



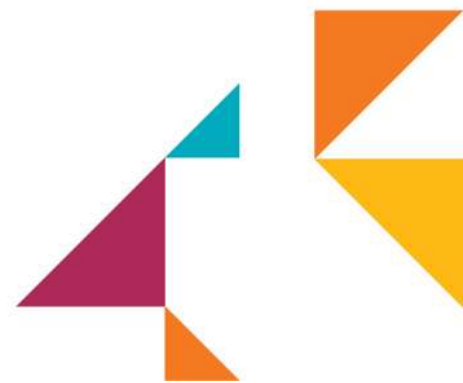
Policy Opinion

Comments on the Draft National Health Data Management Policy

BY

CENTRE FOR THE DIGITAL
FUTURE

NOVEMBER, 2018



1. Digitalizing the health sector can improve vastly the working of the sector in India, bring in innovation, investment, growth and employment along with an improvement of the health status of all Indians. At the same time, the health sector is prone to market failure and hence requires specific policy measures and, often, government's own participation in either financing, or provisioning, or both. Health markets fail for at least two reasons: (a) an individual's health status is not independent of that of others e.g., in the case of infectious diseases, something that in the market failure literature is referred to as externality and (b) healthcare professionals know more about the health of their clients than the clients themselves, referred to in the literature as asymmetric information. Market failure essentially implies that left to themselves, individuals are unable to generate desirable outcomes for themselves through existing market mechanisms. Hence, one needs public policy interventions --- collective, rather than individual, action --- to improve the welfare of market participants.
2. This is important to understand why electronic health records (EHR) and a digital health eco-system are so desirable. These records and their analyses are essential to address the causes of market failure, referred to above. While the easy access to my health history enables the healthcare professional to provide me better services, this same information when aggregated with others' information and then analysed, improves the health status of all. For the latter, it is essential that everyone submits their data for analysis. On the other hand, since health data is sensitive, given a choice, I will not want anyone barring those I allow, to access this data. To address this problem, it is necessary to develop a data collection ecosystem in which no one using the data could relate it back to any individual, unless the individual wants. In other words, the focus of the document should have been on ensuring "privacy", or anonymity of the data principal, rather than on the data principal's ability to exercise "choice".
3. It is important to understand how the implications of the two approaches differ. The approach in the document allows individuals and entities to opt out of the system. This destroys the entire purpose of EHR --- generating value through aggregation of data. My benefit, or incentive, to submit my data for aggregate analysis is less if more people stay out of the system. If the objective was, instead, to ensure privacy, the document would have focussed more on how to "mandate" data collection, storage and use in a way that ensures the anonymity of individual health records. This would ensure that everyone's data was being used to generate the maximum possible value. The objective of any public policy intervention in the market place is to regulate (restrict or enhance) an individual's choice-set; otherwise, there is no need of public policy interventions!
4. The policymakers need to ask themselves two questions: (a) Why are we trying to develop the EHR of all citizens? (b) Are the policy proposals the best way of reaching the target? First, note that the health sector is not like any other sector. The government can make a set of rules for the automobile market that enable investment, employment and growth in that market in a way that improves aggregate GDP. It is perfectly all right if some people do not buy cars out of choice, or cannot afford them. Obviously, the health sector is not the same as the automobile sector. Everyone in the society should be able to afford healthcare and nobody should be allowed to *choose* to stay unhealthy and refuse healthcare as one's ill-health can affect others negatively. Seen in this light, the power of the data principal to prevent his/her data from being used for activities he/she does not approve of is exactly the opposite of what the health policy objective is. In other words, the purpose of any policy on health is very different from that of the sectors that are addressed, say, in most industrial policies.

