ITI Promoting Innovation Worldwide

ITI Comments on the Report by the Committee of Experts on Non-Personal Data Governance Framework

September 9, 2020

The Information Technology Industry Council (ITI) is the premier global advocate and thought leader for the information and communications technology (ICT) industry. ITI's membership comprises more than 70 of the leading technology and innovation companies from all corners of the tech sector, including hardware, software, digital services, semiconductor, network equipment, and Internet companies. Our membership is global and consists of companies from around the world, including the United States, India, Japan, Taiwan, the European Union (EU), and South Korea. Because of the broad nature of our membership, we bring a global perspective that takes into account the diverse views of our membership when we engage with policymakers around the world to work toward policy that promotes innovation and inclusiveness in the global economy.

We are pleased to have the opportunity to formally provide our thoughts on the Report released by the Committee of Experts on Non-Personal Data (NPD) Governance Framework (the Report). During this time of great social and economic uncertainty, this is indeed an opportune moment to be exploring policy areas that are ripe for promoting innovation, development, job creation, and growth. India has an opportunity to show world leadership by embracing a digital policy that encourages investment, innovation, openness, and collaboration.

In the spirit of working with the Expert Committee and the Government of India (GOI) to enable society to derive maximum benefit from NPD, we offer our thoughts on the Report and the recommendations therein because we believe that, in its current form, the regulatory framework outlined in the report would not only negatively impact the Indian data market, but entire sectors of the Indian economy that rely on and benefit from large data sets. The Report does not recognize that, by the GOI's own estimates, India's data-driven economy has grown steadily over the years across sectors and is now the world's second largest start-up hub – all without a data sharing framework in place. Such a regulatory framework would *discourage* rather than *encourage* the very partnerships between foreign and domestic firms that have already contributed to India's successful startup culture. Moreover, far from unlocking the benefits of data, we are concerned that the proposals in the Report would instead create a highly burdensome, over-regulated framework that would have negative impacts – disincentivizing data collection, processing, and sharing – in direct contrast to its stated intentions, while also undermining the security of intellectual property, trade secrets, domestic investments, as well as the privacy rights of Indian citizens.

On behalf of our membership, we provide the below views on the Report and remain committed to engage with the Committee for any follow-up questions. In the first part of this response we will address several general concerns followed by specific comments on the individual recommendations within the Report. Our key concerns with the Committee's Report are that: 1) the underlying premises of the report are incorrect or unsupported; 2) the report creates uncertainty and confusion around India's data policy ecosystem; 3) forced data sharing undermines intellectual property rights; 4) data sharing should not be mandatory; and 5) data localization serves as a barrier to trade and investment.

Global Headquarters
700 K Street NW, Suite 600
Washington, D.C. 20001, USA
+1 202-737-8888

Europe Office
Rue de la Loi 227
Brussels - 1040, Belgium
+32 (0)2-321-10-90

info@itic.org
www.itic.org
@iti_techtweets

General Comments

The Report outlines a new framework to enable – and, in many instances, require – the sharing of data held primarily by multinational companies with companies in India and the GOI. It seeks to achieve this aim by defining types of NPD (including a novel category called "community data"), establishing the purposes under which third parties can request access to NPD, and creating roles and responsibilities for different actors in the NPD ecosystem. The Report recommends creating a new Non-Personal Data Authority (NPDA) that would oversee and enforce the proposed requirements, which would include new company registration requirements and mechanisms through which companies or the GOI can request access to data.

THE UNDERLYING PREMISES OF THE REPORT ARE INCORRECT AND UNSUPPORTED

While the Report correctly points out that the volume of data, and the innovative ways in which companies derive value from it, has grown exponentially in recent history and has transformed the global economy, the Report relies on several sweeping, unsupported statements to make claims of a perceived market imbalance. Furthermore, the Report focuses only on a handful of foreign, consumer-facing companies. This analytical approach overlooks the fact that data collection and use are not the exclusive province of a select group of firms, but in fact are increasingly critical to every globally competitive company. Sectors ranging from manufacturing, to agriculture, to logistics leverage disparate data sets to drive innovations in new products, services, and processes. The legal framework proposed by the Report would impact every major player operating in the Indian economy – as well as their partners – in a manner that undermines the stated objectives of the Report.

