Asia Internet Coalition ("AIC")Submission
on the
Report by the Committee of Experts on Non-Personal Data Governance Framework

Shri. Kris Gopalakrishnan,
Hon'ble Chairperson,
Committee of Experts on the Non-Personal Data Governance Framework, Ministry of Electronics and Information Technology,
Government of India

Cc:
Shri. Ajay Prakash Sawhney
Secretary, Ministry of Electronics and Information Technology

Shri. Rajendra Kumar
Additional Secretary, Ministry of Electronics and Information Technology

13 September 2020

Dear Sir,

Subject: Asia Internet Coalition ("AIC") Submission on the Report by the Committee of Experts on Non-Personal Data Governance Framework

The Asia Internet Coalition ("AIC") and its members express our sincere gratitude to the Government of India for the opportunity to submit comments on India's Non-Personal Data Governance Framework. AIC is an industry association that represents leading global internet companies on matters of public policy. To further its mission of fostering innovation, promoting economic growth, and empowering people through the free and open internet, AIC would like to present our comments on the Personal Data Protection Bill, 2019 to the Joint Select Committee ("JSC"). The AIC was established in 2010 as an industry association that promotes the understanding and resolution of Internet policy issues in the Asia Pacific region. Our membership comprises leading internet and technology companies, and we participate and promote stakeholder dialogue between the public and private sectors, sharing best practices and ideas on internet technology and the digital economy. In this context, we participate in calls for comments on issues impacting key internet and business governance issues, such as the regulation of data-oriented businesses.

In the month of July 2020, an Expert Committee ("Committee") as constituted by the Ministry of Electronics and Information Technology ("MEITY") released the Non-Personal Data Governance Framework ("NPD Report"). In this regard, we are grateful to be able to present our comments and recommendations on the same, and would also like to re-state our continuous support and assistance to the Indian government in its efforts to bring about this transformational change in the privacy landscape in India.

We believe that our comments may assist the Committee in ensuring that optimal value of data can be unlocked and intended intentions of representing all stakeholders in the markets are met.

Should you have any questions or need clarification on any of the recommendations, please do not hesitate to contact me directly at Secretariat@aicasia.org or +65 8739 1490. Thank you for your time and consideration.

Sincerely,

Jeff Paine
Managing Director
Asia Internet Coalition (AIC)

Detailed Comments and Recommendations

1. NPD Legislation

The Committee has proposed a new legislation for regulating NPD[1] to (i) create a framework to generate economic, social and public value from the use of NPD; (ii) incentivize innovation and creation of new products and services in India and encourage start-ups; (iii) make community, public and private data available for social, public, and economic value creation; and (iv) address privacy concerns from processing NPD and to examine the concept of collective privacy.[2]

However, the first three objectives can be adequately addressed through a policy framework. The National Data Sharing and Accessibility Policy (NDSAP), which the government notified in 2012, was designed to enable open sharing and leveraging of the vast quantum of data generated by various government departments and agencies.[3] The government has also set up the Open Government Data Platform under the NDSAP. Thus, there is an existing framework for leveraging public NPD which can be modified to further value creation, and the government, being the largest repository of data, can utilise this to meet the third objective proposed by the Committee.

As for opening up private NPD, the Committee appears to be motivated by anti-trust concerns. The Report mentions the "first-mover advantage" of data driven businesses which has created "significant entry barriers" for start-ups[4] and the need to balance out the bargaining power of "data monopolies" against Indian citizens, businesses, start-ups and the government.[5] This is best addressed through the Competition Act, 2002 (CA) which was enacted with the objective of preventing such anti-competitive practices and promoting the interests of the consumers. The Competition Commission of India (CCI) has in the recent past, probed into anti-competitive practices and made several significant rulings against such practices of data driven businesses. The financial disincentives associated with such rulings serve as a deterrent to data monopolies.

However, the Report fails to consider this entirely and justify as to why Competition law is inadequate and how non personal data regulation will encourage competitiveness including the fact that the report does not explain why an additional sharing-based framework is necessary to ensure competition. The Report also does not take into account the many negatives of regulating private NPD.[6] Curiously, though it recognises the role of "robust IP rights", and "data related privileges" as incentives for new business creation, the Report is silent when it comes to recognising the importance of these aspects for existing data businesses.

While the final objective of the proposed law is to address collective harms from processing NPD and to examine the concept of collective privacy, the Committee merely mentions that

this is an emerging concept that needs to be examined and defined in detail in the future.[7] Without a proper assessment of the subject, it prematurely calls for a legislative framework. While it is clear that all harms relating to re-identification can be dealt with under the Personal Data Protection Bill, 2019 (PDP Bill), a new regulatory framework is unnecessary.

While the Committee mentions that the NPD framework in the Report should form the basis for the new law on NPD[8], the Report only mentions several undeveloped concepts, without sufficiently considering all associated aspects and unintended consequences. Similarly, categorisations, as well as roles and responsibilities of different stakeholders have been included in an overbroad and vague manner. Many start-ups may not necessarily have the capacity to understand the implications of such policies. The report vaguely mentions that sharing of NPD may be useful for Indian entrepreneurs to develop new and innovative services and products. There is no evidence to back this argument.