5. Of course, privacy is an issue and needs to be protected. But that is exactly what the personal data protection laws should ensure. The requirement for a policy on health data is precisely because personal data will have to be used for the greater good. And, hence, the focus should be on how to protect the identity of the data principal and not on the ability of the data principal to exercise choice over the use of his/her health data!
6. And, if the policymakers treat the health sector like any other sector --- like automobiles, say --- even then the policy proposals are flawed. Currently, health data record keeping is at its infancy in India. Various startups and other entities are in the process of experimenting with various systems of data collection. In other words, the digital healthcare market system is yet to evolve and stabilize. In the meantime, we are already making a policy of how the digital health market should operate. The purpose of regulating markets is to stop unfair business practices (unfair to consumers or other competitors both, incumbents and potential entrants) and *not* to decide how a market should develop --- a practice that has more in common with our pre-1991 “planning” experiences.
7. Digitalization of the health sector in India can deliver enormous value not only to individuals, but also to public health establishments, healthcare providers, pharmaceutical companies, diagnostic labs and other stakeholders **without violating privacy or security of data**. This can be achieved by making anonymized data at aggregate and individual level available and easily accessible to various stakeholders. Invaluable insights can be gained into issues like disease prevalence, which treatments/medicines are working or failing and in what conditions, etc. No useful purpose will be served by making such benefits contingent on “individual consent” of people who in any case, are completely unaffected by use of their anonymized data. In fact, individuals benefit only when their data is aggregated.
8. Similarly, enormously beneficial customized information and services can be provided to individuals based on their condition and needs if their individual data is shared. Obviously, this can and should be done if and only if the individual consents to sharing his/her PII. Even in such cases, recourse to reverse anonymization through the anonymizer as a reciprocal operation can enable such personalized services to be provided even without violating individual privacy. This may need to be examined technically and legally.
9. Some specific gaps and observations regarding the draft are listed in the annexure to this note. However, simply addressing the issues listed at annexure will not serve the purposes indicated above. The entire policy will need to be recast based on the following principles:
 - i. Limit protection of privacy of individuals to PII (Personally Identifiable Information as described in the draft PDP Bill). Dispense with the need for individual consent for use of anonymized data.
 - ii. Enable a vibrant anonymized health data eco-system built on sound economic principles to grow naturally
 - iii. Enable easy and secure access to anonymized data to entrepreneurs and innovators
 - iv. Maximize value of anonymized digital health data to public health authorities, caregivers, pharma companies and other stakeholders

Therefore, the overall approach we recommend is to create a vibrant, well-regulated health data eco-system that is built on sound economic principles and so grows naturally, without any compromise on privacy. This alone can deliver on the promise of digital healthcare.

NOTE: A recent CDF publication on Data Sharing (CDF Whitepaper 01/2020) forms the conceptual basis on which the views shared in this paper are founded. That publication is also attached for ready reference.

ANNEXURE

Our more specific comments are:

1. Health is a State subject and States have different State capacities, be they financial or human resources. If the coverage is low, digitization of health records would simply increase the cost of the healthcare system without generating any significant benefits.
2. Overall, the national data management policy has taken the shape of a health data privacy and security policy rather than a comprehensive health data policy. It lays down compliance requirements rather than providing an enabling and supportive framework that aligns the interests of all stakeholders in the digital health eco-system and paves the way for collaborative implementation of NDHB by all players.
3. At places, the policy seems to implicitly assume that there is a single entity (health facility acting alone or in conjunction with another technology provider as a data fiduciary) handling the data of a data principal. The reality may often be of multiple health facilities and multiple data fiduciaries handling data of a single individual. The federated architecture envisaged in the NDHB inevitably has similar consequences. This could lead to the need for a data aggregator who collates the data of an individual from multiple sources on top of which other VAS providers could build personalized health services for that individual. In other words, the approach needs to be based on an assumption of an inter-connected eco-system in letter and spirit as envisaged in the NDHB.
4. Apropos point 3, is it practical to require the data principal to give consent at each and every entity that uses his/her data?
5. The sector has large hospitals with enough resources and single doctor chambers in small towns with little or no financial resources serving clients who are barely able to afford institutional healthcare. The smaller entities would simply bear the costs of the new system, should they opt for it, without being able to extract much of the benefits. This would encourage them to stay out of the system and that defeats the entire purpose of the EHR.



Centre for
The Digital Future

PAPER

DATA SHARING

CDF WHITEPAPER 01/2020

AUGUST 2020



TECHNOLOGY
& DEVELOPMENT

0.1 The purpose of this brief is to lay out the basic economic principles that govern the sharing of data, held by a private entity, with another private or government entity. The fundamental points being made in this brief are three-fold --- (a) whenever private entities can generate value for themselves by sharing data, they will do so (b) there are instances where sharing data among competitors can improve value for all and, hence, no affirmative policy is required; there are other instances where it decreases value for consumers and needs to be restricted, not encouraged and (c) a mandatory data sharing policy runs the risk of discouraging efficient data capture and reduces innovations in the digital economy. There are existing laws in the system that enable the government to access data held by private entities when such data are needed for the greater good of society, e.g., in times of natural disasters or calamities, national security, prevention of money laundering, etc. Indeed, as our recent experience during the Covid19 pandemic has

shown, it is often in the interest of private data holders to make relevant data (like mobility data by Google and Facebook) publicly available to help in policy formulation.