Furthermore, the Report does not clearly articulate the problem that it is trying to solve with the proposed data sharing framework. While it does include some attenuated concerns (such as competition), the Report does not address why existing regulatory frameworks do not sufficiently alleviate the Committee's concerns. It notes a potential imbalance in the economy in terms of the volume of data that companies possess, but neither presents supporting evidence for this claim nor provides justification for why such an imbalance would necessitate a cross-cutting legal framework of the kind described in the Report. Additionally, the Report does not acknowledge the significant risks associated with forced data sharing, including potentially exposing systems to cyber-attack, exposing trade secrets, and violating intellectual property rights. All of these risks combined, and their impact on investments in India, are likely to be so severe as to outweigh any perceived competition benefits from the proposed new regulatory regime.

The Report maintains that certain businesses have gained a "first mover advantage," which, combined with the network effects that have caused imbalances in the economy, creates significant entry barriers for startups and new entrants.[1] This is used as an argument to justify the time being ripe for a regulation on the NPD ecosystem. However, the assumption that network effects automatically lead to monopolies is flawed. Digital markets are dynamic, and consumers can switch to different platforms and service providers with ease. Any ex ante regulation to curb network effects without an assessment of harms to consumers is ill-founded. In any case, competition law and economics are the right tools to conduct this analysis and propose action, if needed.

[1] Page 7, para 3.6, the Report.

Further, we strongly believe that the premise of the report, and its focus on NPD, is flawed and demonstrates a misunderstanding of the practical ways in which data is collected, processed, and shared. The Report does not appropriately consider the significant variances in the types of NPD or the fact that personal data and NPD are often intermingled in data sets and not easily disaggregated, which would almost assuredly lead to over-regulation, mis-matched requirements on the type of data that is sought to be regulated, and likely insurmountable compliance challenges. The non-rivalrous nature of raw data means that, though it does take resources to collect data, the collection of raw data by one entity does not prohibit another entity from collecting the same data. It follows that the value of "using" data in one context does not detract from the value of using the same data in a second or even fiftieth application. Data is also non-fungible, meaning that different data are not necessarily interchangeable, and not all data are relevant for the same applications or outcomes.[2] What this means in practice is that, when considering data governance regimes, governments must recognize that there are many different types of data and ensure that any policy approaches seeking to control or leverage data for specific outcomes must first clearly define what type of data they are addressing. Broad-brush approaches that intend to regulate all data at once will inevitably be overly restrictive, inefficient, and potentially detrimental to innovative sectors of the economy.

For example, data for public use, such as that which pertains to roads, could be broadly valuable and easily sharable with travelers, auto companies, and infrastructure planners and builders. However, other data that might be used for marketing, advertising, or to provide services to consumers in media or technology businesses – such as e-commerce data – are more suited to a relationship between the consumer and the organization, and not a good candidate for being part of a large data set. Context and use of data are salient concerns and since they can vary widely, particularly for NPD, any regulatory framework must be targeted and specific to the type of data it is addressing. The incredibly broad categories of NPD in the Report do not take into account these realities of data and almost certainly would result in the over-regulation and depression of data-driven innovation in India.

Lastly, prescriptive or overly burdensome regulatory and/or registration obligations risk throttling innovation as well as the start-up ecosystem in India, in part by discouraging smaller companies from growing beyond the registration threshold as it would require them to comply with the registration requirements and to open their data sets to third parties. A poorly designed domestic regulatory framework can discourage foreign investment in India – much of which drives innovation and enables the growth of the Indian economy and jobs – while encouraging Indian companies to move their operations to more permissive regulatory jurisdictions abroad.

FORCED DATA SHARING UNDERMINES INTELLECTUAL PROPERTY RIGHTS

Among the most consequential concerns raised by the framework proposed in the Report is the implications of a compulsory data sharing regime on the intellectual property rights (IPR) of private entities, including the potential disclosure of trade secrets which could include the processes that companies use to collect data. Though the Report references "proprietary" data as being out of scope for data sharing, it indicates that this would only apply to highly processed data or algorithms. However, proprietary data sets include far more than highly processed data. All data collection requires significant

[2] "The Rise of Data Capital," *MIT Technology Review Custom + Oracle*, April 2016, http://files.technologyreview.com/whitepapers/MIT_Oracle+Report-The_Rise_of_Data_Capital.pdf?_ga=2.107506433.2062962261.1591387636-1723515635.1591387636.

investments in time and capital resources, and rendering even the most basic understanding of what data has been collected requires substantial human and technical resources and systems to process, analyze, and categorize raw data. The Report's definition of "private NPD" appears to acknowledge this centrality of "private effort" with respect to such data collection and use. For this reason, IPR laws around the world, including in India, recognize that copyright protections apply to databases. This means that even the most basic data sets held by private entities are proprietary. Accordingly, data should generally be shared, if at all, under terms and conditions (including appropriate compensation) set out in contractual arrangements with those private entities.

