Response dated 13 September 2020 to the Report by the Committee of Experts on Non-Personal Data Governance Framework released by the Ministry of Electronics and Information Technology in July 2020

Dvara Research[1] is an independent Indian not-for-profit research institution guided by our mission of ensuring that every individual and every enterprise has complete access to financial services. Our work seeks to address challenges for policy and regulation in India given the waves of digital innovation sweeping financial services, focussing on the impact on lower income individuals in the country. The regulation and protection of data has been a core area of our recent research.

In this document we present our response to the Report by the Committee of Experts on Non-Personal Data Governance Framework (the Report) released by the Ministry of Electronics and Information Technology (MeitY) in July 2020.

This response identifies some overarching concerns with the Report that we seek to convey and substantiate. They are categorised under the following three sections:

Section 1: The Report fails to identify the basis for a separate regulatory regime to regulate non-personal data (NPD).

Section 2: Lack of clarity in concepts and definitions that are foundational to the Report's vision further weaken the basis for it to exist.

Section 3: Assumptions presented in the Report that are not substantiated with suitable evidence.

We are concerned that a sweeping regime with detailed regulatory supporting apparatus has been set envisioned in this Report without a clear articulation of objectives or motivations of such a regime, or why these objectives cannot be dealt with under existing frameworks. Further, it is unclear that the market-wide or consumer-level risks from such a framework have been thoroughly and rigorously considered, or fleshed out using well recognised policy analysis framework.

In the interests of evidence-based policy-making, we hope and humbly submit that any future policy and regulatory decisions relating to NPD be undertaken only after more detailed research and wider public consultation with a range of experts from different disciplines on the costs, benefits and impacts of such proposals.

[1] Dvara Research has made several contributions to the Indian financial system and participated in engagements with many key regulators and the Government of India. Through our recent work we have extended research inputs to bodies including the Committee of Experts on Data Protection under the Chairmanship of Justice B.N. Srikrishna, the Ministry of Electronics & Information Technology (MeitY) and the RBI's Committee on Deepening of Digital Payments. Our primary research on Indians' data sharing attitudes was cited in the 2017 White Paper of the Expert Committee on Data Protection under the Chairmanship of Justice B.N. Srikrishna. Our regulatory proposals on enforcement and the design of the Data Protection Authority (DPA) were specifically acknowledged and relied upon in the Final Report of the Committee dated 27 July 2018.

CONTENTS

Section 1: The Report fails to identify the basis for a separate regulatory regime to regulate non-personal data (NPD)

1.1. Lack of clarity in the needs and objectives driving the creation of a separate regulatory framework call into question its requirement

The Report makes an underlying case for regulating NPD based on three concerns, namely: competition, privacy and public policy (p. 11). In Section 3.8, the Report lays out that "*rules and regulations are required to manage data*" in order to:

(i) create certainty and incentives for innovation and new products/services creation,
(ii) create a framework for the creation of economic/social/public value from data, and
(iii) address privacy concerns.

We find that these objectives are insufficient to form the basis for a separate regulatory framework for the regulation of NPD.

1.1.1 Concerns over Competition

The Report states that there is an imbalance in data and the digital industry in India. It also mentions that a few companies are capable of dominating the digital and data business due to a combination of factors such as first-mover advantage, sizeable network effects and vast quantities of data collected over the years (pp. 7-8). The concerns raised appear to highlight that in a largely unregulated environment, creation of data monopolies could lead to the formation of power imbalance, which could further result in companies having access to large data sets on one side and the rest (Indian citizens, Indian businesses including start-ups, MSMEs and even the Government) having little to no access at all. Accordingly, the Report calls for a new regime and tasks the Non-Personal Data Authority (NPDA) "*with addressing market failures and supervising the market for Non-Personal Data*".

The primary concern highlighted in the Report is that certain characteristics of data markets could inhibit fair competition. It then appears to indicate that the very existence of these characteristics can create significant entry barriers for new entrants and start-ups.

However, well-established economic and regulatory thinking on competition confirms that the exhibition of particular characteristics – such as the dominance of a few players – do not ***automatically*** indicate that the welfare of consumers or the competitiveness of the markets are being negatively affected. Theoretical arguments and empirical evidence from research provide mixed insights as to whether the data advantage of incumbents is insurmountable (World Economic Forum, 2019). In any event, any negative impacts on competition can potentially be dealt with under India's existing framework for competition regulation.

Accordingly, we note that (1) the existence of market power does not automatically indicate abuse of such dominance, (2) several tests exist to understand whether such abuse is taking place in India and (3) the competition and anti-trust issues emanating out of the use of NPD can be addressed by expanding the provisions of the existing Competition Act of India, 2002 (the Competition Act). As such, we find that the competition concerns outlined in the Report are not an adequate basis for the creation of a separate regulatory framework.

Market power does not indicate abuse of dominance

According to provisions in the Competition Act, an entity holds a ***dominant position (dominance, alternatively market power)*** in the relevant market, if:

(i) It is able to operate autonomously, irrespective of the competitive forces existing in the relevant market or

(ii) It can influence its competitors, its consumers, or the relevant market in its favour (Competition Commission of India, 2020).

Dominance, in itself, is not considered anti-competitive. The acquisition of market power by competitive means, such as by providing better products and services or by performing superior to competitors, is not classified as a competition problem. ***Abuse of dominance*** arises when an entity or a group of entities utilises its dominant position in a relevant market in a manner that is either exclusionary and/or exploitative (Competition Commission of India, 2020). Exclusionary acts include conduct that has the objective of terminating competition from existing or prospective competitors. Exploitative acts include utilising dominant position to extort advantages from customers or trading partners, for instance, by charging unfair prices (World Economic Forum, 2019).