While making the case for regulating data, in Chapter 3 of the Report, the Committee also states that "Market transactions and market forces on their own will not bring about the maximum social and economic benefits from data for the society. Appropriate institutional and regulatory structures are essential for a thriving data economy and a well-functioning data society". Instead of creating new regulations, perhaps the way forward is to have frameworks and legislations which facilitate easier collection of user data instead. E.g., the Account Aggregator framework which permits entities and even the user herself to leverage user's existing information for newer services. By making collection of data easier, without diluting the consent preferences of the user, entrepreneurs today have better access to information and therefore will have the ability to create their own data sets.

Further, instead of mandating private NPD to be compulsorily shared, perhaps the concept similar to Creative Common Licenses ("CCL") should be considered for this subject matter. However, this should continue to be on a voluntary basis. A CCL allows creators to share their works with the general public. This will make large pools of data and creation available in the public domain. There have been instances of companies using CCL enabled information from the public domain and creating feed for ML, e.g., IBM creating their "Diversity in Faces" dataset from photos people shared online with a CC license.

Furthermore, the Copyright Act, 1957 recognises databases as protectable literary works in which the author has exclusive rights, subject to its originality and licensing rights. In re Burlington Home Shopping v. Rajnish Chibber and Eastern Book Company v. D.B. Modak [Appeal (Civil) 6472 of 2004], Indian courts have held that the copyright vests with the creator of the datasets on account of the investment, capital and efforts put in by the company. The Copyright Act, 1957 also provides for compulsory licenses of literary works in certain limited situations such as published works that are withheld from the public. This further underscores the point that databases / datasets are proprietary to their owners and their licensing should either be on voluntary/market driven basis and if a scheme of compulsory licensing as the present Report seeks to do should be extremely limited and when a certain use case is not met with. According to this principle, copyright will be assigned to the work that reflects a certain

threshold of intellectual creativity. Even though the application of Copyright law to digital datasets and machine generated data is yet to be tested, the basic principle that Copyright vests with creator of the work should be taken into consideration. The report fails to take into consideration cases where data or datasets could reveal proprietary processes that organisations use to gather or generate raw data. There could also be conflict with legal obligations that consider such processes a trade secret.

Recommendations:
As a first step, the government should modify and implement the existing NDSAP framework to open up large sets of government NPD for public use as it is widely accepted that the government holds the largest amount of NPD which is not currently being utilised. Further, an incentive-based policy could be created for the private sector to share data more freely. Any such policy should be based on voluntary and free market principles as companies spend a lot of investment, capital and efforts for collecting and creating data sets.

The government should at best lay down standards for data sharing and not create multiple adjudicatory bodies for handling data sharing issues as the same will adversely impact innovation and the start-ups. The Government should look at defining non-personal data in a manner that protects intellectual property rights, serves genuine public interest and promotes innovation. There are countries like France who have put in place National Strategy on Artificial Intelligence policy, with the aim to share and pool their data with the state acting as a trusted third party. The European Union's Regulation on the Free Flow of Non-Personal Data[9] also recognises the free flow of non-personal data as a prerequisite of a competitive economy.

2. NPD Authority

The Committee has proposed setting up a new NPD Authority (NPDA) to administer the provisions of the proposed NPD law.[10] This appears to be based on the Committee's belief that (i) the regulatory authority would need specialised knowledge and the ability to keep pace with evolving technologies, (ii) the nature of tasks and focus of the authority would be different from existing regulators.[11] Yet, the Report does not clarify how the NPDA would be better equipped to deal with these tasks as opposed to an existing regulator. It only says that the NPDA will have some members with "relevant industry experience".[12] NPD is an umbrella set of data that can comprise of data from any number of industries. Not all of them can find representation in the constitution of the NPDA, and the Report provides no guidance regarding which industry would be the relevant industry or what experience would be considered as

relevant experience. Hence, it is unclear how the NPDA would have any advantage over existing sectoral regulators in respect of keeping pace with technology.

Most sectoral regulators are best placed to assess the best interests of each sector, be it economic, social or security related interests. They are qualified to understand the problems and the practical difficulties in each sector. In respect of the tasks and focus of the NPDA, while the Committee envisages that the NPDA will have an enabling role in addition to an enforcing role, i.e. enabling data sharing requests, proactively addressing market failures such as information asymmetry, enabling competition and level playing field etc.,[13] these tasks can also be performed by relevant sectoral authorities.

Further, under the CA it is the duty of CCI to eliminate anti-competitive practices and promote competition[14] and addressing monopolistic or exploitative practices resulting from lack of access to NPD is a subset duty, though the CCI may not be able to follow a proactive approach in this regard, currently. However, the CCI has a wealth of experience in dealing with anti-competitive practices and if the CCI were to perform the role of proactively identifying market harms, this experience could be put to use more effectively than if this role is performed by a new regulator.

Moreover, the NPD that is the subject of regulation may already be regulated by existing sectoral regulators that have been introduced keeping in mind the specific requirements of each sector such as the Reserve Bank of India, the Department of Telecommunications (DoT), Ministry of Defence (MoD), the Securities and Exchange Board of India (SEBI), the Insurance Regulatory and Development Authority of India (IRDA) etc. Similarly, e-commerce data is proposed to be regulated by an e-commerce regulator under the Draft E-commerce Policy being developed by the Department for Promotion of Industry and Internal Trade. Currently the technology sector may witness over regulation with at least four regulators Data Protection Authority (DPA), the E-Commerce Regulator, and the Central Consumer Protection Authority and now the proposed Non-Personal Data Authority (NPDA).