Furthermore, the requirement to make metadata freely accessible – a basic requirement repeated several times throughout the Report – raises serious concerns over the potential exposure of company trade secrets. By revealing what data, including metadata, companies collect, how it is organized, and when it is collected, among other details, competitors could extrapolate competitive information on the operations and strategies of the subject company. Asking businesses to put together information about the data they collect in directories and make this available on an open access basis is a significant investment. The Committee's proposed framework would require businesses to spend additional resources to provide details about these datasets at no cost to third parties, which is a significant and intrusive compliance obligation in itself. This could have a severe impact on the willingness of companies to do business in India and would serve to disincentivize data collection beyond the most basic requirements.

The Report does not include a discussion of how these concerns will be addressed, or acknowledge that there would be concerns of this nature. Without robust analysis and discussion of these concerns, the Report seems to undermine the stated objectives of the government to improve conditions for the ease of doing business in India, including through increased protections for IPR. The GOI has been developing improved protections for IPR, a welcome effort to increase investment from abroad. However, we are concerned that the mere publishing of this report in its current form by a committee appointed by the government may have already caused businesses to question current or future investments in India for fear of expropriation of their intellectual assets. We therefore strongly recommend that thorough consideration be given to IPR concerns before any policy action is considered with respect to NPD.

THE REPORT CREATES UNCERTAINTY AND CONFUSION AROUND INDIA'S DATA POLICY ECOSYSTEM

The Expert Committee was constituted and subsequently released the Report during a period in which multiple legislative and regulatory workstreams that will impact the treatment of both personal and non-personal data are under development. First among these initiatives is the Personal Data Protection Bill (PDPB), which has been pending in Parliament for over six months. There are many areas of the Report that would clearly impact personal data and directly conflict with proposed elements of the PDPB, including the personal/non-personal data metadata directories, an overlap in roles and responsibilities of the "data fiduciary" in the PDPB and the "data business" in the Report, the proposed pilot program on healthcare data sharing, and the proposed NPDA. In addition, the proposed consent requirement for anonymized data in this Report fails to consider that it may effectively discourage the anonymous storage of data, which would undermine individual privacy and conflict with the stated goals of the PDPB. Protecting personal data should always be a top priority and as such these and other proposed actions that conflict with the PDPB and would serve to undermine consumer privacy are counterproductive and unjustifiable.

In addition to the PDPB, there are other data policy workstreams that conflict with the Report. For example, the draft National E-Commerce Policy, first released early 2019 by the Department of Promotion of Industry and Internal Trade (DPIIT), seeks to create a new regulatory structure around e-commerce data, a specific type of NPD cited in the Report. There is no discussion in the Report regarding how the GOI would ensure that any resulting policy actions are not in conflict with this forthcoming regulation or others like it that could apply additional rights and responsibilities to actors in the data ecosystem that handle specific types of NPD. The release of this Report has already added to the confusion around who in the GOI is responsible for data-related policies, and risks amplifying policy uncertainty on how data will be regulated in India in the future, effectively chilling investment in India's digital ecosystem.

The Report also appears to undermine existing mechanisms for the sharing of public data, an area where the GOI has been a pioneer. We commend the GOI for programs like the National Data Sharing and Accessibility Policy (NDSAP) and the Open Government Data Platform (data.gov.in) which have sought to provide open access to government data, initiatives that we strongly support and encourage the GOI to develop further. These initiatives and others like them should be expanded to provide access to public data for innovators and entrepreneurs. It is unclear what value the framework in the Report would add given the success of existing public data sharing programs that reflect an open, transparent approach to public data sharing.

For these reasons, we encourage the GOI to focus on providing coherence to its data policy by establishing a consistent governmental lead for data policy and completing its existing workstreams in the most open, non-discriminatory, and innovation-facilitative manner possible.

DATA SHARING SHOULD NOT BE MANDATORY

Data sharing regimes – particularly those that are as broad in scope as the one presented in the Report – should never be mandatory. Forcing the sharing of data with governments and competitors would devalue data sets used by businesses to improve and provide their services, disincentivizing data innovation, and creating a hostile investment environment. More broadly, it could also disincentivize both domestic and foreign organizations from operating in the country imposing such sharing obligations. As such, the Expert Committee should eschew the model of forced data sharing, considering the far-reaching and unintended negative consequences, including but not limited to competition, employment, innovation, and economic growth.