The literature on competition policy indicates that for assessing abuse of a dominant position (e.g. in the EU) or monopolisation (e.g. in the US), authorities have to define a ***relevant market*** in which the firm under investigation competes. This helps to assess if the firm holds any market power and if it wields this power to undertake anti-competitive behaviour (World Economic Forum, 2019).

According to the Competition Act (2002), the ***relevant market*** refers to

> "*the market that may be determined by the Commission with reference to the relevant product market or the relevant geographic market or with reference to both the markets*".

The relevant product is defined in Section 2(t) of the Competition Act as

> "*the smallest set of products (both goods and services) which are substitutable among themselves, given a small but significant non-transitory increase in price*".

The relevant geographic market is defined in Section 2(s) as

> "*the area in which the conditions of competition for supply of goods or provision of services or demand of goods or services are distinctly homogenous and can be distinguished from the conditions prevailing in the neighbouring areas*".

Traditionally, market share held by an entity or group of entities was used to assess dominance. However, presently, additional factors are also considered to ascertain the power held by entities. These include market structure and size of the market, size and importance of competitors, size and resources of the enterprise, vertical integration, dependence of consumers on the entity etc. to name a few (Competition Commission of India, 2020).

On a similar note, whether ***access to data*** provides entities with a ***competitive advantage*** is a contentious issue. It depends on various factors, such as:

(i) Data substitutability: How essential is the incumbents[2]' data?

(ii) Data complementarity: On combining a diverse set of data, is there any effect on accuracy?

(iii) Data returns to scale: On increasing the quantity of data, is there diminishing returns to the accuracy of forecasts? (World Economic Forum, 2019)

The conclusions of available evidence, on the degree to which data provides an advantage to incumbents, is mixed. Therefore, further research is required on the types or features of data that may create entry barriers, as well as on the significance of data in establishing market power.

Given this context, the Report fails to outline the consideration of any of these factors to understand if there is an abuse of dominance in the market that would justify the need to regulate non-personal data.

Jurisdiction of the Competition Commission of India

The CCI is the sole quasi-judicial and regulatory body established under the Competition Act (Lakshmikumaran & Sandeepan, 2020). As a statutorily created expert body, the CCI has been undertaking this duty in accordance with the principles laid out in the Competition Act (Singh M. , 2020).

The Competition Act (2002) in Chapter IV, Section 18 lays down the duties of the CCI. It states that it is the duty of the Commission to

> "*eliminate practices having adverse effect on competition, promote and sustain competition, protect the interests of consumers and ensure freedom of trade carried on by other participants, in markets in India*".

The Act, through provisions laid out in Sections 19 and 20, has equipped the CCI with powers to make enquiries into anti-competitive agreements and abuse of dominance by entities that might have an adverse effect on competition in the country (The Competition Act, 2002). Based on these powers, the CCI has investigated anti-competitive behaviour across diverse sectors of the Indian economy. The authority is indicated to only '*force the hand of the market*' when required, as the generally held belief is that the market corrects itself.

Any assessment undertaken by the CCI includes an investigation by a Director General into the anti-competitive effects, an Appellate Tribunal that reviews decisions, detailed market reports, surveys, and economic analyses. All these measures are undertaken to ensure that an intervention in the market does not affect competition and consumer welfare, or harm innovation (Singh M. , 2020). Therefore, in the Indian context, competition related issues such as anti-competitive practices, abuse of dominance, market distortion and trade barriers created by entities are already being addressed by the CCI.

Further, the Competition Law Review Committee (CLRC) that was set up by the government in October 2018 also dedicated a section to look into competition issues that were arising from digital markets. The CLRC looked into matters such as online vertical restraints, the definition of price, algorithmic collusion, widening the ambit of Section 3 (anti-competitive agreements), control over data and

[2] An incumbent is an entity having a sizeable share of the market

assessment of market power, and new thresholds for notification of combinations. The CLRC found that:

- Section 19(4), which provided the CCI with a list of factors to determine the dominant position of an entity, was inclusive and broad enough to accommodate for situations where "*control over data may pave the way for dominance and its abuse*".
- Section 19(4)(b) explicitly mentions "*resources of the enterprise*" as a factor to determine market power, where the resources may include data.
- Section 19(4) was inclusive enough to also consider *network effects* in determining the dominant position of an enterprise (Ministry of Corporate Affairs, 2019).

As such, competition and anti-trust issues emanating out of the use of NPD appear closer to the jurisdiction of the CCI. Given this context, entrusting the function of addressing market failures and competition issues to the NPDA impinges on the jurisdiction of the CCI, which is a regulator that has both jurisdictional and technical expertise for the same.

1.1.2. Concerns over Privacy

According to Section 3.8 of the Report (p. 11) there is a case for regulating data in order to

> "*address privacy concerns, including from re-identification of anonymised personal data, preventing collective harms arising from processing of Non-Personal Data, and to examine the concept of collective privacy*".

Accordingly, the Report looks to address the harms arising from privacy violations due to re-identification of anonymised data or the derivation of personally identifiable insights from NPD. It highlights that this requires eventual regulation to mitigate against the risks of privacy harms, which does not dilute the protections offered by the PDP Bill. This appears to acknowledge that privacy concerns must be dealt with under India's new draft Personal Data Protection Bill, 2019 (the draft PDP Bill) which aims to protect the privacy of Indians.

Consequently, the expansion of the mandate of the NPDA to address collective privacy harms seems incongruous as this will encroach on the role and functions of the Data Protection Authority (DPA) under the draft PDP Bill which is deemed responsible for the protection of data and information privacy in India. The individual privacy concerns as well as group privacy concerns (as further set out in section 2.3 of this Response) raised in the Report can be addressed under the privacy framework conceptualised by the draft PDP Bill.