0.2 There are two other issues that have been raised on matters pertaining to data generated through transactions carried out by Indian nationals. First, according to some, this data belongs to all Indians and should be treated as a national resource in the same way as we treat oil, coal and other minerals. Second, even though the data is generated by transactions in India, or by Indians, it is held

largely by commercial entities that do not belong to India. Together, they imply that Indian data should be controlled by Indian entities rather than by non-Indian entities as is largely the case at present. Indian control of the data will enable the controller to demand (a portion of the) value generated by the data holder for Indian nationals and commercial entities in a way similar to that of other national resources. Without going into the merits of the argument, it is worth pointing out that even for national resources owned by the government, private parties have

This brief makes three fundamental points

- a. Whenever private entities can generate value for themselves by sharing data, they will do so.
- b. In instances when sharing data among competitors can improve value for all, no affirmative policy is required; in instances where sharing decreases value for customers, it needs to be restricted, not encouraged.
- c. A mandatory data sharing policy runs the risk of discouraging efficient data capture and reduces innovation in the economy.

been allowed to exercise control over their use and distribution while sharing the value from such activities with the government --- be they oil and gas, or coal. In other words, whether or not the value generated by Indian data should be shared with the Indian government is an issue different from the data sharing discussion.

- 1.1 Data sharing is related to, and often referred to, in the contexts of data privacy, data sovereignty, data localization, data portability, etc., all important elements in the nation's overall data policy. There is little doubt that how we decide on each of these has implications for the others in terms of how we can use the new digital technology to extract the maximum benefit for our citizens. However, we feel that it is a good idea to first understand how the policy that governs any one of them, say data sharing, holding the others as given, will generate the social welfare we are after. Once that is done, we can then consider how tweaking policies in the other dimensions will affect the surplus generated through data sharing. E.g., it is important to understand how a particular type of privacy policy affects the nature of data sharing and, hence, the surplus realized by data sharing. Alternatively, disregarding data security issues may nullify whatever we are gaining through data sharing. This white paper on data sharing will follow this approach.
- 1.2 The generation, capture and sharing of data, as well as the intelligence developed from analysing the data, have been going on for a long time between and across governments and non-government entities. For example, macro-data collected and put out by governments have helped businesses plan their long-term and immediate strategies while company disclosures have had implications for policymaking in governments. In both cases information culled from the published (or shared) data was, and is, being used by all parties involved in sharing the data.
- 1.3 While these are obvious, there are more sophisticated data sharing instances in the pre-digital world. In USA, American College of Surgeons started the process of collecting and sharing clinical information among the healthcare professionals to improve the delivery of healthcare services that not only helped patients coming to a particular clinic but also doctors in clinics throughout the country. Such information exchange about patients and the care they received helped evolve healthcare protocols and best practices. This process was started in 1928. The New York Stock Exchange, set up in 1792, is another example of data sharing where

(certain types of information) on share transactions of each company on the exchange is available for everyone to see.

- 1.4 Observe that in both instances, competing entities --- private doctors competing for clients and private companies competing for funds --- were willingly sharing data among themselves. In other words, whenever economic agents and entities find value in sharing their proprietary data, they will. And, hence, one essential element of a data sharing policy is to create an ecosystem where such mutually beneficial exchanges of data are facilitated, rather than hampered. A key element of such value enhancing data sharing is the ability of the entity participating in data sharing, to extract part of the value generated from the dataset generated through data sharing.

Data capture, storage and aggregation are costly activities. Thus without a return on this cost, data will not be generated and datasets will not be created.

Since data capture and data storage are costly activities, without a return on that cost, data will not be generated and datasets will not be created. It is worth emphasizing that data by itself does not create value, structured data, or datasets, do. ¹Creating these datasets does not happen by itself but requires time and effort of economic entities. In the digital world, data may get automatically generated by any transaction but datasets still require investment in time and effort and, hence, dataset creating activity has to be compensated.

- 1.5 There is a distinction that is often made between “raw” data and “processed” data.² This distinction is made to make the point that “raw” data should be shared “free

¹ By structured data we mean any dataset into which some prior thought has gone in to decide what to collect, why to collect and how to access for any type of later analysis. Unstructured data would be random images, PDF files, videos (as in Youtube), etc.

² Anonymised personal “raw” data should also be considered as processed data since anonymization is a process that is often a bit complicated. E.g., the NSSO (consumer) household survey not only anonymises the household but also the village where the household resides. Villages are sampled first and then within villages households are selected from three (second stage) strata. In some villages the households in one stratum could be less than the number to be selected. In such cases, the households can be identified quite easily even when their identities are hidden. Hence, the NSS hides the village name too. It is, therefore, important to keep in mind that anonymization itself is a complicated process depending on the nature of the data and the procedure by which it was collected.