It is the experience of our members that business to business (B2B) data sharing is most effective when conducted through private party contracts – without a government intermediary or specialized authority – which are entered into for the mutual commercial benefit of both businesses. These agreements are limited in scope and carefully tailored to ensure that both companies are able to maintain the security and privacy of their data while upholding their legal obligations to suppliers, consumers, and other third parties. Creating additional sweeping legal obligations as this Report suggests would discourage such beneficial commercial relationships and would assuredly be a mismatch for the contractual complexities and special circumstances of existing commercial relationships. There is no need or demand for a new governmental authority to enforce or moderate B2B data sharing, a role that existing contractual law sufficiently fills.

DATA LOCALIZATION SERVES AS A BARRIER TO TRADE AND INVESTMENT

The ability to move data and access information across borders is essential for businesses regardless of size or sector. Data localization measures serve as barriers to trade and offer governments a false choice between achieving regulatory objectives such as data privacy and security, and data movement. The Report proposes wide-ranging data localization requirements with no evidence for how these would further safeguard or protect the data of Indian citizens. The Report also fails to mention concrete policy objectives for mandating local storage of Sensitive NPD, and instead vastly broadens the scope of localization far beyond what has been proposed under the current draft of the PDPB, in which the storage mandate was narrowed in response to industry feedback. The end result will be a regime that inhibits value generation, reduces exports and foreign direct investment, and results in productivity losses for local companies that rely on a wide range of digital services. At the macroeconomic level, one prominent study assessing the impact of data localization legislation in seven economies estimated significant negative GDP impacts in each instance.[2] At the organization level, such restrictions can have a deleterious impact on the cost and availability of key digital services. Specific analysis undertaken with respect to cloud services found that data-localization policies restrict access to the most cost-competitive global cloud providers, and significantly raise costs for local companies purchasing cloud-computing services.[3]

In the modern global economy, data is already an essential means of widening consumer choice and the affordability of goods and services, helping small and medium-sized enterprises (SMEs) reach global markets, and fostering international production through global value chains, and its uses are widening. [4] Any regulatory measures restricting data flows will therefore also have detrimental consequences for trade and economic development. These consequences are likely to be particularly acute for domestic SMEs, as data-restrictive policies affect access to a range of cost-efficient digital technologies, including cloud and digital communication services on which SMEs rely. The impact of data localization will therefore yield the opposite outcome of what the Report is trying to achieve.

Subjecting such a wide range of datasets to local storage requirements without clear objectives, as the Report suggests, could lead to a host of unintended adverse consequences for innovation, privacy and the digital economy. Regrettably, the Report imports categories from the PDPB and recommends that "critical" and "sensitive" non-personal data be subject to similar localization requirement. In other words, the Report's approach would see a dramatic tightening of India's international data transfer regime, thereby cutting it off from the global economy, and undermining India's full potential as an innovation center for artificial intelligence (AI) and machine learning, deployment of 5G, and other advanced and emerging technologies.

---

[2] Bauer, Matthias, Erik van der Marel and Hosuk Lee-Makiyama, "The Costs of Data Localisation: A Friendly Fire on Economic Recovery," ECIPE, May 2014, https://ecipe.org/publications/dataloc/

[3] "Quantifying the Cost of Forced Localization," *Leviathan Security Group,* February 2015, https://static1.squarespace.com/static/556340ece4b0869396f21099/t/559dad76e4b0899d97726a8b/1436396918881/Quantifying+the+Cost+of+Forced+Localization.pdf

[4] Casalini, F. and J. López González (2019-01-23), "Trade and Cross-Border Data Flows", *OECD Trade Policy Papers*, No. 220, OECD Publishing, Paris, http://dx.doi.org/10.1787/b2023a47-en

Specific Comments on the Report's Recommendations

The Report presents its proposed framework through seven recommendations. In the sections below, we address our specific concerns with elements in each recommendation.

Recommendation 1: Defining Non-Personal Data

Recommendation 1 introduces several categories of NPD, including public NPD, community NPD, and private NPD. It also introduces the concept of "sensitivity" of NPD. At the outset, ITI would like to highlight that if the proposed categories were to be implemented, it would result in an unenforceable framework. The framework fails to acknowledge the complexity presented by mixed datasets, i.e., types of data that are inextricably linked and cannot be separated in order to adhere to these definitions, a complexity that is widely acknowledged in jurisdictions such as the EU. The Report's treatment of datasets represents a fundamental misunderstanding of how corporations collect, store, and process data.