Re-identification risks emerge when NPD records are: (i) isolated in order to identify a data principle (directly or indirectly); (ii) connected to comparable records in other dataset(s) to determine the identity of a data principal; or (iii) used to make inferences about the identity (Article 29 Data Protection Working Party, 2014). In these instances, due to the identification of individuals from the data concerned (both inferred and otherwise), the data would be classified as personal data. This would bring it within the scope of the draft PDP Bill (Singh, Raghavan, Chugh, & Prasad, 2019).

The DPA is responsible for data protection and preserving informational privacy under the draft Bill. It has the power to make regulations and codes to mitigate re-identification risks and allied privacy concerns (Singh, Raghavan, Chugh, & Prasad, 2019). To support this, Section 82(i) of the PDP Bill considers re-identification and processing of de-identified personal data as an offence and lays down penalties to dissuade the same (MeitY, 2019).

Section 8.2(ii) (p. 41) in the Report highlights that the focus of the NPDA would be on "*unlocking the value in Non-Personal Data for India*" only, and not as a privacy regulator. The addition of addressing privacy concerns in the mandate of the NPDA only creates an internal inconsistency with the Report as well as future regulatory overlap and clash with the proposed DPA under the PDP Bill.

Section 2: Lack of clarity in concepts and definitions that are foundational to the Report's vision further weaken the basis for it to exist

2.1. The Report's definition of non-personal data has flaws and inconsistencies, and therefore does not clearly demarcate the subject matter of the proposed framework

The definition of NPD as stated in the Report (p. 13) is "*data that is not personal data (as defined under the PDP Bill), or data that is without any Personally Identifiable Information (PII)*".

The Report in Section 4.1(iii) (p. 13) also quotes the *Guidance on the Regulation on a Framework for the Free Flow of Non-Personal Data* in the European Union, 2019 released by the European Commission to give a general definition of Non-Personal Data (European Commission, 2019):

"*Firstly, data that never related to an **identified or identifiable** natural person, such as data on weather conditions, data from sensors installed on industrial machines, data from public infrastructures, and so on. Secondly, data which were initially personal data, but were later made **anonymous**. Data which are aggregated and to which certain data transformation techniques are applied, to the extent that individual specific events are no longer identifiable, can be qualified as anonymous data*".

Although these definitions of personal and non-personal data appear to be conceptually distinct, in practice they are not watertight and often flow into each other for the following reasons.

- ***Most datasets are "mixed datasets"***: It is well-recognised that the majority of the datasets used in the data economy are ***mixed datasets***. A mixed dataset contains both personal data as well as NPD (European Commission, 2019). For instance, the tax records of a company would also include personal information such as the name and telephone number of the managing director of the company. Other examples include data associated with the Internet of Things where some of the data allows presumptions to be made about identifiable individuals (e.g. usage patterns), and evaluation of log data originating from the operation of manufacturing equipment in the manufacturing industry etc. (European Commission, 2019). Further, personal data protection laws should be applicable to mixed datasets, as personal and non-personal data tend to be inextricably linked within such datasets (European Commission, 2019).

- ***"Identifiability" of data is dynamic and context-dependent and cannot be established in an absolute manner***: Information is deemed to be personally identifiable if it either directly identifies individuals or identifies individuals when it is considered in conjunction with other available information (Medical Research Council, 2019). However, the advent of Big Data, Artificial Intelligence, and IoT has vastly expanded what constitutes as personally identifiable data. Research has shown that NPD, in combination with other datasets, can reveal PII.

  In one study, anonymised data that had information about the movie preferences of Netflix users was used in conjunction with publicly available information from the Internet Movie Database (IMDb) to deanonymise the original data set (Narayanan & Shmatikov, 2008). Another study that used anonymised census data from 1990 found that 87% of the American Population could be uniquely identified on the basis of just three data variables – the 5-digit ZIP code, gender, and date of birth (Sweeney, 2000). As such, what is not personal data now might well be so in the future (Graef, Gellert, & Husovec, 2018).

Consequently, maintaining two separate legal frameworks - one for the regulation of personal data and another for non-personal data - when the two types of data cannot be clearly distinguished from each other could create considerable hurdles. The illusive notion of non-personal data as a starting point for new regulation could create major legal uncertainties and undermine the effectiveness of the objectives that it aims to pursue (Graef, Gellert, & Husovec, 2018).

2.2. The categorisation of data into Public NPD, Private NPD and Community NPD is logically inconsistent, leading to significant overlaps and uncertainty

The Report provides for three categories of non-personal data - **public**, **community**, and **private** NPD. This could prove to be a complex segregation to implement because the definitions of the three categories are not watertight and provide enough space for overlap and confusion. Our analysis suggests that data could fall into one or more of these categories at any point, which would create confusion regarding the relevant rules for their treatment.

Before analysing the differences and similarities among these three categories, it must be noted that a clear distinction between personal and non-personal data may not exist due to reasons considered in Section 2.1. This further complicates any attempts to create categories within NPD.

When specifying the three sub-categories of NPD, the Report mentions the different dimensions of data. Examples include the data's purpose, the sector from which data originates, the source of data, the level of processing involved, who collects the data "*or based on the extent of involvement of stakeholders in the creation of data*" (p. 13). Based on these dimensions, the Report proposes public, community, and private as the different categories of NPD.

- Public NPD, according to the Report, is NPD collected or generated by governments or its agencies (p. 14). The definition indicates that any data that is not personal data and is collected or generated by the government or its agencies is public NPD. Additionally, the Report states that public NPD will be treated as a national resource (p. 48).
- Private NPD is any NPD collected or produced by non-government persons or entities and "*the source or subject of which relates to assets and processes that are privately-owned*" (p. 15). In this case, only the "*raw/factual data pertaining to a community, that is collected by a private organisation may need to be shared*". Any NPD not pertaining to a community and collected by a private organisation would be private NPD (p. 28).
- Community NPD is any NPD or anonymised personal data "*about inanimate and animate things or phenomena – whether natural, social or artefactual, whose source or subject pertains to a community of natural persons*" (p. 15). Unlike public and private NPD, community NPD is defined by the **source** of the NPD instead of by the **collecting agency**. This presents a logical inconsistency in the process of creating categories, as the same dimension is not followed for creating all three categories. Further, the Report states that community NPD also includes datasets "collected by the municipal corporations and public electric utilities, datasets comprising user-information collected even by private players like telecom, e-commerce, ride-hailing companies etc." (p. 15).