Overregulation undermines innovation and leads to conflict especially in cases of overlap of issues and jurisdiction. Example - the DPA under Section 50 of the PDP Bill is expected to regulate non-personal data, which is similar to the role of NPDA. The role of CCI is to address market failures and its role can be strengthened to address antitrust conduct in this sector. Therefore, the role of NPDA is mere duplication of the scope of work of these existing/proposed regulators. Creation of another regulator will not only lead to overregulation but also confusion thereby delaying the process for the companies who are looking to use the anonymised data. The Committee has maintained that the sectoral regulators may make additional regulations in a horizontal fashion.[15] This can result in over regulation and conflicts and the Report does not address how to resolve such issues. Also, the report does not address situations wherein there is an overlap between DPA and NPDA or for that reason specific sectoral regulators, which may lead to conflict.

Recommendations:
From an implementation point of view and in order to maximise efficiency to achieve the objectives outlined by the Committee, it is recommended to follow a sectoral approach instead of setting up a separate NPDA. Depending on the need for regulating NPD in a particular sector, the existing sectoral regulators, if any, or the relevant government department or ministry can issue appropriate rules and regulations. Where needed, new sectoral regulators could be set up. For instance, the DoT and the MoD understand the economic, strategic and security interests for the telecom and defence sectors and should therefore be solely responsible for the data generated within those particular sectors. For health data, it may be appropriate to consider setting up a specific sectoral regulator. The government can consider releasing an overarching policy on voluntary NPD sharing, which can be used as a common minimum guideline for relevant sectoral regulators to seek sharing of certain kinds of data.

3. Assessment of when data is relevant for competition

The main recommendation to regulate non-personal data stems from correcting market imbalance. However, it is pertinent to note that data is not necessarily a barrier to entry in the market. This is because, as the Report also rightly notes that the "world is awash with data" and data is non-rivalrous in nature.

Thus, it is important for any regulation to draw a balance between over-regulating data which may harm innovation and creating checks and balances to ensure that any harmful effects are addressed. Regulation of data for promotion of competition thus needs to be limited to those cases where lack of access to data could cause entry barriers and thus, harm competition. This is supported by developments in other jurisdictions.

Thus, as a starting point, therefore, when creating any framework for non-personal data, it is imperative to fully and comprehensively identify when data as essential input which could result in foreclosure. It is only in this scenario where data regulation may be required. In the absence of this essential framework, data regulations are likely to harm competition.

Recommendations:
Any framework developed should be proportionate to address the competition harm posed by data access. For this, there needs to be detailed study to fully understand under what scenario would data cause competition harm.

4. Addressing the need for a 'Data Business'

The Report introduces the concept of "data business" as one involved in collecting, processing and storing information as per the threshold criteria[16]. It is important to note that the objective of the data businesses is to provide open access to metadata, and regulated access to underlying data. Not only is there "mandatory registration" of data businesses accompanied with tracking the nature of data businesses, kinds of data collections, processing, use, etc. but there are also several operational changes for businesses, which will now have to segregate data collected into different categories of personal and non-personal. Additionally, the Report seeks to mandate consent for use of anonymised data.

The Committee recommends mandatory sharing of sufficiently anonymized raw/factual NPD (private, public and community) at no cost.[17] Data Businesses would be required to submit meta-data about the data user and community from which data is collected, with details such as classification, closest schema, volume, etc. based on a directory of data classification and schema published by the NPDA. Indian citizens and organizations may access such metadata. Subsequently, they may request for underlying data for the following - sovereign purposes such as national security, legal purpose, core public interest such as research and innovation, better delivery of public services or economic purposes such as encouraging competition.[18]

However, the Report does not provide details of how the sharing mechanism would work and has left it to the NPDA to certify rules and technology framework for data sharing. While the Report considers certain checks and balances to sharing[19], they seem to be largely from the perspective of data safety and do not consider the rights of the Data Businesses. In contrast, the Open Data License (ODL) issued pursuant to the NDSAP considers several protective measures such as rightful attribution, disclaimer of warranty for errors and omissions etc., and if the conditions are violated, the license to open data would terminate. The ODL also excludes *inter alia* data that is subject to intellectual property rights and sensitive data (as defined under law). The Report does not consider any such protective measures.

Compulsory sharing of NPD without taking sufficient safeguards to protect Data Businesses from the liability arising out of such sharing is problematic. The Report also does not touch upon such accountability principles that are necessary pursuant to data sharing, which may further disincentivise companies from investing or carrying out business in India. Bad actors can deploy methods to re-identify individuals and breach privacy and use it for irresponsible activities without accountability.

The Report provides that when there is a value addition to data, private companies can charge for the NPD that it shares depending on the level of the value addition. For non-trivial value additions, NPD sharing may be mandated but subject to fair, reasonable and non-discriminatory (FRAND) based remuneration, for an increased value addition the charges will be determined by market forces, and for "high value add" NPD, the private companies will have full discretion.[20] While the Committee recommends affixing costs proportionate to the value

addition, there is a lack of clarity on what would constitute "non-trivial", "increased" and "high" value additions. This is in contrast to how the pricing mechanisms were envisaged under the NDSAP, where the government proposed to open its own data to the public. The pricing under the NDSAP could be decided by the data owners and as per the government policies. The Report does not hold the private stakeholders sharing their data to the same pricing standard. The Report also ignores the costs associated with collecting data and the compliance costs involved in catering to data sharing requests.