COMMUNITY NON-PERSONAL DATA

We have concerns with the inclusion of community data as a category, particularly because the definition of "community" appears to be exceedingly broad. It is not clear how communities will be delineated, nor is it clear how to measure whether and to what extent a community is benefitting from data. Further, the notion of collective harm against a community invites endless dispute over definitions, scope, and intensity of harm. Initial questions could include the following: What is a community? How many people need to be in it? Is it time-bound? Is it based on physical location, family, religion, shared interests, or other categories? Can overlapping communities have ownership interests in the same data? What is different about the harms to this group as opposed to harms to individuals? How do we measure the harms and determine which are meaningful and which are not?

Given that the definition of a community – and the individuals within the community – will be flawed and in flux in every instance, the risk of "community data" being used to make decisions about individual data subjects in that "community" is high. With such broad definitions it is inevitable that such unconstrained sharing will result in detrimental outcomes for the individuals in a community, however defined. Ultimately, data creators need to maintain full control over when and how data is disclosed or shared with third parties. They must not be forced into sharing broad swathes of undefined datasets, especially those about communities of data subjects.

Given the lack of a clear definitions and the array of, in some cases, unanswerable questions, we would caution against introducing community data as a new category of NPD as recommended in this Report.

PRIVATE NON-PERSONAL DATA AND EXTRATERRITORIALITY

Like community data, "private NPD" is also a broad and meaningless category. Other than the fact that it signifies the nature of the data creator, it fails to create a clear sense of the type or categories of data it encompasses. Echoing our argument on community NPD, an overbroad definition carries significant risks. Specific to private NPD, we are concerned about the potential for extraterritorial application of the recommendations in the Report. For example, as a part of the definition provided for private NPD, the Report states that the NPD of Indian citizens, even when foreign bodies collect such data, are subject to the law sought to be introduced by the Report. Further, the Report recognizes that data collected could also include "global datasets" containing data of non-Indians and data which is collected in foreign jurisdictions. This extraterritorial application of the Report not only raises serious questions of overreach

under international law, but also invites other jurisdictions to interfere with India's sovereignty by similarly legislating on activities that take place entirely in India.

The framework proposed in the Report may interfere with the existing contractual agreements between the data principals and the companies, within and outside of India. In jurisdictions where a company and a data principal already have existing agreements subject to the law of the jurisdiction in which the company is doing business, extraterritorial application of this report will create intractable difficulties. It may also potentially conflict with other laws as the framework would be applicable to individuals who may otherwise be subject to protection under the laws of their own jurisdictions, such as the EU's General Data Protection Regulation (GDPR) and the U.S.'s HIPAA Privacy Rule, as well as other regimes, which have an impact on NPD.

CONSENT FOR ANONYMIZATION AND USE OF ANONYMIZED DATA

As referenced in our general comments, the Report does not take into account the complexities of anonymization. In the context of Recommendation 1, the Report recommends regulating anonymized data and creating a mechanism by which the data principal can provide consent for anonymization and usage of NPD. In practice, it is oftentimes impossible to clearly carve out personal data and NPD in a single dataset and privacy risks are exacerbated when the anonymized data leaves the original collector. Including anonymized data in this data sharing framework thus only serves to increase the risks of de-anonymization that the Report later inadequately attempts to address through additional regulatory proposals. Ultimately, we believe that data creators need to maintain full control over when and how data is disclosed or shared with third parties. Mandating sharing rather than the protection of data would only result in negative outcomes for privacy protection of data principals.

Additionally, mandating consent for anonymization may dissuade businesses from holding data in an anonymized form, which could make it more difficult to adequately protect such data. Businesses may already be under obligations to anonymize data for privacy best practices, so mandating consent poses practical issues which may make anonymization difficult. Storage in an aggregated, anonymized or less-precise form is better from a privacy perspective; however, if doing so is subject to a requirement for consent, it could have unintended effects, e.g., where businesses find themselves unable to avoid breaching the overall consent framework and comply with an individual's consent withdrawal request at a later stage because the data relating to that individual can no longer be identified. Moreover, such risks of re-identification and harm vary widely across different uses. For example, data that may impact public policy choices would present fewer risks to any individual. If the NPD poses no risk to any individual(s), it is unclear why such consent would be necessary – the data should be safe to share, and mandating "consent" as a required process would only serve to hamper data-sharing and innovation. In general, the consent requirements introduced for NPD in the Report would be just as onerous, if not more, than those for personal data. This might put at risk the privacy of users in a way not intended by the Report and also in a way that conflicts with the principles within India's pending privacy legislation.

