#NAMA: Why does the Indian government want to regulate non-personal data?

When it comes to data, the primary aim of the government is to earn money, and “the only way they can earn is when they actually sell personal data in a non-personal way”, said a speaker at MediaNama’s roundtable discussion on non-personal data. Speculating about the intentions of the committee of experts that is working on non-personal data, especially Avanti Finance CTO Lalitesh Katragadda, a speaker said that the committee wants to “**ring fence the Indian data economy so that data that is useful for India’s national development can be used by Indians**”. They likened it to rise of similar debates in other parts of the world including Germany (“data sovereignty” and “community data”) and France. Another participant highlighted the few stated objectives of government when it comes to data governance: economic growth, its governance, and privacy protection.

(Note: The discussion was held under the Chatham House Rule; quotes have not been attributed to specific people. Quotes are not verbatim and have been edited for clarity and to preserve anonymity. Also note that this discussion took place before the PDP Bill, 2019, was made public.)

Why does the government want control over non-personal data?

- **Some private data is valuable to community:** Globally, there is now an understanding that some companies dominate the global data market. From that, a speaker highlighted, has emerged an idea that “some of this data is valuable to the community”. While how this community data is defined remains up for debate, “some data that private companies hold can be used for social good/purpose” was potentially the primary goal behind creating the committee of experts, they said.
- **Data trade:** Another speaker guessed that the committee wants to formulate a policy that allows data in certain market segments to be traded in the Indian marketplace. They suggested that the committee could follow the national AI marketplace model, but that is also rife with issues: intellectual property, ownership questions, pricing data sets given lack of economic studies around that, etc. “These are the kinds of discussions that are happening right now: how do you price the data sets, how do you ensure that it is tradable as a commodity, is it tradable as a commodity, if it could be a marketplace, if it would have the ordinary rules of a securities’ marketplace, etc.,” they said.
- **Regulating group behaviour:** A participant pointed out that the government’s policies and general statements, both at national and international fora, about data colonialism and data sovereignty show that the “government is interested in collecting certain forms of data for economic and political purposes of regulating group behaviour and knowing more

about the Indian population". According to the speaker, that is already happening, as proven by "legal mandates such as localisation".

What kind of data does the government seek to control through non-personal data governance?

- **Data from IoT devices:** Remarking that the "government wants to create a governance framework(s) for all data", a speaker said that the government wants "personally sourced data which is not sourced from personally identifying information", and thus means data from sensors and IoT devices. Thus, according to the speaker, "a lot of the argument is about how such data has to be governed, regulated and perhaps shared with and used by start-ups".
- **Behavioural data:** A participant pointed out that the government is looking at behavioural data which can include personal data. "This can be, in their [government's] mind what might be anonymised or well-anonymised data, but it is also data about communities and groups and how to model behaviour using data," they said. They highlighted the government's concerns that data-based behavioural models are currently created and implemented by large technology companies that are based outside of India. "There is a very large economic rationale for why they want to regulate and localise this kind of data, and ensure more control over the kind of behavioural models that are created," they said. The speaker cautioned against using data modelling to regulate human behaviour and the need to limit access to it, and its use: "The whole identity crisis that the Open Data Movement had was because they did all the groundwork to create an open maps infrastructure, which was ultimately misused by builders to demolish slums."
- **Consumer spend and health data:** A speaker pointed out that in a 2014 paper, iSPIRT, a Bangalore-based lobby of technology companies in India, had said that consumer spend data belonged to the community. In a 2019 paper, they said that about health data. As a result, the committee on non-personal data would also want to regulate consumer spend data. "Consumer spend data is found in e-commerce, it is found in the financial sector, it is found in the non-banking financial sector. Incidentally, these are also sectors where you will see increasing technical mandates," the speaker highlighted.

Factors to consider while governing non-personal data

A participant speculated that the committee on non-personal data would probably propose a policy that "incentivises return of certain kinds of data sets to a marketplace of sorts".

- **Use must govern policy:** A number of participants agreed that the governance of non-personal data should be governed by how the data is used. That should include establishing how the data would/should be used, who its owner is, what kind of risks it poses to users and establishing thresholds for it, etc.

- **Ownership of data determines policy:** As consent over NPD is considered, it boils down to who owns the data, a participant said. “When ownership comes into play, IP comes into play, as do anti-trust and competition issues. There has to be some sort of regulation now who should regulate these considering these are the issues that come out of use of NPD data and ownership,” they said.
  - For a lot of corporates, data is a business asset. A speaker used the example of a truck driver to illustrate their point: “If I am a truck driver who always comes to the mining operation drunk, my actions are behavioural data. Is it corporate data? Does it belong to me? Should it be used punitively against me? Should it be used to prevent other people from coming to work drunk? Should it be used to look out for signs in other people?” They also highlighted how if this data is traced back, it could be used to persecute people.

