Governance of Non-Personal Data – A Tricky Balancing Act

Introduction

Eric Schmidt, one-time CEO of Google, quipped back in 2010 that "there were 5 exabytes of information (data) created between the dawn of civilization through 2003, but that much information is now created every 2 days". The comparison gets even more ridiculous when you realize that the world is well on its way to generating about 90 zettabytes of data in the year 2020 alone. There is no doubt that we are living in the age of big data. IoT (Internet of Things) devices have proliferated abundantly and are penetrating deeper in the society - collecting more and more data each day. Artificial intelligence is being implemented across the board and the demand for the raw material in the form of datasets to feed the furnace of machine learning is increasing exponentially.

While debate rages on about the regulation of personal data in India and the Personal Data Protection Bill, 2019 ("**PDP Bill**") is presently languishing with the Joint Parliamentary Committee, the country has started contemplating a regulatory framework around governance of "non-personal data". With this view, the Ministry of Information and Technology ("**MeitY**"), constituted a committee on September 13, 2019, to create a set of rules which would govern non-personal data ("**Committee**"). The Committee is chaired by Mr. Kris Gopalakrishnan, co-founder of Infosys. The mandate of the Committee was to study various issues relating to non-personal data and to make specific suggestions for consideration of the Central Government on the regulation of non-personal data. The Committee has, on July 12, 2020, presented a draft report on 'Non-Personal Data Governance Framework' ("**Report**") and sought feedback from interested stakeholders on the Report by August 13, 2020.

Not all Non-Personal Data is created equal

Before discussing the recommendations of the Report, it may be relevant to briefly understand what constitutes "non-personal data" and why does the same need to be regulated in the first place. Simply speaking, non-personal data constitutes all such data which is not 'Personal Data' (as defined under the PDP Bill), or the data without any personally identifiable information. The Committee in the Report acknowledges that defining non-personal data properly would be very critical in regulating the same. Non-personal data has been further categorised as follows by the Committee:

(i) **Public Non-Personal Data** - means non-personal data collected or generated by the governments, or by any agency of the governments, and includes data collected or generated in the course of execution of all publicly funded works. E.g. anonymised data of land records, public health information, vehicle registration data etc.

(ii) **Community Non-Personal Data** - means non-personal data (including anonymised personal data) about inanimate and animate things or phenomena – whether natural, social or artefactual, whose source or subject pertains to a community of natural persons. E.g. datasets comprising user-information collected by telecom, e-commerce, ride-hailing companies, etc.

(iii) **Private Non-Personal Data** - means non-personal data 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 person or entity, and includes those aspects of derived and observed data that result from private effort.

The Report also highlights 3 key non-personal data roles:

(i) **Data Principal** - in case of Government and Private Non-Personal Data, the data principal will be the corresponding entities to whom the data relates. In case of Community Data, a community would be the data principal.

(ii) **Data Custodian** - undertakes collection, storage, processing, use, etc. of data in a manner that is in the best interest of the data principal.

(iii) **Data Trustee** - data principal group/community would exercise its/their data rights through an appropriate data trustee.

Separately 'data trusts' are the institutional structures, comprising specific rules and protocols for containing and sharing a given set of data and 'data infrastructure' corresponds to technical-material elements required for data sharing, like actual databases, APIs, organisational systems, etc.

The Committee has also articulated the following legal basis for establishing rights over different types of non-personal data. The Community Non-Personal Data collected in or from India would belong to the community concerned. The rights over this data collected in India would vest with the trustee of that community, with the community being the beneficial owner. In case of Public Non-Personal Data, since this data is derived from public efforts, the datasets created would partake of the characteristics of a national resource. In the case of Private Non-Personal Data only the raw / factual data collected by a private organization may need to be shared subject to the well-defined grounds; and proprietary knowledge would not be required to be shared.

Big Data - Bigger Ambitions

Atul Butte of the Stanford School of Medicine has, not hyperbolically, stated that "Hiding within the mounds of data is knowledge that could change the life of a patient or change the world". The importance of such data is recognised the world over. First world countries like USA and EU and giant corporations like Google, Facebook and Amazon have taken significant strides in identifying, analysing and harnessing the immense potential of such data. European Union Policy on Digital Single Market states that "Free flow of non-personal data is a pre-requisite for a competitive data economy within the Digital Single Market. To fully unleash the data economy benefits we need to ensure a free flow of data, allowing companies and public administrations to store and process non-

personal data wherever they choose in the EU".