Mandatory sharing of data is based on a wrong presumption that it will drive competition. It will in fact create an anti-competitive environment. Further, sharing of data by businesses may possibly erode the value of such data. It will have adverse business and cost consequences to established businesses operating in India. Mandating opening access to data ignores the significant cost incurred by businesses to collect such large amounts of data. Furthermore, the powers of NPDA to mandate data sharing can be misused in ways to undermine confidentiality and integrity of NPD. For instance, if entity A shares a dataset with entity B under contract and confidentiality; any requester can mandate even entity B (through NPDA) to provide the requested data, in case entity A refuses to furnish the requested data. Further, datasets created and owned by entities may constitute valuable trade secrets that are subject to a separate legal regime and need to be protected as such. The commercial interests in such data and the requirements of the industry must be considered, similar to how the NDSAP takes into account the pricing of data to be shared and allows the data owners to decide the pricing themselves.[21]

Technology companies invest heavily in data science technologies to make use of raw/factual data and a mandatory open data sharing policy will negate the competitive advantage held by these companies and make their investment worthless. Companies will be disincentivised from investing in technological advancements and R&D relating to data science, which will have a negative impact on the economy of the country. It will also impact new investments in India as similar open data sharing policies do not exist in jurisdictions comparable to India. This may lead to a flight of capital out from India to such countries, which recognise the right of companies to proprietary data. The disclosure of such data will result in a loss of competitive advantage and competitors all around the world can leverage any proprietary data shared to advance their position. This will starve the start-up ecosystem from much needed foreign funds as investors will flock to other investment destinations. They are also better placed to decide the pricing for any data that is to be shared.

Diluting the existing protections for proprietary information can also be considered a breach of India's obligations under WTO's Agreement on Trade-Related Aspects of Intellectual Property Rights and the General Agreement on Trade in Services.

Mandating compulsory sharing of NPD for private organisations, is in fact creating compulsory licensing provisions for copyright through a new legislation. India already has a comprehensive Copyright Act, 1957. A new legislation for this subject matter is therefore not required. Also any incremental laws on compulsory licensing etc, should be continued to be covered under Copyright Act, 1957. A parallel can also be drawn to the provisions of compulsory licensing under the Patent Law in India. The Patent Act allows the government to permit someone else to produce the patented product or process without the consent of the patent owner under certain circumstances. Therefore, just as with patent laws, any compulsory licensing provision

on the subject-matter of copyright should be under the Copyright Act, 1957 not a separate legislation. Also, the exception of compulsory licensing for patents was also included in the WTO's agreement on intellectual property – the TRIPS Agreement. However, a similar exception is at present not provided for copyright. India is a signatory to TRIPS, which extended copyright protection to computer databases. Therefore this regulation poses a challenge of demarcation between non-personal data that cannot be shared, and non-copyright non-personal data that can be used as a public resource. It is vital from an ease of doing business point of view that the international community also recognise this principle and adopt it holistically. Therefore, the IPR laws are a more appropriate legislation for deliberation and decision on provisions relating to compulsory licensing of copyrightable datasets.

Our concerns with this are as follows:

i. Concerns on Proprietary Rights over Data: We submit to the Committee that the purpose, the manner and the use of the data collected is a proprietary information, associated with such decisions arrived made by the company on the basis of their market dynamics and operations. Thus, requiring businesses to disclose such important information would heavily impact innovation as now businesses would no longer have an incentive to devise a unique data-processing and use methods and techniques for the information that is essentially important to them. Give this, the need for disclosures relating to processing and use of data is unclear.

ii. Metadata Directory and Mandatory Registration: An important factor to be considered in the above framework is that the Report does not provide guidance on how the "threshold" for "data businesses" will be determined. There are several compliance requirements that have been suggested for such 'data businesses' and these include mandatory registration, open access to metadata, disclosures relating to collection, storage and processing of data. It is very likely that since most companies deal with large volumes of non-personal data, a very large section of businesses will come under the definition of 'data businesses' and be subject to these compliances. Any onerous mandatory registration framework of this kind is a significant disincentive to businesses in India and we recommend that it should be reconsidered.

iii. Operational Changes: With the changes recommended by the Report with regard to the different operational requirements, businesses will have to undertake a complete overhaul of their current operational practices in relation to how they collect and process the data. As mentioned in the Report, data will have to be segregated into categories of personal data and NPD and this then is further divided into sensitive NPD and critical NPD, etc. This will be in addition to tracking value additions to the data for the purpose of data sharing. Further, the businesses will also have to appoint relevant personnel as required by the Report (a data officer for compliance with the proposed framework).

Recommendations:

Private NPD should not be brought within the ambit of mandatory sharing. Self-regulation by industry/ voluntary sharing is a much more appropriate mechanism when it comes to NPD owned by private businesses. A proposal on this front is to incentivise data portability so that data subjects/principals are in complete control of their data and can take it out of an entity / transfer from one entity for use in different use cases, thereby meeting/ supporting the needs of start-ups etc. The Account Aggregator framework, for e.g., is a regulation that enables entrepreneurs and users to leverage existing data to help create new datasets.

Over the years, private companies, organisations and institutions have voluntarily shared large volumes of data sets available for processing and use by the public without any intervention or compulsion from the state to disclose them to maximize their potential benefit for the public.[22] These organizations have internal checks and balances to determine the sensitivity of datasets owned by them and to ensure that their own proprietary interests are not compromised when making decisions on sharing. It is not possible for governments to replicate these checks without accessing proprietary information and tools belonging to these organizations which they do not wish to share.

