

UNDERSTANDING NON-PERSONAL DATA SHARING A Principle-First Approach

Understanding Non-Personal Data Sharing

A Principle First Approach



Amrita Nanda and Astha Kapoor July 2021

Acknowledgements

This report was produced as part of Aapti's Data Economy Lab (DEL) initiative. In addition to contributions from the wider Aapti team, this report also draws from the expertise of numerous academic and industry experts and practitioners. We are grateful for their input during interviews and rigorous feedback sessions.

Research assistant: Deepa PadmarEditor: Yana BeyReport design: Meher RajpalCover illustration: Ayesha Dias

Aapti is a public research institute that works on the intersection of technology and society. It examines the ways in which people interact and negotiate with technology both offline and online.

This is a work of independent research, funded by **Google India.**

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 India License.

To view a copy of this license, visit <u>http://creativecommons.org/licenses/by-nc-sa/2.5/in/</u>.

Glossary (CONCEPTUAL AND ABBREVIATIONS)

Data	Any representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means. (as in IT Act, PDP Bill).
Personal data	Data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, including any inference drawn from such data for the purpose of profiling (as in the PDP bill).
NPD: Non-personal data	Data that either never related to an identified or identifiable natural person, or data which may have initially been personal data, but was later anonymised through transformation techniques to the extent that individual- specific events are no longer identifiable (as in the NPDR). Data sharing under regulations that combine a voluntary and mandatory approach. This may be defined by purpose, sector, or other thresholds under legislation.
Data principals /users	Persons, both natural and legal, to whom any data relates (as in the PDP Bill).
NPDR: Non-Personal Data Report	Revised Report by the Committee of Experts on Non- Personal Data Governance Framework , 2020

Ecosystem	Ecosystem is a composite term referring to the various stakeholders within the data economy, enabling infrastructures for data sharing, enabling legislation to data sharing, and the network of relationships between each of these - working together to make data sharing a possibility.
Data-driven / data-centric	Refers to business, economic, academic and innovation activities for which data access and usage is central in development, deployment, growth or sustenance.

Defining data sharing approaches

Mandatory	Data sharing mandated by state regulations, as defined under legislation. In implementation, this approach varies in which stakeholders are mandated to share data.
Voluntary	Data sharing, not mandated by regulation, where the other party receives monetary or other direct benefits from the initiative.
Altruistic	Data sharing where the sharing party receives no monetary or other direct benefits from the initiative.

Contents

01	Executive Summary	7
02	Introduction	10
03	Methodology	13
04	SECTION I: Value of Data Sharing	15
05	SECTION II: First Principles and Global Pathways	25
06	SECTION III: India's Status and the NPD Report	36
07	SECTION IV: Recommendations to the Committee	45
08	Conclusion	65
09	Annexe	67
	A : Bibliography	68
	B : Further analysis of policy pathways	73
	C : List of experts interviewed	83

Executive Summary

The purpose of this research has been to understand the landscape of, and consequently present evidence around non-personal data sharing and governance - particularly for India. While numerous countries are taking steps toward data governance and carving pathways toward societal value from data - there is a dearth of literature and research to unpacking these pathways. With non-personal data governance still a novel arena, research around the modes and approaches of such pathways are largely limited to theoretical underpinnings, and the ecosystem awaits more qualitative, empirical assessments of various frameworks. This gap in research has in part defined both the premise and outcome of this report, and it is the authors' hope that such research may be built upon as the various approaches to data sharing make their impact known.

Globally, there is a marked need to rebalance power across data economies. With the exponential rise of big tech, there has been a recognition of the immense public value latent in data and its related computational tools. Consequently, there is a recognition of the access problem, with the majority of valuable datasets locked in privatelyheld silos, and used primarily for market interest. Equitably distributing the value of data as market power itself is also a matter of concern, as a means to enable greater entrepreneurship and data-driven innovation, with the aim of creating a more level playing

field for smaller players. The issue of rebalancing power has framed the discourse around not only top-down policy moves across the globe, but of increasingly prevalent bottomup data governance structures like data stewards. Aapti's Data Economy Lab works primarily to understand, instantiate and create supporting ecosystems for various models of stewardship, untangling the design choices and governance decisions that suit different data types, sectors, and regions. This research, however, is limited to an evaluation of policy-led approaches to governing non-personal data and how such approaches may best solve for the aforementioned power imbalances, responsibly opening up avenues for public value from data.