ITI recommends that any core concept or terminology used to frame data policies be adopted from existing global standards such as ISO/IEC. Since the notion and definition of non-personal data is not yet developed internationally, any regulatory proposals based on non-personal data is premature and will hinder global interoperability and cross-border trade.

Recommendation 2: Define Non-Personal Data Roles

The Report defines four roles in the context of NPD: data principal, data custodian, data trustee, and data trust. Ostensibly, these definitions are intended to lend clarity to different recommendations in the Report, but, unfortunately, the definitions as drafted are artificial constructs and do not reflect the on-the-ground realities of liabilities and accountability between data creators, data principals, and recipients of the data. It would be manifestly unjust, if not impossible, for data custodians to take on the responsibility of: (a) ensuring that the data – which was collected for an *initial purpose* – is repurposed in a manner that is consistent and fair to both the creator of the data and the principal; (b) accounting for all the unforeseen and unpredictable downstream implications of data being repurposed and reconstituted; (c) accounting for issues like inaccuracy of the data and bias; and (d) accounting for potential losses to the business, especially if the data is proprietary or sensitive, resulting from misuse by the data recipient or a security accident.

Secondly, how can a data trustee be reasonably expected to represent the community if the data is non-identifiable and could be misused down the line? How data trustees are chosen, who funds them, and how their communities are delineated needs to be further considered. Further, it remains unclear how a specific community would "exercise its rights," as a community is not an individual or legal entity. The concept of "data trust" is also new and is at an experimental or theoretical stage in most jurisdictions. As such, it would be premature to start introducing these concepts into the framework of a proposed law at this stage. Finally, the co-ordination between these entities, and how they will interact in the best interest of the data principal, is unclear. It also appears that a data principal need not be a natural person; however, the utility of creating rights for non-natural persons under this framework is ambiguous, both in principle and in practice.

The Report also introduces the concepts of "beneficial ownership/interest" in community NPD, where the "community" would be the beneficial owner of the data, and would exercise its primary economic rights through a "data trustee." Besides the concerns raised above on how a "community" is to be delineated and a "data trustee" is to be chosen, among others, the concept of "beneficial ownership/interest" also introduces legal/equitable concepts of property ownership. Accompanying restrictions on dealing with the property would create tremendous uncertainty in how companies can deal with the data they collect. The related and additional imposition of a "duty of care" on data custodians also creates potential legal liability under common law, and even greater uncertainty and legal risk for companies in dealing with data that they collect.

It is evident that such artificial constructs will not enable safe sharing, but rather create a false sense of security that data can somehow be used "reasonably" and "in a controlled manner." We reiterate the need for data creators to maintain full control over when and how data is disclosed or shared with third parties.

Recommendation 3: Articulating a Legal Basis for Establishing Rights over Non-Personal Data

Given our broad objections with the premise of the Report, we do not have comments to provide on Recommendation 3. We do not see a benefit to examining the legal basis for establishing rights over NPD when, in our view, the creation of the categories of NPD itself is unfounded.

Recommendation 4: Defining a Data Business

The Report recommends the creation of a new category of "data business," which collects, processes, stores, or otherwise manages data, and meets a certain threshold criteria of data processing. It has been envisaged that data businesses will provide open access to metadata and regulated access to the underlying data. As a general matter, all businesses collect, store, process, and/or manage data to some extent, and could thus be described as a data business, barring the development of threshold criteria. Should threshold criteria be developed, it is unclear how it would be determined, evaluated, or enforced.

Notwithstanding our broader position that there should be no mandatory frameworks for "data businesses," we want to highlight through a number of specific examples how impracticable the proposed framework would be, including:

MANDATORY REGISTRATION

Data businesses are subject to mandatory registration in the Report after they meet the threshold of data processing determined under the framework. However, as mentioned above, the Report does not provide guidance on how the threshold criteria for data businesses will be determined. Moreover, there are compliance requirements set forth, including disclosure obligations relating to the nature of data business, kinds of data processing, and the designation of a "data officer," etc. Depending on how the threshold criteria is determined, it is plausible that the mandatory registration would create an onerous and unnecessary compliance requirement that could apply to every company operating in the Indian economy and, in practice, could become a market entry barrier for new players.