The government can instead consider incentivizing the sharing of NPD sets by private parties by various policies and programs. The NPD law must also provide detailed guidelines for framing data sharing requests to ensure that these are reasoned and specific, and do not impose onerous obligations on private companies to routinely transfer large quantities of NPD. There is also a need to protect and shield the Data businesses against any liability that may arise as a consequence of sharing of NPD beyond the indemnification against vulnerabilities that the Report contemplates for adhering to standards.[23]

a. Registration of businesses should be reconsidered as most businesses would now fall under this definition and will have to follow onerous compliance requirements.
b. There should be no disclosure requirement for businesses as there may be adverse impacts of disclosing 'business-sensitive' information and trade secrets.

5. Mandatory Nature of Data Sharing

The Report submits a case for mandating sharing of NPD with the broader public. The reason for the same is the need to enhance innovation. The Report suggests three purposes for which data can be shared: sovereign, public interest, and economic purpose. The Report provides different types of data sharing based on value-addition. Raw data is shared on a no-remuneration basis, whereas if the value-addition is non-trivial, there may be FRAND (fair,

reasonable, and non-discriminatory) remuneration. At a high level of value-addition, the private organisation can decide how to use the data.

Our concerns with this are as follows:

i. The concept of "mandating" data sharing: We recognise the efforts made by the Report to incentivise open data which is mandating of data sharing. We would like to bring attention to the point that such mandatory data sharing has not even been incorporated by global NPD governance frameworks. We believe that making it mandatory and not voluntary will have a significant impact on proprietary information of businesses which can be subjected to malicious data requests as a result of mandatory data sharing policies. Post this framework, any company can access metadata and request for the underlying data, which may possibly contain sensitive business information and trade secrets.

ii. The broad scope of mandatory private data sharing: B2B Infrastructure-as-a-service (IAAS), software-as-a-service (SAAS) and platform-as-a-service (PAAS) providers simply furnish other enterprises, including start-ups, with on-demand platforms, tools and services instead of carrying out data processing activities of their own accord. They do not collect the types of raw[1] data that this recommendation mandates sharing. Enterprise service providers have a contractual obligation to their clients (data fiduciaries) to store their customers' data (both personal as well as non-personal) securely and ensure no third-party access. Hence, if the NPD governance framework mandates them to register with the NPDA and share the raw data of their clients (data fiduciaries), it would require a complete overhaul of existing contracts, compliance protocols and increase the overall ease of doing business in the entire ecosystem. Therefore, B2B enterprise service providers (IAAS, SAAS, PAAS) should be explicitly excluded from the scope of the non-personal data governance framework. In addition to the concerns related to enterprises, people may not partake in certain services such as social media if they know their data will be provided.

iii. Ascertaining the "value" of data by a regulatory authority:

   a. As mentioned in the Report, the 'raw data' is to be shared without remuneration, however we submit that this would create disincentives for businesses from incurring costs for collection of such data, as the costs of such collection or cleaning up the data to make it usable are not considered at all.

   b. There is also no clear legal framework for determining what "non- trivial" value is add. The regulatory authority determining this and mandating sharing of this information on a FRAND basis may turn out to be a major concern, as it may not be equipped make such a determination relating to

whether the value addition is trivial or not. Further, the FRAND concept is not easy to import in this situation because it is a very specific idea that arises in particular cases – such as for standard-essential patents where the patent holder agrees to license the patent on fair, reasonable and non-discriminatory terms. Without any of the context surrounding this concept, it should not be imported into a new framework.

We submit that the process of value-addition for any company is based on its capabilities, resources, business strategies, etc. Such valuation also varies from sector to sector and transaction to transaction. This flexibility should not be compromised.

c. Any such mandatory data sharing regime may affect the rights of investors to their own property and this might heavily undermine any incentives to anonymise data or scale up data operations in India. Technical challenges with Data Sharing: The Report does not consider or even delve into details pertaining to the accuracy, completeness and consistency of the data that have to be checked by both the parties sharing and collecting the data. There are various challenges pertaining to anonymisation that would be rampant and the aggregation of data, especially in the context of sharing anonymised NPD that has the risk of re-identification. We submit that the creation of shareable data sets, especially those that are usable by the entity receiving them is a huge technological concern as any raw data that is actually cleansed and labelled and placed in a certain format actually entails a high degree of value addition, in the absence of which, such data is unlikely to be usable.

iv. Overburdening of the NPDA: The Non-Personal Data Authority (NPDA) has been now been vested with the power to determine the public interest benefit of data sharing, in case a request for data sharing by a company is denied. We submit that the parameters against which such "public interest benefit" will be ascertained are not clear and not mentioned in the Report. Further, we submit that the NPDA may not possess the required expertise to make the extremely subjective as well as technical assessment of such a benefit and how it compares with the investment and value addition made by businesses in developing the datasets in question.

Recommendations:

- Mandatory data sharing should be reconsidered, and voluntary open data policies should be encouraged;
- Enterprise service providers should be excluded from the scope of mandated raw non-personal data sharing;
- The NPDA should not be tasked with ascertaining value addition to data held by companies;
- Any marketplace for data should be industry led and valuation of data in different contexts should be left to be determined in context specifically by the parties to the transaction.
- The scope of regulations (if any) should only be limited to creating parameters in determining such value. This can only be done post a detailed understanding of how economic value can be attributed. Detailed assessment also needs to be undertaken prior to arriving at any such framework.
- The report states that in the spirit of "maintaining checks and balances" for data sharing, the contract between the cloud provider and data business must comply with terms of storage, processing and usage of data as specified by NPDA. There is no clarity at this stage on how the NPDA is going to ensure this is done (if they wish to inspect contracts, which is eventually onerous). The Bill also sets out the "tools" and "expert probing (procedures)" that the NPDA is going to have the power to wield, to ensure compliance, which is very invasive as this would require a huge degree of access into companies' systems.