These definitions seem to suggest that **there is a possibility for data to be identified under two of the NPD categories simultaneously**. For instance, NPD collected by a municipal corporation about a housing society could be categorised as public NPD. However, the current definitions allow for it to also be categorised as community NPD by a resident welfare association.

The categorisation is important because the control over data and the subsequent data sharing arrangements depend on these categories. The consequences of overlapping definitions will complicate the application of the rules under the framework, including the proposed data sharing arrangements. Separately, the current definitions of public and private NPD do not account for NPD generated or collected by public-private ventures or by private organisations funded by public enterprises.

This foundational definitional issue has repercussions for the entire proposed framework. It further calls into question the robustness of the proposals and the need for such a separate framework.

2.3. The understanding of groups, group privacy, and collective harm in the Report is narrow

The Report makes a case for regulating NPD "*to generate economic benefits for citizens and communities of India*", to address collective privacy concerns, and to "*prevent collective harm*". (p. 11).

To this end, the Report states (p. 10),

> "*non-personal data, including anonymised personal data, could provide collective insights that could open the way for collective harms (exploitative or discriminatory harms) against communities*".

The Report provides instances of probable group privacy harms, such as data emerging about a group of people with a specific sexual orientation frequenting certain places. These groups and places can be identified and targeted by those groups of people who oppose that sexual orientation. Similarly, people with certain diseases may be ostracised by their locality members. Therefore, according to the Report, there is a need to protect the privacy concerns of individuals at a group or collective level because of the ubiquitous nature of data.

However, data that identifies or relates to "group" identities of individuals (such as individuals' sexual orientation and health) is already protected under Clause 36 of the draft PDP Bill in the "*sensitive-personal data*" category (Government of India , 2019, p. 5). Accordingly, data identifying and relating to the sexual orientation or health of groups of individuals would also be protected under the privacy framework provided by the draft PDP Bill. In such cases, addressing collective privacy or privacy of groups of individuals would be governed by the privacy framework provided by the draft PDP Bill.

Separately, the Report fails to take account of some types of groupings that create risks for people purely on account of the operation of technology. Big data analytics, machine learning, and data mining enable data processors to collect, organise, and analyse vast amounts of data to draw critical insights about the behaviour and choices of individuals (Taylor, Floridi, & Sloot, 2017). Data is, now, gathered and analysed about "*large and undefined* groups" (Kammourieh, et al., 2017). They enable governments and businesses to identify patterns in groups, and profile or make decisions about them. This can create harms for people at the collective level because of their identification as part of a group, rather than purely as individuals.

The Report identifies communities or groups in a "*traditional*" sense, as it defines groups based on "*common interests and purposes, and involved in social and/or economic interactions*" (pp. 14-15). This indicates that the members of these groups share some pre-defined socially constructed, physical and/or behavioural characteristics (Kammourieh, et al., 2017). In addition, the members of such groups are self-aware and proclaim the existence of such groups (Mittelstadt, 2016).

Mittelstad (2016) identifies three types of groups based on membership,

i. Collectives: Individuals forming a group based on common background, interests, traits, and/or purposes. Eg: Labour unions and trade unions.

ii. Ascriptive groups: Individuals that come together due to inherent or incidentally developed characteristics. Eg: Ethnic and racial groups.

iii. Ad-hoc groups: Groups whose membership is assembled according to perceived, and sometimes new and imperceptible, links between members, often for a time or purpose-limited period for third-party interests. Eg: Market segments and profiling groups.

Collectives and ascriptive groups fit into the concepts stated in the Report i.e. their respective members willingly and consciously come together. Collectives and ascriptive groups have a collective identity ("*refers to reducible identity aggregated from individual members and non-reducible identity held by the group itself*") and self-awareness which are observed as moral reasons or minimum requirements for claiming rights (Mittelstadt, 2016). However, the Report does not seem to account for privacy risks or harms that arise through profiling within databases, where "ad-hoc groups" are created without the knowledge of the individuals by whom they are constituted themselves. This limits the robustness of the Report's conceptualisation of group privacy.

In any event, group and personal privacy concerns raised by ad-hoc groups can also be addressed under the privacy framework conceptualised by the draft PDP Bill. As argued in Section 2.1 and Section 2.2, having two regimes regulate mixed datasets to protect the privacy concerns of the same data principals/subjects may not be an effective regulatory approach.

2.4. The model of data trustees suggested by the Report is limited in its authority as a representative of the communities it is supposed to represent

The Report identifies multiple stakeholders in order to enable and develop a robust NPD ecosystem. They are the Data Principal, Data Custodian, Data Trustees and Data Trusts. In the context of community data and communities, the Report envisions that groups or communities will exercise their data rights through a suitable data trustee. Such a data trustee is responsible for representing the best interests of the community to the data regulator in the cases of mandated data sharing and recommending enforcement of soft obligations of data custodians.

This model of data trustees is problematic for two reasons. One, the ambiguity in how the data trustees interact with the data trust calls into question the actual authority that data trustees may have. Two, the suggestions made for bodies that will act as trustees indicate that they are not genuine representatives of the community (or even their more vulnerable members) but could merely entrench government control or other power structures.