The distinction made between 'raw' and 'processed' data is not sound. Any data that is captured, as well as those that were not, contain informative signals regarding the data collector's potential business strategy. This is a 'secret' business enterprises may want to protect. Indeed, if this secret cannot be protected, market research companies will go out of business!

of charge" while "processed" data can be paid for. What data are captured by an entity, as well as what are not, are both informative signals regarding the data collector's potential business strategy and, hence, a "secret" that the business would want to protect. So, it makes very little business sense for them to let everybody know what data they are capturing. Indeed, this might kill all market research companies as they might run out of clients asking them to conduct surveys!

- 1.6 A second point we would wish to make here is the fallacy behind the thought that if some entity is willing to pay for data, the holder is obligated to sell it as long as its cost is covered. One of the basic determinants of the market price of any item is the ability of the potential supplier to refuse sale. It is the market price that determines the volume of sales and, in most cases, the number of sellers. If the government intervenes and decrees that certain types of data if processed must be shared with all at a price, it will generate an unhealthy incentive for all and sundry to collect data, process them and put them out in the public domain at a price to be paid by the government. This will be an extension of the logic behind the existing minimum support price (MSP) in agriculture and the cost-plus administered price regime we had till the end of the 1980s. Both have had disastrous results in

A key determinant of market prices and the efficiency of markets is the ability of the potential supplier to refuse sales. The market price thus determines the volume of sales. Government intervention, in setting prices for certain types of data, creates perverse incentives to collect and process data that is not useful, since sales are guaranteed.

- making India an inefficient and high-cost economy. Pricing determined by governments and not markets, results in an innovation-proof economy. In other words, mandatory sharing of processed data at government controlled prices is not a panacea for the data economy but a death-knell for all data based innovations.
- 1.7 There are situations where data sharing may actually be value destroying. E.g., potential competitors sharing pricing strategies among each other leads to cartelization and reducing consumer surplus even when improving the surplus

enjoyed by those sharing such data. Another instance of restricting information flow (or preventing data sharing) is the mandatory requirement of anonymous trading on stock exchanges (excepting when takeover rules kick in) because market knowledge about who the buyer or seller of shares is prevents the stock market from functioning effectively.

- 1.8 It is important to understand that the data sharing curbs referred to here have very little to do with society's privacy requirements! E.g., a company trying to cartelize would prefer to have its identity revealed along with its pricing strategy, but it is not allowed to do so. Similarly, in stock trading, information about who is transacting on the exchange dissipates information in a way that harms the generator of the transaction and, hence, prevents such value enhancing transactions. In other words, maintaining anonymity is crucial to the functioning of efficient markets. The same would be true for personal health data also. E.g., information that leads to identification of a HIV positive person, could lead to discrimination by others who come to

It is not always the case that data sharing is desirable. A company trying to cartelise would prefer to have its identity and pricing information revealed, but it is not allowed to do so. Such concerns have little to do with privacy. Keeping these data private improves economic efficiency and social well-being.

know of her/his health status. This is quite independent of whether or not a HIV positive individual wants to be identified. In these instances, anonymity is derived from the requirement of efficient markets; anonymity here is a means to efficient markets and not a goal. In many Northern European countries, individual tax information is public knowledge, whereas individual trading in the stock exchanges is not. In India, neither of the two is public knowledge. Anonymity in stock transactions is required for market efficiency; whether or not to allow others to see one's tax returns is a function of a society's preference. In the debate on privacy, it is important to distinguish between the two.

- 2.1 But these are all examples of practices that were already there in the non-digitized world. Why are these questions coming up again now? There are, at least, three clear reasons. All of them are related to the digitalization prospects in a digitized world. First, never has society been able to capture such large volumes of usable data. This volume is not only in terms of the accumulated stock, but also in terms

of the rate at which they are being accumulated and transmitted (flow). And, most importantly, we now have the machines and the mathematical capacity to analyse this “big data” in ways we never could imagine before. Second, much of this data

Data sharing predates the digital economy. But three features distinguish the current setting from the past.