6. Data Localisation and Sensitivity of NPD

The Report mirrors the principle as recommended in the Personal Data Protection Bill, 2019 (PDP Bill). It suggests that different categories of NPD be treated differently in terms of localisation and transfer. For example, 'sensitive' NPD and 'critical' NPD may only be stored in India, processing of the latter may only take place in India.

The Committee has recommended that NPD should inherit the sensitivity characteristic of the underlying personal data from which it is derived.[24] Accordingly, there would be 3 categories of NPD - (i) general NPD, (ii) sensitive NPD and (iii) critical NPD, derived from personal data, sensitive personal data and critical personal data respectively.

The Report provides that general NPD may be stored and processed globally, however, Sensitive NPD can be transferred overseas but must be stored in India.[25] Critical NPD can only be stored and processed in India.[26] It further states that NPD may be sensitive based on (i) national security or strategic interests such as vital infrastructure; (ii) business sensitivity or confidentiality; or (iii) risk of collective harm.[27] The Report envisages that the NPD regulation will continue to apply to the regulation of community NPD and public NPD transferred

overseas.[28] Businesses that transfer community NPD or public NPD overseas will be responsible for complying with data sharing requirements.

The scope of sensitivity of NPD under the Report is very broad. It could depend on dynamic factors such as national / security or strategic interest and risk of collective harm, which leads to uncertainty around the classification and treatment of such data. The inheritance of sensitivity from personal data could also involve a large compliance burden since the current draft of the PDP Bill envisages that the government has the ability to expand the scope of sensitive and critical personal data. Further, given that NPD implies that it is already no longer relatable to the data subject, there is inherently neither a need to categorise such data as 'sensitive' or 'critical NPD nor to regulate them as such. This would also lead to several ambiguities regarding the treatment of any given data set, given that any data set is bound to include several data points. For instance, given that the sexual orientation of a data subject is considered 'sensitive personal data' under the PDP Bill, any dataset on purchases made by users for a given product, which includes gender/ sexual orientation of the individual as one of the data points, would likely, unduly subject the entire dataset as 'sensitive NPD'.

Further, the Report recommends that the consent of data principals should be obtained for anonymization and usage of the anonymized data, while obtaining their consent for processing their personal data, and that appropriate standards of anonymisation be defined to prevent or minimize the risks of re-identification.[29] If this recommendation is implemented, the notice given under the PDP Bill will have to also include a specific provision to obtain consent for anonymization and the uses of anonymized data. The draft Report overlooks the fact that all NPD is not based on underlying personally identifiable data, hence obtaining consent would be irrelevant. Separately, the Committee needs to consider that obtaining consent from individual data principals / subjects runs the risk of converting the NPD to identifiable information that would be subject to the PDP Bill. However, this recommendation is unnecessary since de-anonymized data would be subject to the PDP Bill and the PDP Bill has specific provisions prohibiting re-identification of anonymized information. Further, the PDP Bill contemplates that the Data Protection Authority (DPA) would stipulate standards of irreversibility for personal data to be considered anonymized, Therefore, if the NPD regulations also stipulate standards for anonymization, there is a possibility of conflict / overlap in the event these regulations are not harmonized. This may lead to uncertainty for businesses and increased compliance costs.

Our concerns with this are as follows:

There are severe economic impacts of localisation as a policy instrument. It is pertinent to note that whenever the categories of data are to be localised, if at all there are any perceived advantages to such localisation, such categories must be narrowly defined. The categories of data to be localised in the NPD Report are broad in nature, as will be discussed in later sections, which is principally against the idea of localisation being a policy tool to be used sparingly. The broad categories of data that are likely to be included within the ambit localisation measure would result in large amounts of data requiring localisation. This would ultimately be barriers to free movement of data that would have allowed businesses both Indian and otherwise to make use of global infrastructure, all the while increasing costs for

users. As explained there are unintended impacts of creating barriers to entry for new players attempting to enter the market, thereby affecting competition in India. The possibility of retaliatory localisation measures by other countries also cannot be discounted.

Recommendations:

- Localisation requirements should not be included in the NPD framework. Alternatively, the NPD framework should look at having a a copy of Indian citizens' data in country, instead of a blanket data localization requirement.
- The Committee should clarify whether there are other objectives to defining NPD as sensitive or critical and consider removing the concept from the NPD framework entirely. The NPD framework should not incorporate a requirement to obtain consent for usage of anonymised data.