2.4.1. The ambiguity in the relationship and interaction between data trusts and data trustees calls into question the authority that data trustees may have

In some models of data governance, data trustees have been envisioned as intermediaries that manage data for data principals and protect their privacy by assessing how data should be shared and processed (Kerber, 2020). These data trustees operate using data trusts - infrastructures that are designed with technical and legal expertise - and manage data sharing between different entities (Blankertz, 2020).

In this Report, data trusts are described as "*institutional structures, comprising specific rules and protocols for containing and sharing a given set of data*" (pp. 21-22). Such an infrastructure would contain data from multiple sources and would be responsible for managing and sharing data within it with different sectors for specific uses. However,

- instead of a data trustee or a group of different data trustees operating data trusts, the Report suggests that they can be operated by public authorities;
- data trustees appear to have limited rights or levers to impact the management of data trusts;
- data trustees only interact with trusts through the regulator and therefore has limited capacity in how well they can protect the privacy interests of the community they represent.

Separately, the entities indicated to represent communities on page 20 of the Report seem to be Government bodies, and in one case, a non-governmental organisation (NGO). It is not clear if such bodies will act in accordance with the interests of ALL community members, rather than entrench existing power structures or worse act in their own interests rather than that of the community. No democratic, representational or other safeguards are described in the Report, seriously calling into question the real representativeness of the data trustee structure.

2.5. Asking users for consent for anonymisation of their personal data does not provide suitable empowerment to them, as envisioned by this Report

In the Report, any data that is not personal data is considered as NPD. Any personally identifiable data that has been anonymised is also considered NPD. The Report recommends that data principals be asked for consent for the anonymisation and usage of such data while providing consent for the collection and usage of their personal data (p. 14). While consent is a good first step, the same problems of consent-taking for collection and processing of personal data persist in the case of NPD.

Several empirical research studies show that there are severe cognitive limitations that impair individuals' ability to make informed and rational choices about the costs and benefits of consenting to the collection, use, and disclosure of their personal data (Solove, 2012). Acquisti (2004) has highlighted how immediate gratification bias - due to which individuals overvalue the immediate period as compared to all future periods - can cause individuals to make suboptimal privacy decisions.

Due to advancements in data analytics, it is often difficult for providers to state at the time of obtaining consent the purposes for which anonymised data will be used. Separately, in a country like India, where several individuals are first time users of technology, anonymisation is a new and incomprehensible concept. Accordingly, asking data principals for consent for anonymising their personal data as well as for its usage is inconsequential and fails to provide data principals with any level of protection from the potential harms that may arise immediately or in the future due to the processing of their anonymised data.

Section 3: Assumptions presented in the Report that are not substantiated with suitable evidence

3.1 The assumption that data can be "priced" is not tested. The value of data is context-specific, and currently, it is widely recognised that data may not be accurately "priceable"

The Report in Section 7.4 (iii) (p. 37) suggests mechanisms for the sharing of data based on its perceived economic value. These include:

- sharing of raw or factual data regarding the community that was collected by a private entity, **at no remuneration**,
- mandated sharing of community data sets with significant value addition through **processing on fair, reasonable, and non-discriminatory (FRAND) based remuneration**, and
- Sharing of data sets with a higher value addition in a well-regulated data market, with **a price being determined by market forces**.

These mechanisms appear to assume that the value of a given data set can be easily defined and converted into monetary terms.

However, research indicates that there are several factors that can help determine the value of data but based on the context of its use. This makes it difficult to estimate the price of data based on its value alone.

Data are heterogeneous goods whose value depends on the context of their use, with different implications for individuals, businesses and policymakers (OECD, 2019). It has no intrinsic value, as its value is contextual and may be based on numerous attributes such as frequency of usage, content, cost of creation, security requirements and so on. The value of data may also alter over time as a result of changing priorities, litigation, and regulation, all of which are pertinent factors but are difficult to quantify (Short & Todd, 2017). Additionally, as a result of its context dependency, it is almost impossible to assess the value of data ex-ante (before its use) (OECD, 2013).

The '*Quality Framework and Guidelines for OECD Statistical Activities*' (OECD, 2011) advises that data quality and therefore value needs to be considered as a multi-faceted notion. It defines seven dimensions that affect the quality and value of data, namely: relevance, accuracy, credibility, timeliness, accessibility, interpretability, and coherence (OECD, 2011). Two of these factors particularly affects the value of data: (1) accuracy and (2) timeliness (OECD, 2015). The more relevant and accurate the data is when used in a particular context, the more its value and usefulness increases. Alternatively, this implies that the value of data can expire over time based on the use case (Engelsman, 2009).

In addition, the extraction of information from data is contingent on not just data but also the capability to link different datasets in order to derive insights. As such, there are factors beyond data itself that help determine its value. These are as follows:

i. Data linkage***:*** Insights that can be derived from data depend on the organisation and structure of the underlying data. Therefore, the same data set can result in providing different information based on its structure as well as its linkages to other (meta-)data (OECD, 2015). For instance, housing data can be linked to data on examination results in order to understand the impact that socio-economic factors have on education. This impact can also be studied by linking income data to exam results (Scottish Government, n.d.).

ii. Analytical capacities: All potential information and insights that can be derived from the data are extracted and interpreted differently by different receivers of such data. This is because the receivers of data might differ in their skills and prior knowledge, the techniques and technologies available for the analysis of data etc. (OECD, 2015).

Since, the value of data is context-dependent, it questions the appropriateness of value-based pricing. The pecuniary value of the same data set can be different depending on the market participant in question. Research conducted on measuring the monetary value of personal data indicates that the monetary value attributed to the same data set can differ vastly among market participants. Economic experiments and surveys conducted in the United States of America (USA) showed that individuals were ready to provide the details of their Social Security Numbers (SSN) for an average of USD 240. However, the same data sets were available through data brokers such as Pallorium and LexisNexis for less than USD 10 (OECD, 2013).