OPEN ACCESS TO METADATA DIRECTORIES

The Report recommends that data businesses be required to provide open access to their metadata directories, where any private party may access the metadata and request further access to more detailed datasets. However, if the determination and organization of metadata is conducted at the entity-level, the access and use of this metadata would be unlikely to result in any benefit to any other entity. Again, different data are not necessarily interchangeable, and the "demand" side of data intuits that what may be useful and usable in one context may not bear true for another context. Without proper and robust security infrastructure and cybersecurity resources in place, the party accessing the data may also be subject to security violations, such as confidential data leaks or attempted re-identification of anonymized or aggregated data. Further, it is concerning the Report contemplates that there should be a harmonization of personal and non-personal data-related directories, which would also subject metadata relating to personal data, and possibly even the personal data itself, to such risks under the proposed open access requirements. Requiring this type of access therefore disincentivizes

companies from scaling up data operations in India as they would become more vulnerable to legal risks and regulatory non-compliance.

Recommendation 5: Defining Data-Sharing Purpose

In Recommendation 5, the Report lays out several purposes for requiring data-sharing, including "sovereign purpose," "core public interest purpose," and "economic purpose." We recommend that the Commission reconsider these purposes, as they cast such a wide net that almost any rationale could be used to justify mandating data-sharing, and many justifications could generate conflicts with the Report's stated objectives. This Recommendation also discusses ways in which the Expert Committee understands value to be attached to data, another potentially problematic area. Below are specific items for consideration.

RECONSIDER DATA-SHARING PURPOSES & MANDATORY NATURE OF DATA-SHARING

The purposes for data-sharing laid out in the Report capture an incredible breadth of activities. Between the three purposes, it is difficult to imagine a situation in which data could not be requested. This is concerning, particularly because "mandatory" data-sharing is discussed in many areas of the Report.

Also troubling is that if a request for data is denied, the NPDA would be empowered to resolve the issue by ascertaining the "public interest benefit" of such data sharing. However, it is not clear whether such a body would have the expertise to adequately assess the "public interest," especially as such a body would also have to consider how the public interest benefit would be balanced with the value addition and resource investment by a private company to the dataset in question. There is also the question of whether such a public body can render truly impartial and objective decisions in weighing the "public interest" against the interest of the private company. Such decisions could potentially risk the IPR of companies as well as rights under bilateral investment treaties and free trade agreements. In some cases, mandatory data sharing may violate the terms of a company's agreements with IPR licensors or its confidentiality agreements.

Further, the carte blanche requirement for the private sector to provide any data on request to the GOI may drive legal challenges which will in turn hamper business, economic growth, and erode confidence in the Indian market. Any business to government (B2G) data sharing regime must be subject to the rule of law with clear standards for data requests, a predictable and transparent process that respects the rights of data subject and the entities handling the data, and a robust and responsive redress mechanism. The "sovereign purpose" must be clearly defined both in purpose and in scope before it is used as a justification for data requests. The principle of equality under the law must be central to any such recommendations.

RECONSIDER HEALTHCARE AS A PILOT AREA

The Report recommends healthcare as an area in which the NPD Governance Framework could be piloted. First, it is premature to consider healthcare or any other area as a pilot area until the fundamental issues outlined above regarding the flawed underlying premises of the Report are addressed. Second, with regard to healthcare specifically, we understand there is great potential for data sharing in the health sector to fight disease, support clinical trials, and produce laboratory results. However, we caution against choosing this sector as a pilot area, particularly because health-related data presents a unique set of challenges and sensitivities – not the least of which is the fact that

oftentimes much health data is not only personal data but sensitive personal data. It is hard to overcome privacy concerns related to such data – or to easily disaggregate sensitive personal data from NPD in such data sets – and rate the degree of sensitivity. Sharing anonymized health data can never be risk-free, especially in the context of data security and the risks of leaks and data breaches, as well as ethical considerations. In recent years, there have been several cases in which cybercriminals have initiated re-identification attacks on healthcare databases in hospitals and research institutes. Since there is no anonymization method currently available to ensure that the re-identification risk is zero, and considering the heightened sensitivity of health data, any initiatives to promote health data sharing must consider the unique consequences.

DETERMINING THE "VALUE" OF DATA

The Report creates a scheme for enabling startups, businesses, data trustees, and governments to request access to data from data businesses. However, we have significant concerns with the way in which the Report proposes "valuing" data. For example, mandatory external valuation standards will make it difficult for companies to value their data correctly in the context of any sharing or transaction because it removes flexibility for engaging with the data. This again contradicts the fact that all data are not interchangeable, and therefore different data have different values to different users. Additionally, the Report proposes that no remuneration be paid for raw/factual data, but it does not account for the resourcing and monetary costs of collecting, storing, protecting and analyzing this type of data. This is problematic because the data was collected with a particular intent, and benefits should accrue to the data-collector. Without such benefits, there would not be any incentive to participate in data collection. We also note that the fair, reasonable, and non-discriminatory (FRAND) concept emerged in the highly specialized context of standard-essential patents and is itself not free of controversy. Without proper established jurisprudence for the use of the FRAND concept in the context of NPD, it will not serve any intended purpose and is not practical.