However, a lawyer said that “**it is not about who owns the property, but what uses are made of that property which should determine how it is regulated**”, with respect to data collected by instrumentation in smart cities. They clarified that even though the basic principle of jurisprudence, governed by John Locke’s theory of labour, says that “anybody who puts effort into collecting certain information or collecting certain data becomes the owner”, in this case, it should be regulated by use.

Benefits of processing non-personal data

- **Certain level of granularity can help in emergencies:** In certain situations, more personally identifiable data is more useful, a person said. Taking the example of a flood, they said, if a government passes a law that says they need location information at an aggregated level for the lake for the last 24 hours. “We will store it in a particular manner, and once the situation is sorted, will delete the information. In emergency situations, certain kinds of location data, if governed well, is of immense use to society. And if that data is deleted immediately after that and not used for anything else, it is a positive net gain,” they said.
- **Track spread of diseases:** Another person said that if the government wants to map out spread of disease, it would be useful to see in which direction it spreads, so that it can take steps to mitigate and halt the spread.

Harms of processing non-personal data

- **Biased, inaccurate data collection can lead to ‘horrific’ outcomes:** If the aim is to create a kind of public commons of non-personal data, we need to be especially careful about the kind of inherent biases that may be built into databases, a participant warned. Cautioning that use of AI and ML would mean that a human might not even look at some of these decisions, they said that “if these data sets are not properly curated, if the data that is fed into these data bases is not accurate, significant decisions could be taken about entire populations and communities with inaccurate and biased data”.

- **Non-personal data, combined with other data sets can be used to target (groups of) people:** Taking the instance of a popular browser, a speaker highlighted how it collects an “astonishing amount of telemetry data”. A lot of this data, which is not associated with a personal identifier and the related IP address is not “regarded”, is publicly available on GitHub as open data set for people to work with.
  - This data includes: how long people take to connect to the website, the websites themselves, which websites are working, how long an application is open, etc. If this non-personal data is combined with other data sets, it could be used to block the internet in particular regions or decide which websites are popular or not. Another speaker said that when fintech data “comes together”, “it is quite beautiful from a marketing perspective, but from an individual’s perspective, it can be extremely dangerous”.
- **Misuse by government through sale and merging of data sets:** Taking the example of the government selling vehicular data to private companies, a participant said how the Bulk Data Sharing Policy allows that. As a result of this policy, if there is any misuse of personal identifying information, violators can only be looked at under the Information Technology Act, but data protection itself is not a tenet of it. Another speaker drew attention to the fact that the Bulk Data Sharing Policy “actively discourages merging different data sets” and the onus of misusing such data would be on the companies who buy it.
  - The previous speaker also brought up a new committee formed under the Insurance Regulatory and Development Authority of India (IRDAI) that is looking at whether traffic violations can be linked to motor insurance premium. The committee, they said, is headed by a person from HDFC bank and consists of many private and public insurance companies along with officials from the Delhi Police.
- **Sale of aggregated data:** A speaker commented on how a lot of aggregated data is sold by PR companies. The trends that then emerge are used to profile and specifically target specific people for a whole host of purposes.
- **Group privacy remains ignored:** A speaker cited the Sidewalk Labs project in Toronto to highlight how community data can be used for behaviour modelling and to profile communities. “Google’s open infrastructure and technology arm, Sidewalk Labs, entered into a public-private partnership with a neighbourhood in Toronto. Under this project, Google embedded passive sensors within urban infrastructure to decide how streets

and traffic management is going to happen, what kind of neighbourhoods get developed, how do you locate the most essential communities to allocate resources to and so on," they explained. While this may not look at individual data, but data is being processed to formulate public policy, group privacy is not considered and such schemes allow for greater state surveillance as well.

What can be done to prevent some of these harms?

A lot of the harms associated with non-personal data are also possible with PII, a speaker said. To that end, they suggested some 'Lean Data Practices' that can be implemented to mitigate these risks. These include norms around data collection, data storage, data processing, and deletion of useless data.

Not share certain personally identifiable information (PII) with government: A speaker suggested that we could come up with a list of PII that should never be present in data sets that are shared with the government.

How will consent work?

A speaker pointed out that the thus far, consent has not been taken into account when considering anonymised data. Another participant said that as we consider consent, it all boils down to who owns the data.

- **Consent doesn't come into the picture:** Since data is irreversibly anonymised by the data processor, it becomes the IP of the processor, and thus the question of consent does not really arise, argued one speaker.
- **Without consent, there can be other harms:** Taking the instance of demographic information, a person asked if they could refuse to consent to it. "I don't want things like my religion being used even in an aggregated way because aggregated anonymised non-personal data, which was originally PII, can lead to significant harms," they said. They took examples of databases in Andhra Pradesh and Telangana that can be analysed for religion and yield gram panchayat-wise percentage of Hindus, Muslims, etc.

*

Read our coverage of the our discussion on Non-Personal Data in Delhi here. The discussion was held in New Delhi on November 28, 2019, with support from Amazon Web Services, Facebook and FTI Consulting.