It might be possible to define meaningful markets for data with a single price when the classes of data varieties are narrow[3]. However, data is not homogeneous and can be differentiated on the basis of a large number of attributes. Therefore, the degrees of complementarity or substitutability of data will have an effect on its price. Different varieties of data might be ***complementary***[4] because merging them might result in the discovery of relationships that could not have been inferred from examining the data in isolation. Data might be ***substitutable***[5] if it shares certain general characteristics, such as describing similar populations along a common aspect of interest (IMF, 2019).

3.2 The assumption that data can be "owned" or be treated like "property" is not tested. Various frameworks support broader control-based approaches to data rather than defaulting to ideas of "ownership"

According to Section 5.1(ii) in the Report (p. 23), for intangible assets like data, the term ownership relates to "*a set of primary economic and other statutory rights*". Additionally, Section 5.3(iv) (p. 25) states that the allocation of these rights has to be operationalised through the notion of "*beneficial ownership/interest*" to safeguard the interests of the community to which the data belongs. However, the Report has conferred this new right without taking into consideration the various dimensions of data ownership and the subsequent effects of conferring such rights. There are different forms to data ownership, and it is congruous with a variety of normative targets and background assumptions which the Report has failed to examine in detail (Hummel, Braun, & Dabrock, 2020).

It is found that creating a data ownership right is extremely laborious and encodes a variety of concerns (Hummel, Braun, & Dabrock, 2020). Primarily, the creation of such a right would require data to be defined for the purposes of its application (Scassa, 2018). As outlined in Section 2.2. of our response, the definitions of the three categories of data as stated in the Report are not conclusive and

[3] The variety of data indicates unstructured data sets from diverse sources such as social media, financial transactions, mobile communications etc. Variety is concurrent with the capacity to link diverse datasets (OECD, 2015).

[4] A complementary good or service is an object used in combination with another good or service. When the price of a particular good rises, then the demand for its complement is likely to fall as consumers are unlikely to use the complement in isolation **Invalid source specified.**.

[5] A substitute is a good or service that can be used in place of another, as they are similar. When the price of a particular good rises, then the demand for its substitute is also likely to rise **Invalid source specified.**.

have scope for overlaps. This has implications for the ownership rights proposed by the Report, as it becomes difficult to assess to whom the rights should be assigned.

Secondly, ownership rights are considered a blunt tool to address competing interests. Competing interests on data forms one of the fundamental problems in creating a data ownership right, which makes it difficult to 'locate ownership' (Farkas, 2017). There arise competing interests even in the case of personal information. For example, the medical history of an individual, which includes information about their DNA, could also be considered as personal information of that individual's children. Other instances of competing interests include interests of the entity that collects personal information, and interests of the subject of the personal information.

There also arises the question of whether the rights should be based on the source of the information or the resources invested in defining the parameters and collecting the information. Another instance would be in the context of smart cities. The competing interests there would include the interests of: the company providing the hardware that collects the data, the entity that derives data from the data that has been captured, the source of any other data used in the procedure of creating new data, the city that gives access to common spaces from where the data is collected etc. There needs to be some advanced considerations regarding such complexities before a new right is created (Scassa, 2018).

Additionally, ownership rights do not account for all the rights and interests associated with data (Scassa, 2018). For instance, the different types of interests in agricultural data involves farmers who could benefit from data on crop sowing, whereas environmental interests might be served in developing safe and sustainable agricultural practices (Joshi & Ruhaak, 2020).

Critics of data "ownership" also highlight there are certain significant disparities between data and the paradigm cases of property. According to Zech (2012), possession or ownership of data presupposes tangibility. However, unlike tangible entities, the ownership of data does not indicate that one is the sole owner and user of the data. Several people can use data at the same time as it can be duplicated, and therefore it is onerous to exclude third parties. Moreover, data cannot be depleted as it can be used repeatedly without loss in quality (Hummel, Braun, & Dabrock, 2020).

Further, ownership comprises of a bundle of rights as opposed to a single right. It can be considered as a proxy for certain access, usage, and control rights (Hummel, Braun, & Dabrock, 2020). According to Honoré (1961) 'ownership' can comprise of the following rights and duties: "(i) the right to possess, (ii) the right to use, (iii) the right to manage, (iv) the right to the income of the thing, (v) the right to the capital, (vi) the right to security, (vii) the rights or incidents of transmissibility and absence of term, (viii) the prohibition of harmful use, (ix) liability to execution, and (x) the incident of residuarity". These rights and duties are considered to be jointly sufficient but not individually necessary. Full ownership would be satisfying all of them; however, (less-than-full) ownership can come from the satisfaction of some of these stipulations (Honoré, 1961). Therefore, ownership imparts imprecise information on the control that an owner has over her resource (Hummel, Braun, & Dabrock, 2020). For example, although the owner of a building in a historic district might have certain entitlements to control and use the resource, certain restrictions would still apply wherein he might not be able to tear down the building and replace it with a skyscraper (Waldron, 2017). Similarly, even if a patient owned their health data, the state might still hold certain rights to non-consensual access, especially if such access was in the legitimate interest of the public (Evans, 2011). Therefore, what becomes essential are the particular rights implicated by ownership rather than ownership per se (Hummel, Braun, & Dabrock, 2020).

Therefore, the idea of “ownership” for data creates expectations and attitudes which are worrisome. This is a misunderstanding that might arise from unduly simplistic models of data ownership, transferability, and the value of data (The British Academy & The Royal Society, 2017). For instance, useful and new applications cannot be supported by the raw health information of a patient, as it is not valuable in itself. The creation of useful data resources involves considerable inputs from both human and infrastructure services. Therefore, data in itself is fruitless unless it is combined with the necessary services (Evans, 2011).