- a. Volume of data being captured and accumulated, and our ability to transmit and analyse them, have grown manifold.
- b. Individuals to whom this data pertains have little choice on how, or whether, their data are stored, or captured, or used.
- c. Data are an essential infrastructure in a digitalised world and, hence, are instruments of commerce. To realise the full potential of an infrastructure, access to it cannot be restricted to any economic activity.

relate to various transactions carried out by individuals who have little, or no, choice on how, or whether, their data are stored, or captured, or used. An individual making an online purchase cannot prevent that data being generated and stored by the platform on which the purchase is being made. If the individual does not want this, s/he has to desist from such a transaction, an option that is becoming increasingly scarce in a digital world. The third reason is that data are being considered as an essential infrastructure in a digitalized world. Infrastructure is an instrument of commerce, its ability to create value depending on the resourcefulness and

innovative abilities of those engaged in commerce. Therefore, to realize the full potential of an infrastructure, every entity must have access to it.

- 2.2 Data of all transactions on a platform carried out by 10 individuals in one year has no value. If the same data are available for a million, or 10 million individuals, this data has immense value. An entity that has the larger dataset has an immense advantage over rivals with smaller datasets and can successfully prevent the latter from maintaining their market shares. This leads to market dominance by one player. This dominance by incumbents with large datasets prevents entry by new players with no datasets. That hampers competition as well as innovation. But, is mandatory sharing of the incumbent’s data with all potential entrants the way to go? It assumes that there are no new data to be generated, or new datasets to be

created. Mandating data sharing with competitors will kill the incentive to generate new data and create datasets. These are costly activities for the private sector. The only way they can get a return from investing in such activities is by excluding others from using the data and dissipating away the surplus that could be enjoyed by the generator and creator of datasets. Mandatory data sharing is a myopic approach based on the erroneous presumption that all data (and data sets) have already been created. On the other hand, the competition authority is already there to monitor whether dominance

Mandating data sharing to counter future market dominance is a myopic approach. It assumes that all data and datasets that could be created have been created. Should an entity dominate a market because of its proprietary data, the competition authorities can always deal with it to prevent abuse of that dominance. Dominating a market is never forbidden; if it leads to abuse or restricts competition, the competition authority is the right institution to deal with it.

in the market place by the data holder is affecting competition and innovation. Mandating compulsory data sharing is not only an overkill, it destroys the very process of data creation.

- 2.3 The second reason for the current discourse regarding the data economy (mentioned in 2.1 above), is the fact that this is personal data from which the data holder is extracting value while the individual whose transaction generates this data is unable to extract any part of this value. This is an erroneous approach.

Individuals whose data are being collected and aggregated do not automatically have a claim to a part of that value. This is because value is being created through the processing of this data. By itself, the individual's data is not valuable to others.

Suppose a private hospital generates, stores and analyses its client data to improve its healthcare services. This allows it to attract more clients and generate more profit. If we argue that the client's data is owned by the client and cannot be used by a hospital, the hospital will have no interest in collecting the data. In as far as patient data enables the hospital to provide better care to its

clients, this value generation is lost. Just as my data is being used to generate value for someone else, the use of others' data generates value for me the next time I visit the hospital. This argument holds even if the hospital generates value from the data by selling it to some other entity like a pharmaceutical company. Most

importantly, observe that if data is mandatorily shared, it does not guarantee any value to the individual.

2.4 Market research, loyalty points, TRPs on television programming, etc., are all various ways data are used to generate value for the company. Market research could help in better targeting customers and in formulating strategies regarding prices, product development and investment. Generation and analysis of such data are costly. As long as it improves the value of services received by the customers, such data are valuable. However, if companies were forced to share this data with their competitors, because it is data on individual clients, or personal transaction data, or for any other reasons,³ the ability to extract a part of the value generated by the company in undertaking the research is lost as its competitors will acquire the same market intelligence without having to pay for it. Competitors become free-riders and, hence, no one company will undertake such research. At the other extreme are TRPs which a company will want to share with everyone, most notably the advertisers on TV programmes because it tells them where their advertisement will have the greatest reach.

2.5 The limited point being made here is that depending on how economic activities are organized, data will either be shared or not. Since data generation is costly, economic entities generating and storing the data are best suited to decide whether the data should be put out or kept hidden.

2.6 Consider the situation where Google through the usage of its maps and, Uber through the usage of its services, each has its own datasets that are being used by them to generate value. Suppose that it is possible for Google to create more value by combining the two datasets. There are two cases to consider here. First, the increase in value to Google from combining the two datasets does not diminish the value that Uber was getting by

If two companies can benefit by combining their data then such data sharing will happen. The two will bargain over the share of the increased value. The ecosystem should be devised so that it keeps transaction costs during the bargaining process to a minimum.

itself. In this case, Google and Uber will bargain between themselves on how this additional value (from combining the two datasets) is to be divided between them.

³ This holds independent of whether the data are anonymised or not.

