Comments on the draft data protection bill

Subhashis Banerjee
Computer Science and Engineering, IIT Delhi
New Delhi 110016

September 10, 2018

B. N. Srikrishna committee’s draft data protection bill and the accompanying report have been both heartening and disappointing. 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. This perhaps is also a failure of the way the discourse has taken shape. A data protection framework is unlikely to be grounded in reality without first formulating a data usage policy.

The committee has got several things right. The most noteworthy are its inclusive functioning style and seeking of public opinion at all stages. Also welcome are the recommendations for a user centric design, setting up of an independent Data Protection Authority (DPA), and the recognition that along with the private sector, the government too needs to be regulated. The committee has recommended that intelligence gathering for national security cannot be unregulated and it 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 are the recommendations that seem to suggest that the UIDAI be both the data fiduciary and the regulator for Aadhaar; and the curious suggestion that even though personal data can be transferred outside India, all data fiduciaries be required to store a local copy within our borders. Does this benefit the data subject 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 where digital literacy is low.

But the most disappointing have been the clichéd vocabulary and the superficial treatment. 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 the unacceptable levels of identity theft in the US, the unethical profiling of individuals, and the corporate panopticons like Google or Facebook from growing more or less unchecked. Dictums such as “personal data shall be processed in a fair and reasonable manner that respects the privacy of the data principle” 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 the 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. The devil, as they say, lies in the details; and 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 a detailed analysis of the extent to which personal information needs to be revealed for conducting businesses and obtaining services, both during transactions and during
eKYC processes. In addition, one needs to develop a deep 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 a detailed study of the 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 at length, but fails to define clear-cut guidelines for their safe use. It correctly observes that implementing collection limitation is difficult in the context of big-data, but ends up only vaguely suggesting that no processing of personal data should result in taking decisions about a person without consent.

It seems obvious that improving state efficiency in delivery of welfare and health services using personal data will have to be based on 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 and electronic health records will be crucial to such systems. There has 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 definitely needs to try 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. Again, 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 white-papers on all aspects of data usage and data protection for the data protection framework to be successful.