



Policy Opinion

Response on the Kris Committee Report on NPD Governance

BY

CENTRE FOR THE DIGITAL FUTURE

SEPTEMBER 2020



The focus of the Kris Committee Recommendations seems to be on three things:

- a. protection of opportunity for digital startups in India
- b. protection of SMEs from “digital marauders” and
- c. preserving “data sovereignty” to ensure that the economic opportunities that arise from digital data remain within the country.

The report categorizes non-personal data into three groups:

- a. Public Non-Personal Data
- b. Community Non-Personal Data
- c. Private Non-Personal Data

In particular, it identifies (irreversibly) anonymised personal data as non-personal data.

The basic approach taken by the Committee is that non-personal data available with an entity must be made accessible to all entities in a form determined by policy. Moreover, if an entity holds personal data, it must ‘convert’ it into non-personal data by a suitable process of anonymisation that is irreversible by anyone who accesses this data. The report also suggests that even if there are no immediate takers of the data, it has to be submitted to the government, in case some entity finds a use for it in the future. Finally, the government is willing to cover the cost, or fix a price, of converting the data in the possession of an entity in a format that can be used by others.

Three conceptual problems with the recommended framework

There are three conceptual problems with the Committee report. **First**, it ignores the incentives of the data holder to capture and store data --- the essential first step in generating a data driven economy. India is a “data poor” economy. This is not in terms of its potential to generate data, but in terms of the actual capturing of data. The data economy requires an innovative group of data suppliers who will feed the demand for data. The Committee focuses on the demanders of the data and, in the process, destroys all incentives for data suppliers to generate newer and newer datasets. The Committee’s approach would have been appropriate if we had all the data that could be generated in our economy; it is inappropriate and flawed if there are more data to be generated.

The **second** problem with the Committee report is its misinterpretation of the purpose of policy. The objective of policy is *not* to improve the welfare of a particular group within society but to improve the welfare of everyone in society. In case the improvement of one group has a spillover that improves the welfare of other groups or, at least does not harm them, then, of course, policy for a group within society is acceptable. Unfortunately, the focus on potential startup activity today, is hampering the startup activity of the future by destroying the incentive to generate new, or hitherto non-existent, data or datasets.

This second issue is detrimental not only in the long run, but could be so in the short run also. Consider a two-sided platform entity that holds raw/factual data on individuals/entities that have been generated by the transactions carried out by the players on each side of the platform. Suppose that this data or, datasets created from subsets of this data, generate value V for a second activity. According to the recommendation, given such a possibility, the platform data needs to be shared. It is also obligatory on the part of the platform to irreversibly anonymize this data. This last process is costly and, suppose, that cost is C . The question that comes up immediately is who bears the cost.

Suppose the cost C is less than V . Then it is immediate that the platform can negotiate with the entity that wants to use this data to distribute the surplus $V - C$ between them. In other words, there are enough incentives in the market place for such a value enhancing sharing of data to take place without a policy in place.

Now, consider the second alternative, wherein C is greater than V . In such a situation, the demander of this dataset will not have enough value to compensate the data generator and still enjoy a positive surplus. In such a situation, the market solution is one where the cost C will not be incurred and the value V will not be generated. It is the right solution as the data should not be processed as the cost is greater than the benefit. However, according to the Committee's recommendation, the platform will still have to make the data available to the second entity. This second entity will generate the surplus V for itself. The platform loses C and the rest of society, as a whole, will lose value equal to $C - V$. If the government reimburses the cost to the platform of anonymizing the data set, taxpayers will lose the amount C to generate value V for the second entity. Even if we count the second entity as being among the taxpayers, the taxpayers as a whole are still losing an amount $C - V$. If we mandate that the platform must make the data available to all, then there is no stopping such an outcome.

The **third** conceptual flaw is the concept of data sovereignty a) sovereignty is to be protected from other countries and not from own citizens and businesses. The paper seems to protect India's data sovereignty from Indians or those doing business in India; b) data sovereignty in the paper slides into sharing non-anonymous data with law enforcement agency. We understand that law enforcement agencies have a problems of accessing data. But non-anonymous data sets will not solve that problem at all.