A good ecosystem will ensure that the transaction cost of such bargaining is as low as possible.⁴ E.g., if the policy mandates when and how such data are to be shared, and that is done without a careful cost-benefit analysis of the mandate, it could make compliance too costly and, hence, restrict the realization of the societal surplus it could generate.

2.7 If two data holders gain from “data sharing” between them, they will, with or without any data sharing mandate by policy. What the government needs to do here, however, is check whether a third party, like the users of Google maps or Uber cab services, are losing value or not. Once again, this is in the realm of competition authorities. Mandatory data sharing is an ex ante “one size fits all” approach while a regulatory approach (competition authority) is an ex post mechanism that fits the regulation to the specific structure and type of market being regulated. Thus, we have restrictions on pricing strategy data being shared among competitors (to prevent cartelization) while we have SEBI forcing disclosure of certain types of company data to all and sundry.

2.8 One of the five properties⁵ of data, like all digital goods, is that it is recombinant. Data generated from a particular type of transaction can be used to generate value in some entirely different activity.⁶ This issue has come up repeatedly in Indian policy discussions. The argument given is that start-up activity in India is being hampered because they do not have access to data that incumbents have. Empirical support for this argument is weak. A recent publication from ADB⁷ mentions India as having the third largest start-up ecosystem globally, totalling

If the incumbent’s data can be used by a start-up to create value, without loss of value to the incumbent, then the two would be able to bargain to come to a mutually beneficial arrangement. And if the arrangement results in loss to the incumbent then how it is to be compensated to preserve its incentives to further collect and aggregate data is best left to the bargaining process in the market-place. Policy needs to step in only if the market mechanism is not working.

⁴ Similar considerations would apply if Uber could benefit from using the data stored by Google regarding the use of its maps.

⁵ The other four are non-rival, infinitely expandable, a-spatial and discrete. See Quah, Danny (2003), “Digital goods and the new economy”, CEPR Discussion Paper No. 3846

⁶ A contemporary example of this would be the Google and/or Facebook mobility data that can be used to understand the behavioural response by individuals to the current pandemic.

⁷ <https://www.adb.org/publications/startup-environment-and-funding-activity-india> (accessed July 18, 2022)

36 billion USD of investment with 26 unicorns. In general, suppose that data possessed by Amazon can be used by a new fashion designer (start-up). In this situation, it makes no sense for Amazon to prevent the fashion designer from using the data unless (a) Amazon loses value because of this (b) Amazon has some strategic interest in preventing the fashion designer to come on-stream. If (a), then it is not clear why mandating data sharing is desirable (see the next paragraph for where this line of reasoning may be questioned). If (b), then the data sharing policy ecosystem needs to enable Amazon to give the data to the start-up. Mandating data sharing destroys Amazon's incentive to collect, store and maintain the data that can be used by the fashion designer. Mandating destroys the very possibility of surplus generation as it takes away the incentive of the data holder to collect, store and maintain the data.

2.9 One situation where (a) above may still warrant Amazon sharing its data with a start-up is when the start-up coming on-stream increases the value to a third party which is some other commercial entity. Once again, as long as all three together produce a surplus that adds up to more than what they could independently produce, they will negotiate with each other without any government mandate.

2.10 The only time such negotiations may not happen is when no part of the value generated by data sharing is extractable by the data holder. A classic example of this is the data on mobility, during the pandemic, put out in the public domain by Google (and Facebook). It may not generate

any direct money value to the publisher of this data, but it creates great value to society as a whole. It enables the health and economic authorities to design policies of containing infections and triggering economic recovery. It is important to note that the mobility data is a dataset that has immense value that has been created from raw data

with zero value. That required the data

The data holder needs to be given incentives to generate socially valuable datasets where private value is difficult to extract. The incentive needs to be generated through a data market. If that is not possible, government should provide the necessary incentive. Forced data sharing, unfortunately, destroys any such incentive.

holder to spend resources on creating the dataset from the raw data and if the government wants such value generating datasets, it should reward the data

holder to give it an incentive to do so. This incentive has to be generated in a data market where the supplier of the data has the right to refuse sale if it is not happy with the price being offered by the buyer.

3.1 It is the responsibility of the policymaker to establish beyond doubt that data sharing is socially beneficial --- i.e., everyone gains or that the policy design has an in-built mechanism that ensures that the losers are compensated sufficiently. This brings us to the “third reason”, mentioned in 2.1 above, why data is important in the digital economy. Data has properties that are similar to “infrastructure” in the non-digital world. The most important aspect of infrastructure is that while it generates value through increased, and new, commerce, it is difficult for the builder of the infrastructure to extract value from those who are creating value from the infrastructure. As far as infrastructure is costly to build, this discourages the private sector from building infrastructure. The government, on the other hand, has taxation powers to raise revenues and, hence, plays an un-substitutable role in infrastructure building.

