited digital compliance reports, exhibit good faith and have best-effort internal processes in-line with the best of industry standards, are to be indemnified against any vulnerability found as long as they swiftly remedy it."*
- While the Report does provide for indemnification of the party at whose account the breach has been caused (if they are able to swiftly remedy the breach), it does not provide for any penal actions as part of breach of liability
- The Report seems to have provided a long rope to the parties whose actions result in breach of protection of non-personal data

ALG's Recommendation - *"The Non-Personal Data Protection Framework should envisage and mention specifically that there would be penal consequences in cases of unauthorized usage of non-personal data [including non-consensual anonymisation of personal data (and usage thereof)]."*

4. Algorithms / proprietary knowledge should not be considered for data sharing, unless voluntarily disclosed

- The Report states - *"Algorithms / proprietary knowledge may not be considered for data sharing."*
- Usage of the word 'may' results in ambiguity on the subject on whether or not Algorithms / proprietary knowledge (some of which could be the subject matter of copyright and trade secrets) could be disclosed as non-personal data
- Public disclosure of such data without the consent of the data custodian will lead to the effort (in the form of time and money) of the data custodian being non-utilized
- It would not serve as a motivating factor for the Indian entities to invest in innovation and research
- Deterrent for individuals and companies to invest in R&D when they would be under constant threat that their algorithms / proprietary knowledge could be made public

ALG's Recommendation -*"The disclosure of non-personal data of the data principal in the form of Algorithms / proprietary knowledge should not be allowed, unless express consent to this effect is obtained from the data principal."*

5. Government should consider some other sectors (other than the health sector) for the pilot project

- The Report states – *"large anonymised data sets of health data, could lend community level insights into diseases, epidemics, and community genetics – leading to better tailored health solutions for the community. Accordingly, the Committee considered health as a pilot use-case for the Non-Personal Data governance framework".*

- Privacy concerns and the detrimental effect on individual(s) and communities as a whole cannot also be undermined

- There are frequent instances of discrimination with person(s) (and communities as a whole) suffering from diseases/ailments by the public at large

- Since there might be concerns whether sufficient safety standards are in place to ensure anonymisation of data or not, the healthcare sector might not be a suitable sector for conducting the pilot project

ALG's Recommendation – *"A sector (such as the FMCG sector) should be chosen for the pilot project since it may contain less sensitive personal data and the chances of harm being caused by misuse of such data are much lesser."*

ALG's Further Recommendations

- Lack of clarity on the stakeholders with whom consultations were held by the committee and the process adopted in respect thereof
- The concept of collective privacy as well as provisions in respect of governance of data pertaining to collective privacy should be elaborated
- Lack of specificity in 'social/ public/ economic benefit perspective' which is to be adopted by the Non-Personal Data Authority in deciding cases entailing refusal of data sharing requests
- If categorisation of data is done on the basis of its creation, the extent of protection of data needs to be specified
- Non-Personal Data need not necessarily inherit the sensitivity characteristic of the underlying Personal Data from which the Non-Personal Data is derived
- Graduated sharing obligations should not apply to sensitive and critical data
- Adequate mechanisms should be put in place for protection of data collected during registration as a data business

Thank You!

Questions?

Sri Lekha Rayapati, Associate
Gaurav Bhalla, Partner

Disclaimer: Views, opinions, and interpretations are solely those of the presenters, not of the firm (ALG India Law Offices LLP) nor reflective thereof.

This presentation hosted at: https://www.algindia.com/wp-content/uploads/2020/11/Presentation_Non-personal-data_V7-Copy.pdf