Especially in India, following the *Justice Puttaswamy & Anr v. Union of India & Ors (2017)* matter, informational privacy has been enshrined as a fundamental right, placing it on par with human rights which do not fit well with a framework of extreme commodification. Any regulatory approaches to personal or non-personal data governance therefore must move away from notions of “ownership” which are not well-suited to the reality of how data flows, or to the claims of control that various parties – and especially individuals and groups – will have with regard to information that represents them.

References

Abrahamson, Z. (2014). Essential Data. *Yale Law Journal*, *124*(3). Retrieved from https://digitalcommons.law.yale.edu/ylj/vol124/iss3/7

Article 29 Data Protection Working Party. (2014). *Opinion 05/2014 on Anonymisation Techniques.* European Commission . Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Blankertz, A. (2020, February). *Designing Data Trusts*. Retrieved from Stiftung Neue Verantwortung: https://www.stiftung-nv.de/sites/default/files/designing_data_trusts_e.pdf

Competition Commission of India. (2020). *Provisions Relating to Abuse of Dominance.* Retrieved from https://www.cci.gov.in/sites/default/files/advocacy_booklet_document/AOD.pdf

Court of Justice of the European Union. (2016). *Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland.* Retrieved from https://curia.europa.eu/jcms/upload/docs/application/pdf/2016-10/cp160112en.pdf

Determann, L. (2018, February 14). No One Owns Data. *UC Hastings Research Paper No. 265*. Retrieved from https://dx.doi.org/10.2139/ssrn.3123957

Dvara Research. (2020). *Comments to the Joint Parliamentary Committee (JPC) on the Personal Data Protection Bill 2019.* Retrieved from https://www.dvara.com/research/wp-content/uploads/2020/03/Dvara-Research-Final-Submission-Comments-to-the-Joint-Parliamentary-Committee-on-PDP-Bill.pdf

Engelsman, W. (2009). Information Assets and their Value. University of Twente. Retrieved from https://www.semanticscholar.org/paper/Information-Assets-and-their-Value-Engelsman/a9c3f38f978b124d077fa99fd169e5c2dfb28a65

Erickson, K., & Sørensen, I. (2016, June 30). Regulating the Sharing Economy. *5(2)*. Internet Policy Review. Retrieved from https://policyreview.info/articles/analysis/regulating-sharing-economy

European Commission. (2019). *Guidance on the Regulation on a Framework for the Free Flow of Non-personal Data in the European Union.* COM(2019) 250 final, Brussels. Retrieved from https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-250-F1-EN-MAIN-PART-1.PDF

Evans, B. J. (2011, June). Much Ado About Data Ownership. *Harvard Journal of Law and Technology* , *25*. Retrieved from https://ssrn.com/abstract=1857986

Farkas, T. (2017, August 01). Data Created by the Internet of Things: The New Gold Without Ownership? . *23*. Revista La Propiedad Inmaterial. Retrieved from https://ssrn.com/abstract=3012155

Government of India . (2019). *The Personal Data Protection Bill, 2019.* New Delhi: Government of India .

Graef, I. (2016). Data as Essential Facility: Competition and Innovation on Online Platforms . KU Leuven Centre for IT & IP Law. Retrieved from https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1711644&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1

Graef, I., Gellert, R., & Husovec, M. (2018). Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation. Retrieved from http://ssrn.com/abstract=3256189

Honoré, A. M. (1961). 'Ownership' in Guest, A. G. (ed.), Oxford Essays in Jurisprudence. 107-147. Oxford University Press.

Hummel, P., Braun, M., & Dabrock, P. (2020, June 15). Own Data? Ethical Reflections on Data Ownership. Retrieved from https://link.springer.com/article/10.1007/s13347-020-00404-9#citeas

IMF. (2019). *The Economics and Implications of Data: An Integrated Perspective.* Strategy, Policy, and Review. Washington, DC: International Monetary Fund. Retrieved from https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2019/09/20/The-Economics-and-Implications-of-Data-An-Integrated-Perspective-48596

Joshi, D., & Ruhaak, A. (2020, August 11). Too Many Questions Remain Unanswered in India's Proposal to Regulate Non-Personal Data. *The Wire*. Retrieved from https://thewire.in/tech/india-non-personal-data-regulation-amazon-facebook

Kammourieh, L., Baar, T., Berens, J., Letouzé, E., Manske, J., Palmer, J., & Vinck, P. (2017). Group Privacy in the Age of Big Data. In L. Taylor, & L. &. Floridi, *Group Privacy New Challenges of Data Technologies* (pp. 37-67). Springer International Publishing.

Kerber, W. (2020, August 20). *From (Horizontal and Sectoral) Data Access Solutions towards Data Governance Systems*. Retrieved from Philipps-University Marburg: https://www.uni-marburg.de/en/fb02/research-groups/economics/macroeconomics/research/magks-joint-discussion-papers-in-economics/papers/2020-papers/40-2020_kerber.pdf

Koutroumpis, P., Leiponen, A., & Thomas, L. (2020). Markets for Data. *Industrial and Corporate Change*, 1-16. Retrieved from https://www.researchgate.net/publication/338411973_Markets_for_Data

Lakshmikumaran, C., & Sandeepan, N. (2020, June 19). Regulatory Tussle: Competition Commission of India v. Controller of Patents & Ors. Lakshmikumaran & Sridharan Attorneys. Retrieved from https://lakshmisri.com/insights/articles/regulatory-tussle-competition-commission-of-india-v-controller-of-patents-ors/#

Medical Research Council. (2019, September). GDPR Guidance Note 5: Identifiability, Anonymisation and Pseudonymisation. Retrieved from https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-5-identifiability-anonymisation-and-pseudonymisation/

MeitY. (2019). *Personal Data Protection Bill, 2019.* Retrieved from http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

MeitY. (2020). *Report by the Committee of Experts on Non-Personal Data Governance Framework.* Retrieved from https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf

Ministry of Corporate Affairs. (2019). *Report of the Competition Law Review Committee.* Retrieved from http://www.mca.gov.in/Ministry/pdf/ReportCLRC_14082019.pdf

Mittelstadt, B. (2016). From Individual to Group Privacy in Big Data Analytics. *Springer*, 475-494.