3.2 Infrastructure has three important features:⁸ (a) it is an input into many commercial activities, both extant and yet to be commercialized (b) it is non-rival in its use and (c) it is used mainly by downstream producers making it an instrument of commerce rather than commerce itself. Data is generated by private party transactions but it becomes valuable only when the data holder creates structured data or datasets. In other words, unlike in the case of non-data infrastructure, private parties have a large role to play in creating the data infrastructure. And, hence, commercial incentives to private players⁹ to develop this infrastructure is going to decide the quality of, and often whether, this infrastructure is generated. Riding roughshod with a universal and mandatory data sharing policy will be detrimental to all future business activities.

⁸ Frischmann, B. (2012), *Infrastructure: The Social Value of Shared Resources* (Oxford University Press).

⁹ In non-digital infrastructure, we have mechanisms like PPP, BOT and its variants, leases on toll collection, etc., to incentivise the private entities to participate in building maintaining infrastructure.

3.3 How and whether data is to be shared should be left to the interplay of market forces. If markets fail to develop, or the data holder's controlling the data creates market power and harms competitors, it is a matter for the Competition Commission of India and not the policymaker. Ex ante regulation of data sharing and data access prevents entrepreneurs from seeking new opportunities. One could always use the

Whether and how to share data should be left to the market. Ex ante regulation of data sharing and data access prevents entrepreneurs from seeking opportunity. Ex post regulation of data sharing prevents anti-competitive outcomes

Competition Authority's regulatory powers to ensure that there are no anti-competitive outcomes due to an entity's proprietary data being excluded from others.

3.4 There is one class of instances where access to data held by a private party may be desirable. Such a situation arises when the use of this data serves a social goal which is often not amenable to monetization and, hence, to the realization of any commercial value. It is best to make laws, or policy, that specifically allows the government, or some independent regulatory authority, to step in and gain access to this data (e.g., for national security, prevention of money laundering, during national calamities, etc.). India already has these laws.

4.1 A society's policy on data sharing must be able to bring benefits to its members that would not be forthcoming without the policy. It is not enough to argue that data sharing is good for any particular sub-group in society. The fact that data held by one entity could be used by another entity to generate value for it and its customers, is not a sufficient argument for asking one to share data with another. If one loses value to another entity by sharing its data, or if no part of the value generated is extractable by the original data holder, no new data entrepreneur will ever emerge in the future, negating the possibilities of generating either data or digital goods.

4.2 Some of the arguments offered for mandatory data sharing are (a) the data are already collected (b) sharing it does not destroy the data (c) another entity will be able to use the data more effectively than the current data holder. While all these may be factually correct in specific sectors and particular types of data, mandating the sharing of all data has serious long-term implications that none of these “facts” take into account. If we mandate sharing of all current data, no new types of data will be generated as the holder of this new data runs the risk of being forced to share this with others and dissipate the rent the innovative data collector was hoping to gain. In trying to develop an innovative economy today, we could be destroying the innovative potential of the economy in the years to come.

The arguments for mandatory data sharing are based on a very static approach. If we are to consider a more dynamic setting, which is necessary to build an innovative economy, we will realize that while mandatory data sharing may lead to an enhancement of, current value, it will have serious adverse effects on the generation of future value. Ex ante regulation of data sharing prevents entrepreneurs from seeking new opportunities. Ex post regulation of data sharing can always be used to prevent anti-competitive outcomes.

4.3 Finally, data sharing has often been linked to three other aspects connected to data --- privacy, data localization and data sovereignty. This has led to the erroneous conclusion that if all data are shared with the government then all these aspects are taken care of.

4.4 First, privacy as a fundamental right (as decided in the Supreme Court) precludes the government from accessing personal data, excepting for specific well-defined purposes. Anonymized data, on the other hand, is no longer traceable to an individual and privacy, therefore, is no longer an issue. The government still does not have a right to it.