Additional Concerns:

The report is weak on law and adherence to established economic principles and rests on a shaky conceptual framework

- a. It is not clear whether data is being treated as a resource, IP or infrastructure.

- b. The proposed regulation is needlessly complex in many places, creating confusion about categories of data, who is responsible for what, and how any regulation would function within existing system of regulations
- c. The data sharing provisions, propose sweeping powers for the government to demand data from companies for purposes of “economic development and domestic industry support” which could fundamentally injure data-driven business models.
- d. The proposal for another regulator seems unnecessary if there is a Data Protection Authority created via the Personal Data Privacy Bill – the data protection authority should be able to regulate issues related to personal and non-personal data, and creating more regulators would not lead to better regulation.

Other concerns are elaborated below:

1. Data definitions: India is probably the only government trying to create separate policy regimes for personal data and non-personal data. Definition of community data especially the inclusion of anonymized and aggregate personal data is problematic and would lead to overlaps in jurisdiction with PDP regulation. An integrated approach to data is preferable. The term “raw data” has not been defined and seems to have been assumed to be self-evident, which is not the case (see section on ‘Data Sharing’ below)
2. Stakeholder/ Authority definitions: The non-personal data policy is long and confusing, with multiple designates of holders of data that are not well defined and are possibly overlapping. Delineations of Data Principal, Data Trustee, and Data Trust are not clear. Nomination/selection of Data Trustee and Data Trust could lead to contention and conflict. For a given community, there could be multiple legitimate claimants to each of these roles.
3. Data Sharing: No clear definition of “raw” data. Separation between data and value addition is nearly impossible in practice – collection, standardization, collation, and organization of data itself is value addition. Moreover, if a company is reluctant to share data, it can adopt a valid definition of raw data that would make it useless for the recipient, thereby nullifying the stated intent of the report. The biggest concern for any company with data analysis as an important part of the business model is that the government may be looking to enshrine into law the ability to take proprietary data from companies, potentially calling into question the ability of some companies to run a business model related to data. Failure to distinguish between “public data”, and proprietary data of companies vitiates the recommendations of the committee. Fear of what might be defined as “public data” or “community data” and required to be mandatorily shared at no cost, would inhibit, not encourage data collection which is the foundation on which “Data Business” rests. An unintended consequence of the recommendations would be to undermine the entire data economy and business – the very things that the report seeks to stimulate and enable!

4. Mandating data portability is not always good. For example, Amazon in the US has supported some forms of data portability that would allow it to get access to data of supermarkets in the US that it could use in a better way given its already strong advantages in this space. The importance of protecting IP of all companies is important, including smaller companies with less sophisticated data operations. Such portability may well end up benefitting the large companies and perpetuating their dominance – the very ailment that the report purports to address. Conversely, any injury to individual SMEs caused by large platforms to enrich themselves at the cost of the former could be addressed by CCI. Defining IP of small companies with respect to data may be difficult and impossible to protect if it is readily portable by mandate.
5. Regulators: Multiplicity of regulators (PII and NPD) will complicate matters with jurisdictional conflicts. Getting regulators involved in ex-ante business decisions (eg. validating and then mandating a data request as legitimate) can never work. Besides, in India, regulation and court proceedings are already time-consuming and with technical complexity thrown in, will take even longer. Ease of digital business, both for the intended recipient and donor of the data will be the first casualty.
6. Competition: Enabling competition in the digital world is a priority. The inherent nature of the digital economy is that it is centred around large digital platforms. Dominance results from the network effect and cross-sectoral leveraging of data value. Vigilance on the part of the competition regulator to prevent abuse of that dominance, rather than trying to counter the network effect is a better way to capture the full potential of the digital economy and maximize innovation. Any ex-ante regulation may create, unwittingly or otherwise, a non-level playing field. Discrimination on the basis of Indian ownership ignores the possibility or even likelihood of subsequent sale to a foreign entity. Conversely, banning such sale will be counterproductive and discourage entrepreneurship.

We would like to conclude that the none of the stated purpose would be served by following the recommendations of the committee. In conclusion, the report of the committee is seriously flawed, suffers from many infirmities and will undermine the very objective that it espouses, namely, to nurture the data economy in the country.