The Committee has recognised the benefits which could accrue to India and its communities and businesses from such non-personal data. The PDP Bill states that the Central Government can direct a data fiduciary or a data processor to provide anonymised personal data or non-personal data "to enable better targeting of delivery of services or formulation of evidence-based policies by Central Government". Now the Committee has gone further in the Report and have stated that sharing non-personal data collected by both government and private organizations with citizens is likely to lead to increased transparency, better quality services, improved efficiencies, and more innovation. The shared non-personal data may be useful for Indian entrepreneurs to develop new and innovative services and products, which citizens may benefit from. Non-personal data may also be used by researchers, academia and governments for creating public goods and services like an Indian genome repository, data for training natural language translation systems on Indian languages, etc.

The massive population size of the country and significant cell-phone and internet penetration allows India to have an untapped repository of databases which may have unforeseen benefits. The big data companies will have to carefully evaluate the obligations which might be imposed on them under any such regulatory frameworks considering such companies are the largest collectors and repositories of such data. However, access to such data may prove useful to small business and start-ups.

The Bugaboo of Big Data

However, all is not well in the wonderland of non-personal data exploitation. Like every asset, non-personal data is also prone to misuse. Therefore, it is extremely important that such non-personal data is regulated appropriately and used responsibly. The Report does not shy away from highlighting the possibilities of collective harm related to non-personal data. Misuse of data belonging to a group or community may arise from inappropriate exposure or handling of such data. The Report acknowledges that non-personal data is an emerging concept which will need to be examined and defined in detail in the future.

Some areas of sensitivity around misuse of non-personal data are as follows:

- Such data could have an impact on national security or strategic interests of the country and in the hands of enemies of the state, could be help influence strategic military decisions against India;
- Such data bears risk of collective harm to a group (collective privacy etc.) e.g. ghettoization or institutionalised bigotry on the basis of race, religion, sexual orientation etc.;
- Such data could be business sensitive or may include confidential information and may have been derived using proprietary technology developed by such companies;
- Such data, even if anonymised, bears a risk of re-identification.

Key Recommendations:

The Report touches upon the following aspects of non-personal data and makes

following key recommendations in relation to the governance of the same:

1. **Consent of Data Principal** – The Committee recommends that the data principals should provide consent for anonymisation and usage of this anonymized data while providing consent for collection and usage of his/her personal data. Appropriate standards of anonymisation must be defined to prevent / minimize the risks of re-identification.

2. **Data Business** - A new category / taxonomy of business called 'Data Business' must be created with certain data threshold criteria. Existing businesses in various sectors, collecting data beyond a threshold level, will get categorized as a Data Business and such Data Business would have to register once they reach a certain data related threshold. Every Data Business must make prescribed disclosures pertaining to their business and nature of data collected, processed and used by them, including the manner and purpose.

3. **Access to Meta-Data** - The meta-data about data being collected, stored and processed by Data Businesses should be stored digitally in meta-data directories in India. Such directories should be freely accessible to Indian citizens and India-based organizations. Subsequently, data requests may be made for the detailed underlying data.

4. **Data Sharing Purpose** - Data may be requested on following grounds – (i) Sovereign: security, legal, law enforcement and regulatory purposes; (ii) Core Public Interest: community uses / benefits or public goods, research and innovation, for better delivery of public services, policy development, etc.; and (iii) Economic Welfare Purposes.

5. **Data-Sharing Mechanisms** – The Report sets out detailed recommendations on data sharing mechanisms to ensure appropriate access balanced with suitable protections.

6. **Non-Personal Data Authority** - The Non-Personal Data Authority must be created and tasked with enabling legitimate sharing requests and requirements; regulating and supervising corresponding data sharing arrangements involving Data Businesses, data trustees and data trusts; addressing market failures and supervising the market for Non-Personal Data.

7. **New Legislation** - The Committee recommends that the proposed Non-Personal Data Governance Framework becomes the basis of a new legislation for regulating Non-Personal Data to ensure effective enforcement of the framework on a national scale.

Conclusion

The world has incontrovertibly changed in the last few decades. We are living in the world of Bid Data now and there is no going back. Peter Sondergaard, senior vice president, Gartner Research stated that "Information is the oil of the 21st century, and analytics is the combustion engine". In this backdrop, careful and considered approach to non-personal data is of paramount importance. There are many lessons to be learned if we are

to harness Big Data for the betterment of the world while avoiding the misuse of the same and protecting the rights of the data principals.