Comments on the draft data protection bill

Subhashis Banerjee

September 10, 2018

B. N. Srikrishna committee's draft data protection bill is expected to soon be tabled in the parliament after final touches. However, while several of the recommendations are welcome and laudable, the committee has ultimately failed to develop an effective vocabulary to deal with the complex subject. A data protection framework is unlikely to be grounded in reality without first formulating a data usage policy, and this has been a major lacuna in the data protection discourse.

The committee's inclusive functioning style and seeking of public opinion at all stages are commendable. Also, particularly welcome are its recommendations for a user centric design and setting up of an independent Data Protection Authority. The committee has further recommended that along with the private sector, the government too needs to be regulated; that intelligence gathering for national security requires a new law; and that the Aadhaar Act requires several modifications and provisions for regulatory oversight. Recognition of data portability as a right is another step in the right direction.

There are also the inevitable aberrations. Among the most glaring is the suggestion that the UIDAI be both the data fiduciary and the regulator for Aadhaar. There is also the curious suggestion that even though personal data can be transferred outside India, data fiduciaries be required to store a local copy. Does this benefit the individual or is it a surveillance requirement of the state? Besides, as many have pointed out, over dependence on consent and notice is unlikely to succeed in a country with low digital literacy.

But the clichèd vocabulary and the superficial treatment have been the most disappointing. For example, the concepts of fair and reasonable processing, purpose and collection limitation, notice and consent, data quality and data storage limitation, etc., are not new. They have largely failed to prevent identity thefts, unethical profiling and other privacy violations, anywhere. Dictums such as "personal data shall be processed in a fair and reasonable manner" are non-specific, and they do not adequately define the contours of the required regulatory actions. Moreover, as episodes like Cambridge Analytica demonstrate, 'harm' is often not immediately obvious, and causal links of 'harm' are not always easy to determine. Hence, ex-post accountability and punitive measures of the kind the committee has recommended may be largely ineffective, as they have been elsewhere, and the committee has not explored the ex-ante preventive measures adequately. There are three broad areas of omission where more due diligence is required.

First, a data protection framework is incomplete without an investigation of the nuances of digital identity, and guidelines for the various use cases of authentication, authorisation and accounting. It is also incomplete without an analysis of the extent to which personal information needs to be revealed for conducting businesses and during eKYC processes. In addition, effective protection requires an understanding of the possible pathways of information leaks; of the limits of anonymisation with provable guarantees against re-identification attacks; and of the various possibilities with virtual identities. Also required is an analysis of the possibilities with privacy preserving tools, techniques and protocols from computer science including hash functions; symmetric and public key cryptography; trust as negotiable protocols; selective disclosures; k-anonymity, unlinkability and untraceability; one-time, anonymous and dynamic credentials; zero knowledge protocols; and quantifying information leak about individuals using techniques of differential privacy.

Second, the committee does discuss AI and big-data analytics, but fails to define clear-cut guidelines for their safe use. It ends up vaguely suggesting that no processing of personal data should result in taking decisions about a person without consent, but does not provide guidelines about enforcement.

Most theories for improving state efficiency in delivery of welfare and health services using personal data will have to consider improved data processing methods for targeting, epidemiology, econometrics, tax compliance, corruption control, analytics, topic discovery, etc. This, in turn, will require digitisation, surveillance and processing of large-scale personal transactional data. Acquisition, storage and processing of personal health data will be crucial to such systems. There have to be detailed analyses of how purpose limitation of such surveillance - targeted towards improving efficiency of the state's service delivery - may be achieved without enabling undesirable mass surveillance that may threaten civil liberty and democracy. Much of the popular discourse seems to assume that no such balancing is possible, but naively and without basis.

Moreover, it does not appear that the committee has carefully evaluated the data processing requirements of the diverse private sector, spanning across health care, insurance, social media, e-commerce, etc., and how they may infringe upon privacy. While nobody wants episodes like Cambridge Analytica, Facebook and Twitter do have some redeeming features, and many of us do like the book recommendations of Amazon.

The committee needs to balance the seemingly conflicting requirements of individual privacy and the benefits of large-scale data processing, and it is not obvious that a trade-off is inevitable.

Third, a data protection framework is incomplete without defining the requirements and standards of access control, and protection against both external and insider attacks in large data establishments, both technically and legally. The computer science sub-areas of security and automatic verification will certainly have a lot to offer.

The participation of the civil society in the data protection discussions has been exemplary, especially in the wake of the Aadhaar debates and the privacy judgment. In contrast, it is the response from our institutions engaged in economics, public policy and computer science that have been muted. They have to now wake up and produce comprehensive studies and whitepapers on all aspects of data usage and data protection for the framework to be successful.