Narayanan, A., & Shmatikov, V. (2008). *Robust De-anonymization of Large Sparse Datasets.* 2008 IEEE Symposium on Security and Privacy.

OECD. (2011). *Quality Framework and Guidelines for OECD Statistical Activities.* Retrieved from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs(2011)1&doclanguage=en

OECD. (2013). *Exploring the Economics of Personal Data A Survey of Methodologies for Measuring Monetary Value.* OECD Digital Economy Papers. OECD Publishing. Retrieved from https://doi.org/10.1787/5k486qtxldmq-en

OECD. (2015). *Data Driven Innovation Big Data for Growth and Well-Being.* OECD Publishing. Retrieved from https://doi.org/10.1787/9789264229358-en

OECD. (2019). *Data in the Digital Age.* Retrieved from https://www.oecd.org/going-digital/data-in-the-digital-age.pdf

OECD. (2019). Enhancing Access to and Sharing of Data Reconciling Risks and Benefits for Data Re-use across Societies. Retrieved from https://doi.org/10.1787/276aaca8-en

Patrick Breyer v Bundesrepublik Deutschland, C-582/14 (2016). Retrieved from https://curia.europa.eu/jcms/upload/docs/application/pdf/2016-10/cp160112en.pdf

Peter Nowak v Data Protection Commissioner, C-434/16 (Court of Justice of the European Union December 20, 2017). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CJ0434

Pitofsky, R., Patterson, D., & Hooks, J. (2002). The Essential Facilities Doctrine under US Antitrust Law. *Antitrust Law Journal , 70,* 443-462. Retrieved from https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1342&context=facpub

Richter, H., & Slowinski, P. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *International Review of Intellectual Property and Competition Law, 50,* 4-29. Retrieved from https://doi.org/10.1007/s40319-018-00777-7

Scassa, T. (2017, November). Sharing Data in the Platform Economy: A Public Interest Argument for Access to Platform Data. *54, 4*, 1017-1071. University of British Columbia Law Review. Retrieved from https://www.researchgate.net/publication/321427046_Sharing_Data_in_the_Platform_Economy_A_Public_Interest_Argument_for_Access_to_Platform_Data

Scassa, T. (2018, September). Data Ownership. *CIGI Papers No.187*. Centre for International Governance Innovation. Retrieved from https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

Scottish Government. (n.d.). Retrieved from https://www2.gov.scot/Topics/Statistics/datalinkageframework/Whatdatalinkageis

Short, J., & Todd, S. (2017). What's Your Data Worth? . *MIT Sloan Management Review, 58*(3). Retrieved from https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/short-whats-your-data-worth.pdf

Singh, A., Raghavan, M., Chugh, B., & Prasad, S. (2019, September 24). The Contours of Public Policy for Non-Personal Data Flows in India . Retrieved from https://www.dvara.com/blog/2019/09/24/the-contours-of-public-policy-for-non-personal-data-flows-in-india/

Singh, M. (2020, August 19). Non-Personal Data Governance Framework and Competition Act. Retrieved from https://www.livelaw.in/columns/non-personal-data-governance-framework-and-competition-act-161663

Solove, D. J. (2012, November). *Privacy Self-Management and the Consent Dilemma*. Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2171018

Stucke, M., & Grunes, A. (2016). *Big Data and Competition Policy.* Oxford University Press. Retrieved from https://www.researchgate.net/publication/308970973_Big_Data_and_Competition_Policy

Sweeney, L. (2000). *Simple Demographics Often Identify People Uniquely.* Carnegie Mellon University. Retrieved from https://dataprivacylab.org/projects/identifiability/paper1.pdf

Taylor, L., Floridi, L., & Sloot, B. v. (2017). *Group Privacy New Challenges of Data Technologies.* Springer International Publishing.

The British Academy & The Royal Society. (2017). *Data Management and Use: Governance in the 21st Century.* Retrieved from https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf

The Competition Act. (2002). Retrieved from https://www.cci.gov.in/sites/default/files/cci_pdf/competitionact2012.pdf

Tucker, C. (2018). Network Effects and Market Power: What Have We Learned in the Last Decade? . *(Spring 2018)*. Antitrust Magazine, American Bar Association. Retrieved from http://sites.bu.edu/tpri/files/2018/07/tucker-network-effects-antitrust2018.pdf

Tucker, C. (2019, January 31). Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility. Review of Industrial Organisation Special Issue on Antitrust and the Platform Economy. Retrieved from https://ssrn.com/abstract=3326385

Varian, H. (2018). Artificial Intelligence, Economics, and Industrial Organization. *NBER, Working Paper No. 24839*. Retrieved from https://www.nber.org/papers/w24839

Waldron, J. (2017). Property and Ownership. *The Stanford Encyclopaedia of Philosophy, In E. N. Zalta (Ed)*. Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/entries/property/

World Economic Forum. (2019). *Competition Policy in a Globalised, Digitalised Economy.* Platform for Shaping the Future of Trade and Global Economic Interdependence. Retrieved from http://www3.weforum.org/docs/WEF_Competition_Policy_in_a_Globalized_Digitalized_Economy_Report.pdf

Zech, H. (2012). Information als Schutzgegenstand. *XXV*, 488. Jus Privatum Contributions to Private Law. Retrieved from https://mohrsiebeck.com/buch/information-als-schutzgegenstand-9783161521621?no_cache=1