4.5 If we interpret data localization as storing data in servers that are physically on Indian soil, it will be highly cost ineffective. The “heat and dust” coupled with the demand for costly energy to keep the servers running 24x7 will be a huge economic cost.¹⁰ More importantly, it is near impossible to prevent the duplication in some other place of the localized data. The main reason for data localization is access to the data generated in India to other Indians, especially when such access serves a public purpose. An example of

Data localization can be very costly both to individual businesses and to the economy as a whole. In other words, not only does it affect the company's bottom line, it also reduces the data localizing country's GDP.

this would be the requirement that personal data on financial transactions of a terror suspect has to be given by the financial institution (domestic or otherwise) to the authority investigating the suspect. Another example would be rushing medical help to one who has an emergency need for it. If data localization is interpreted as access when needed, and if “when needed” is appropriately well-defined by our laws, then this type of data localization will be much less costly to the economy.

4.6 One may argue that data that originates from India should be subject to the laws of the land whether for economic rights, technology development or taxation, apart from privacy and security. This is, indeed, the sovereign right of any nation. However, such laws are intended to improve the welfare of the nation's citizens. Mandatory sharing of data, in most cases, does not do so. Instead, governments need to step in only when there is evidence of market failure and that too, ex-post and not ex-ante (as explained in paragraphs 2.2, 2.6 and 3.3, above). In particular, data sovereignty should not be equated with the government having an over-arching control of and access to data. Indeed, going back to privacy as a fundamental right, sovereignty could also mean denial of any right of government over citizens' data. One should not confuse data sharing with a government's right

¹⁰ A study by The European Centre for International Political Economy concluded that for China, data localization will cost the country 1.1% of its GDP. This would translate to a 13% reduction in a citizen's salary. For the European Union, GDP will fall by 0.4% and total welfare loss will be \$193 billion. When it comes to an individual company, a report by the Leviathan Security Group concludes that data localization will increase data storage costs by at least 30% to as high as 60%. <https://www.itic.org/news-events/techwonk-blog/the-costs-of-data-localization> (accessed August 29, 2020)

to get a hold of its citizens' data. Whenever such access is necessary for improving the aggregate welfare or, preventing its erosion (like information on tax frauds, money laundering, terror financing, etc.) the government can exercise its right to access its citizens' data but that too with transparently legal supervision.

5.1 Mandatory data sharing will prevent innovation in both, data capture and the use of data. There are enough incentives, in the data market to help facilitate different entities to come together even when they exclude others and generate value not only for themselves but also for consumers. The government needs to enable this coming together for mutual benefit, rather than mandating it. India's digitization is still at a

Mandatory data sharing will prevent innovation in both, data capture and the creation of valuable datasets. Rather than a mandate, the government should enable the market for data. Markets can, in most instances, create enough incentives to facilitate different entities to come together and share their data for mutual value generation.

relatively nascent stage and its digital segment is yet to mature and stabilize. Regulating the data economy at this stage of its transformation process will close off innovative avenues of development. It may, at a stretch, help some incumbent businesses but will hamper the entry of potentially innovative game changers.

5.2 If the data held by a commercial entity generates privately extractable value by sharing it with some other commercial entity, a well-functioning data market will make it happen. Well-functioning markets are determined by the rules that govern markets --- e.g., how contracts governing data exchanges among entities are written and enforced. To know what type of market rules the economy should have one must first generate knowledge about how data markets operate in different sectors. It is only after we know this, can we think of designing policies that make these markets efficient. It is this approach that led to rules under SEBI and even these are changing as financial markets keep innovating. Data markets are in a state of flux and generating knowledge on these through research is an essential first step towards designing optimal rules of the market.

5.3 This research first and policy design later (but based on the research findings), becomes even more relevant when sharing the data held by a commercial entity generates privately non-extractable social value. In such cases we need to design policies that enable markets or, develop market type incentives to encourage data

generation that adds social value. We need to know how the business strategies get translated into observable firm behaviour and how these affect the innovation market that generates employment- and growth-inducing investments.

- 6.1 In conclusion, if data held by an entity can generate value for another commercial entity, without reducing the value of the holder, will be traded in the data market. If such sharing reduces the value of the holder, but the data shared with another entity generates a positive surplus even after the holder is compensated for its loss, the data will be traded in the market place. Rules governing data markets should be such that datasets are accessible to whoever can generate value from them and, at the same

More research is needed to figure out the enablers of data markets and the regulatory regime under which such markets generate the maximum welfare for the country's citizens. Forced data sharing removes the market incentives necessary for innovation and efficiency in a digital world.

time, the entity generating the datasets must be able to share in that value to compensate it for its cost of generating them. Data markets will differ according to the type of data. the process of generating the data and, of course, in the use of the data. More careful and systematic research is needed to figure out the enablers of such markets and the regulatory regime under which such markets generate the maximum welfare for the country's citizens. Using broad and hurried brushstrokes like mandatory data sharing will curb the innovations necessary to make India an innovative data driven economy.