7. Data Custodians

The Report introduces the concept of a data custodian, whose role is similar to that of data fiduciary under the PDP Bill. Data custodians must collect, store, process and use NPD in the best interest of the data principal. Data custodians have a duty of care, in respect of community NPD, to the concerned community.[30] Such duty of care will be specified in the proposed NPD law and will include anonymisation requirements, protocols for data sharing etc. The Report provides that 'duty of care' must be operationalised through the best interest standard in order to prevent harms to communities and individuals from processing NPD.[31] Businesses may make data sharing requests to the data custodian[32]. If the data custodian refuses to share the data, the request can be made to the NPDA and if the NPDA determines that the request is genuine and beneficial it can require the data custodian to share the raw/factual data.[33] Equating the obligations of a data fiduciary under the PDP Bill which relates to protection of personally identifiable data to that of obligations of entities that process NPD is not justified since the nature of data being processed is different. NPD may be used for business development or market strategies which do not directly benefit the data principal and may therefore hinder the ability to process such data effectively. Further, the standard for ensuring the "best interest" of the data principal is presently unclear, and the Committee has recommended that this be detailed in the NPD regulation. The best interest principle and a duty of care principal are not the same, yet the Report equates the two. The duty of care is a minimal standard of care owed to the person concerned. In tort law, it is established as a negative obligation to ensure that no acts are done which may affect the principal negatively. The best interest standard in administrative law or trust law requires acting in a manner to benefit the person concerned, and not just ensuring that no negative impact is felt by the principal. While this may work for personal data protection, it may not work in the context of NPD where the context is to unlock the value.

Recommendations:

Data custodians should not be treated on the same footing as data fiduciaries. They should not be imposed with an obligation to act in the "best interest" of the data principal. Especially in the case of community NPD where there could be multiple communities involved, positing a fiduciary obligation on data custodians would unfairly burden them.

8. Extra Territorial Application:

Given that data collected by foreign bodies pertaining to Indian natural persons and communities will be governed by Indian law, the future regulation proposed in the NPD Report will likely have extraterritorial application. The Report does not consider conflicts between Indian and foreign law, to the extent that non-Indian entities will be governed by data governance regulations in their native jurisdictions as well as Indian law. Additionally, contracts entered into by businesses and data principals generally contain choice of law provisions, the Report does not delve into this concept. Extraterritorial application of law will impact the freedom of entities to enter into such arrangements and will therefore impact India's objective to be a business and "investment-friendly" destination.

Recommendations:

We recommend that the extra territorial application in the proposed law should be reconsidered altogether in the proposed law.

9. Definitional Concerns

The Committee recognises that NPD can be defined in multiple ways. Therefore, it has adopted the approach of defining categories of NPD - (i) public, (ii) private, and (iii) community NPD.

NPD collected or generated by the government / agency of the government will amount to public NPD as per the Report[34] with the exception of data that is explicitly treated as confidential under any law. However, confidentiality obligations in the commercial context are largely driven by contractual obligations. Without a clarification that information that is contractually confidential is an exception to public NPD, any information provided to any government department or agency in the course of collaboration with the government by a

private entity or individual is also at risk of being categorized as public NPD. This could disincentivize skilled private players from bidding for government projects.

The Report defines private NPD to mean NPD collected or produced by persons or entities other than the governments, the source or subject of which relates to assets and processes that are privately-owned by such persons or entities, and includes those aspects of derived and observed data that result from private effort, insights involving application of algorithms or proprietary knowledge, and data included in a global dataset and which is collected in foreign jurisdictions.[35] This would mean that all Indian data custodians, including those collecting NPD of foreign individuals, and all foreign data custodians collecting NPD of Indians would be subject to these regulations. Therefore, the presence of Indian NPD in a dataset will invoke applicability of the NPD regulations, even for datasets created by foreign entities which largely contain NPD of foreign individuals. This would also act as a strong deterrent for several service providers that operate at a global level, or even smaller players from offering their services to customers/ users based in India, since even any incidental collection of NPD of users in India is likely to subject the entire dataset(s) of such organisations to the proposed requirements and obligations.

The Report further defines a community as any group of people that are bound by a common interest / purpose, and involved in social and/or economic interactions such as a geographic community, a community by life, livelihood, economic interactions or other social interests and objectives, and/or an entirely virtual community.[36] It goes on to define community NPD as NPD about inanimate and animate things or phenomena – whether natural, social or artefactual, whose source or subject pertains to a community of natural persons. It excludes private NPD and qualifies that only raw/factual data without processing or derived insights is considered community NPD. However, the examples of community NPD presented in the Report include datasets collected by the municipal corporations and public electric utilities which should have been considered as public NPD. The examples also include information collected by private players such as ride-hailing companies. While the Report qualifies that in such cases, the raw/factual data without processing /derived insights will be characterized as community data, it would still be considered as private NPD going by the definition in the Report. Thus, there is considerable ambiguity and potential for overlap in the categorization of community data.

Further, just as there can be an infinite number of NPD data sets, there can be an even greater number of communities since each data set could result in interest from multiple communities. However, it may not be necessary and also practically not possible to give rights to all such communities. There can also be several subsets of communities, within any given 'community' as defined. Without identifying a community, it would not be possible to identify the rights that such communities should exercise as not all communities can be equal or enjoy equal rights. The obvious overlaps amongst different 'communities', would invariably lead to disputes over the manner in which NPD of those communities are to be handled. Moreover, this would also lead to contesting claims from different bodies/ organisations over the right to be the 'data trustee' of any given community. The lack of clear guidelines with respect to the definition can lead to undue interference with established business practices and create uncertainty since there is no practical definition for any business to implement. Notably the idea of regulating

community data was introduced into mainstream discourse on data regulation by the Committee headed by Justice B.N. Srikrishna.[37] However, even this reference was in the limited context of potential collective privacy harms.