CONSIDER AND ADDRESS TECHNICAL DATA-SHARING CHALLENGES BEFORE RECOMMENDING DATA-SHARING SCHEMES

There are various technical challenges that must be considered before creating a data sharing framework, such as the difficulty in ascertaining the viability of a company to create a dataset that is readily available for data sharing. Since data-sharing in isolation is extremely difficult, the entities which are required to share data will inevitably be forced to expose details revealing components of their competitive advantage, trade secrets, business models, and corporate strategies.

Furthermore, we encourage India to facilitate the removal of technical barriers for data sharing, beginning with a robust G2B data sharing environment. As first step, India should open up large sets of government NPD for public use as it is widely accepted that governments hold the largest amount of NPD that is not currently being utilized. Even if governments make data available, it does not mean the data is easy to find and ready for use. Other barriers to G2B data sharing include lack of business awareness, lack of technical knowledge, incompatibility of unlike data sets, and poor data quality. Those factors often make it difficult to share and aggregate data in a way that is valuable. We encourage the GOI to make high-quality public sector data available to the public for re-use, and respectfully recommend that it first narrow the focus of this consultation on a G2B data-sharing strategy instead of advancing premature discussions on NPD.

Recommendation 6: Defining Data-Sharing Mechanisms and Checks and Balances

This Recommendation attempts to establish the terms under which data is to be shared, the process for doing so, the role of the new NPDA, and additional mechanisms to govern the sharing of data. Many of the concerns raised in the sections above are also relevant to this recommendation, but there are also several new concerns that arise.

VALUE ADDITION

This recommendation details the data that would be subject to the data sharing framework, including raw/factual data being shared for free with various levels of additional restrictions as the perceived value of the data increases. However, there does not seem to be any attempt at defining what a "value-add" event means, or who would determine when data has undergone sufficient value-add to be considered worthy of increased remuneration or protection from sharing requirements. While we welcome the recognition that data may undergo several layers of value addition that in turn imbue it with more value and sensitivity, even the collection of raw data is very expensive. Requiring that it be shared for free severely diminishes the commercial incentive to collect such data in the first place. These requirements would create a severe free-rider problem in the Indian data ecosystem and ultimately hamper – rather than encourage – data-driven innovation.

SENSITIVE NPD AND DATA LOCALIZATION

The Report emphasizes the need for "data sovereignty," and proposes that Indian laws and regulations apply to all data collected in/from India or by Indian entities. It recommends that a certain category of data characterized as "sensitive NPD" be stored within India, but transferrable outside India, while "critical NPD" can only be stored and processed only within India, and general NPD can be stored and processed anywhere in the world.

The Report's definition of "sensitive" NPD is broad and far-reaching in nature, particularly as it includes national security or "strategic interests." There is no indication as to what may be considered a "strategic interest," making it unclear what categories of data would be subject to such localization requirements. That said, it is our view that even if "strategic interests" was to be defined in more depth, that data still should not be subject to localization. The Report also leaves "critical NPD" undefined, which remains problematic for practical implementation. Moreover, these arbitrary definitions would be practically unworkable and unenforceable. They do not account for the complexity of mixed datasets which cannot be separated.

Recognizing that this requirement is closely linked to the PDPB currently pending in Parliament, we would refer the Committee to our comments[5] on that legislation for a fulsome discussion of our concerns with forced data localization. Given that the primary personal data regulation on these concepts is still not clear nor in effect, seeking to simultaneously regulate NPD using a similar approach will only create greater regulatory and business uncertainty. Subjecting such a wide range of datasets to local storage requirements without clear objectives could lead to a host of unintended adverse consequences for innovation, privacy, and the digital economy. We recommend against mandating data localization, especially because the costs of adhering to such localization measures may be insurmountable for new domestic and international players interested in entering the market. This would in turn significantly affect competition and economic growth in India.

[5] https://www.itic.org/policy/ITIResponsetoPDPBill2019.pdf

Recommendation 7: Defining a Non-Personal Data Authority

Given our objections with the premise on which the Report is based, we do not have comments to provide on defining a NPDA as we do not see a need for such an authority to be created.