Personal Data Protection Bill strikes a discordant note on ‘non-personal data’

““Anonymised data” under the Bill refers to data from which markers of identity have been “irreversibly” removed,” writes Divij Joshi.

On December 11, 2019, the Personal Data Protection Bill was introduced in the Lok Sabha as a landmark legislation meant to safeguard the constitutional guarantees of privacy for Indian citizens and provide a just and equitable vision for the future of India’s digital economy. However, an incongruent provision in the Bill departs from this expectation — clause 91 enables the central government to direct any of the regulated entities under this Act to provide anonymised personal data or non-personal data to enable “targeted delivery of services” or “evidence-based policy making”. The implications of such a provision for India’s digital economy must be carefully considered as Parliament reviews the proposed legislation.

“Anonymised data” under the Bill refers to data from which markers of identity have been “irreversibly” removed. However, anonymisation should not be seen as a silver bullet for the use of large datasets without compromising privacy. Recent research shows that the common methods of anonymisation applied today are imperfect and data released as “anonymous” can be re-identified, particularly with the use of modern machine learning techniques. This renders large anonymised datasets vulnerable to “re-identification attacks”, where data from other sources can be combined to re-identify anonymised data and link it back to individuals.

OPINION by Udbhav Tiwari | Don’t rush into bad law: Giving India the data protection law it deserves

In the UK, personal location information has been extracted from anonymised datasets of public transit, while in Australia, individual health records have been mined from anonymised medical bills. These examples should caution Parliament from allowing the government to acquire anonymised data without further protections — if anonymisation techniques are imperfect, then forcing companies to create and share insecure datasets with the government increases the vulnerability of personal information and undermines data protection. Going forward, the approach to regulation of anonymised data must be contextual and sectoral, targeting sensitive areas such as healthcare and finance, and focus on data aggregators, which are often used for reidentification attacks.

The proposal to acquire non-personal digital data must also be seen in the context of the Centre's push towards using "big data" and "artificial intelligence" technologies within governance and planning systems. Indeed, the use of these technologies has the potential to increase government capacity and transparency, as well as provide insight for making informed decisions about economic and social planning. However, the provision ignores the multiplicity of existing and inchoate rights and interests that exist within non-personal data, particularly those which are created by private firms.

OPINION by Akriti Gaur | Deconstructing the proposed draft data protection law

While the Bill assumes that any such data held by any entity should be open to acquisition by the government, similar to a power of "eminent domain" over land, this is in conflict with existing legal systems such as copyright law and trade secret protections. Such databases are commercially significant to private companies, and a law to acquire them must consider how it will affect their commercial exploitation. Moreover, the overlap of these existing rights within government systems can jeopardise accountability and transparency by limiting the ability of citizens to participate in, understand or interrogate government decisions. The RTI Act, for example, may not apply to private databases protected by intellectual property law.

The unregulated use of private datasets in governance also has consequences for the people and communities who are being made more visible, or are being invisiblised, through the use of this data. While the government has historically relied on qualitative methods like the census for understanding populations and their needs, the shift to quantitative methods and "big data" relies upon private datasets, which were created and used in a completely different context and for different purposes.

OPINION by M Sridhar Acharyulu | When it isn't right to forget

Inevitably, such data will be incomplete for the purpose of governance, and replete with the biases of the private entity creating and analysing the data. In the absence of regulation which carefully considers its limitations, using such data to target beneficiaries or for economic planning can have hazardous consequences — including arbitrary denial or exclusion from critical government services; or through biased and discriminatory planning which replicates biased data and risks undermining important legal principles such as the right to equality before the law.

The regulation of non-personal data must take into account both the potential harms to individual privacy as well as the wider social and political consequences of such “datafication” of government. This is ostensibly why an expert committee was established to look specifically into the governance of non-personal data, even while the PDP Bill was expected to limit its scope to personal data of individuals. Instead of jeapordising both these goals and putting the cart before the horse, as the PDP Bill has done, the Gopalakrishnan committee must be allowed to deliberate and inform a public consensus on the appropriate models of governance of non-personal data.

The writer is a Mozilla Technology Policy Fellow