Comments on the distinct definitions:

a. Sensitive & Critical Data: The Report introduces the concept of 'sensitive' NPD and 'critical' NPD. Sensitive NPD has been understood to mean NPD relating to natural security, strategic interests, business-sensitive or confidential information, or anonymised sensitive personal data which bears a risk of re-identification. We submit that the report does not accommodate any distinction in the treatment of sensitive NPD as any different from other NPD. It is not clear why these separate categories are created, considering that it is not for the purpose of exemption from data sharing and there is no specific purpose that is served by this distinction except additional localisation mandates.

b. Community Data: Community data is defined as the data which originates from a community of natural persons. We submit that this concept has not been implemented with a reasonable amount of success in any other jurisdiction as of now, and the concepts associated with community ownership of data require further scrutiny on account of being inoperable in the present form.

c. Data Custodian: A Data Custodian is defined as being responsible for collection, storage, processing, use, etc. of data in a manner that is in the best interest of the data principal/ community. However, the NPD Report fails to clearly define "best interest". Thus this opens up the possibility of misuse, it also leads to communities becoming politically mandated with the benefits not accruing to the all sub-communities.

d. Data Trustee: A Data Trustee is to exercise the data rights of a data principal group/community and appropriate community data trustee. The definition of "closest and most appropriate representative body" is vague and liable to misuse and confusion. Further, we submit that there is a possibility that most of the Data Trustees would be government entities or would reflect the interest of the dominant groups within the community, which the NPD Report does not seem to have considered.

Recommendations:

Any framework for governance of NPD should provide sufficient and clear guidelines for data classification that would enable businesses to implement the classification. Given the complexity involving community data, it is advisable to remove the concept from any policy regarding NPD. As and when required, specific communities can be identified based on policy objectives and the government and the relevant sectoral regulators can still seek appropriate NPD as required by such policy. Foreign data sets should be excluded from the NPD framework since the object of the framework is to unlock the value of data related to India. It also needs to be clarified that information that is treated as confidential contractually is a valid exception to public NPD.

- The categorisation of sensitive and critical NPD must be reconsidered, unless it is self-determined by industry and is exempted from data sharing norms.
- The framework for community data are worthy of a separate and detailed assessment, but their current inclusion in the Report is not sufficiently clear in terms of rationale, framework, and exercise of rights. The Report should observe the evolution of this concept globally to assess the need and relevance to the Indian market in the longer run, and in consultation with all stakeholders.
- There should be better clarity provided on the relationship between a data trustee and a data custodian.
- An enabling framework defining the nature and purpose of data trustee should be created with no mandatory requirement of their creation

10. Conclusion And Recommendations

We recognise that the NPD Report is well intentioned in wanting to create a framework for the regulation of NPD. However there are certain aspects and under- defined concepts that could be detrimental to such regulation We recommend that open data policies, like Data for Good by Facebook, and an open digital economy should be encouraged, which allows free movement of data and technology in a way that is compatible with the needs and interests of industry.

[1] Paragraph 8.3 (ii) of the Report.
[2] Paragraph 3.8 of the Report.
[3] Paragraphs 1.3, 3 and 4 of the NDSAP. Also see https://dst.gov.in/national-data-sharing-and-accessibility-policy-0
[4] Paragraph 3.7 (iv) of the Report.
[5] Paragraph 3.7 (v) of the Report.
[6] See the discussion in paragraph 5 of this position paper.
[7] Paragraph 3.8 (iv) of the Report.
[8] Paragraph 8.3 (ii) of the Report.
[9] https://ec.europa.eu/digital-single-market/en/free-flow-non-personal-data
[10] Paragraph 8.1 of the Report.
[11] Paragraph 8.2 of the Report.
[12] Paragraph 8.2(x) of the Report.
[13] Paragraph 8.2 (iii) of the Report.
[14] Section 18 of the CA.
[15] Paragraph 8.2 (ii) of the Report.
[16] Paragraph 6.1 of the Report.
[17] Paragraph 7.4(iii) of the Report.
[18] *Ibid*.
[19] Paragraph 7.4 of the Report.
[20] Paragraph 7.4 (iii) of the Report.
[21] Paragraph 11 of the NDSAP.
[22] Over the years, private enterprises have voluntarily made large data sets available, which helps empower other organisations to innovate and develop programs that produce socio-economic value. For example, firms make data sets available to local governments and other partners to inform sustainable mobility projects and to help in urban planning. Businesses have also made open data sets and tools available to third parties, free of charge, so they can leverage the power of emerging technologies like AI/ML and cloud technologies. These tools help produce valuable insights around image processing, video annotations, natural language processing and search trends around the world. It is important to note that these tools and datasets have been made freely available to third parties after a careful review of the sensitivity of the data involved and to ensure conformity with intellectual property laws.
[23] Recommendation 6, Para ix of the Report.
[24] Paragraph 4.5 of the Report.
[25] Paragraph 7.6 of the Report.
[26] *Ibid*.
[27] Paragraph 4.5 of the Report.
[28] Paragraph 7.6 of the Report.
[29] Paragraph 4.6 of the Report.
[30] Paragraph 7.4 (iii) of the Report.
[31] Paragraph 4.8 (iii) of the Report.

[32] Paragraph 7.5 (iii) of the Report.
[33] Paragraph 7.5 (iv) of the Report.
[34] Paragraph 4.2 of the Report.
[35] Paragraph 4.4 of the Report.
[36] Paragraph 4.3 of the Report.
[37] Paragraph IV, Page 45 of the Report on A Free and Fair Digital Economy, by the Committee of Experts under the Chairmanship of Justice B.N. Srikrishna.