India's recent Non-Personal Data Governance Framework, an expert committee report, offers recommendations for a governance structure to unlock non-personal data (NPD) for public good. The approach recommended is mandatory data sharing from businesses to the government. Given not only the untested nature of most data governance approaches, but also India's uneven distribution of digitisation and infrastructural capacity, the mandatory lens requires some unpacking. As various other countries have done, it is important to first crystallise what the root of any data governance in India must be, and for what purpose, before codifying by what means. This has informed

the basis of our research - to adopt a principle first approach and while outlining the guiding jurisprudence for NPD governance, simultaneously identify how the ecosystem may be better prepared for eventual legislation. Through our analysis, it is found that the uni-directional mandatory approach presents potential problems for India - disincentives to innovation, vague contours for community data rights, impending regulatory uncertainty and potential harms associated with a premature ecosystem. This report includes a series of recommendations (salient points below) to the expert committee, the core of which is adopting an ecosystem-oriented approach that works to mitigate existing challenges and establish a voluntary data sharing system for the country.

Key recommendations to the NPD expert committee

01	Establish an ecosystem-based voluntary approach
02	Clarify and delineate first principles
03	Map high interest sectors and empower sectoral regulators
04	Invest in building blocks for technical, infrastructural capacity
05	Provide incentives for stakeholders - toward public value
06	Test NPD sharing through sandboxes and pilots - to codify accountability and spur innovation

While this research recommends an approach framed by India's current status and maturity for data sharing, it is clear that increased data sharing efforts are required from the private sector in order to meaningfully challenge the status quo. In this regard, the state's intention toward increasing data sharing is welcomed. However, it is worth considering whether a framework such as the NPDR's may usher us from a status quo of private-held data silos, to similar silos held by the state, with neither ensuring data-driven innovation for public value. The framework specified by the NPDR may be premature; while a staggered approach that doubles down on incentives and enabling ecosystems holds a superior chance of instilling an enduring, trusted and transparent data sharing structure for India. It is pertinent to build a sustainable ecosystem for India and, having done so, return to the drawing table to refine NPD governance based on learnings along the way.



Introduction

In 2019, the Ministry of Electronics and Information Technology (MEITY) set up the Committee of Experts on Non-Personal Data (NPD) Governance Framework to work on a governing structure to regulate all data not covered by the ongoing conversation on personal data protection. The Committee, headed by Kris Gopalakrishnan of Infosys, comprised members from academia, tech policy think tanks and legal experts. The objective of the NPD governance framework was to understand how community and anonymised data can be accessed, if at all, to generate public good and value for the citizenry, to catalyse data-driven innovation and entrepreneurship - and how this may interface with personal data discourses.

Since then, the Committee has put out two draft reports in the public domain, the first, in July 2020 and a revised report in December 2020. They were open to public comment and have received considerable engagement from both the private sector and civil society. The debate around the NPD Report has been significant as it impacts the private sector, civil society organisations, and individual and community data principals in equal measure. The report (NPDR) has a fairly large remit since non-personal data is defined by exclusion - comprising everything that is not personal data. This means it includes all data that does not have any

personally identifiable information. It also introduces new concepts such as community data rights, duty of care and data ownership and more functional suggestions such as data trustees, high-value datasets, data businesses and a new regulator, the NPD Authority (NPDA), who will define these ideas further. The report is well-intentioned and makes a progressive contribution to thinking about data as a resource of public value. However, it leaves considerable room for interpretation and consequently has led to concern and uncertainty.

While the effort of the Committee to reimagine and rebalance power in the data economy is welcome, and the second draft report does include remedies for some of the issues brought into relief by the first, key concerns linger. For instance, the report does not necessarily protect the interests of the data principals, even though it proposes to do so by extending new community data rights and making data trusts available to exercise these rights. Data principals are likely more vulnerable since anonymisation and consent for anonymisation, as suggested in the report, are insufficient protection against privacy infringements. Duty of care, especially of data trustees that are supposed to safeguard the interests of the principal groups, is unclear.

¹ Accessible at <u>https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf</u>

² Accessible at https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

³ <u>https://hasgeek.com/fifthelephant/impact-of-npd/</u>

Another fundamental issue, and the focus of the current research, is that the report mandates data sharing by businesses with the government in public interest, to ensure greater datadrivenness. It recommends that highvalue datasets (HVDs) be created by companies for sharing, based on sectors / issues identified by the government or NPDA, which will then be shared more broadly to enable citizen-centric innovation and distribution of the value of data more democratically among communities. However, the report does not adequately engage with the implications of this mandatory sharing: will the private sector lose incentives to anonymise, will there be loss of privacy for data principals, and will small, Indian data-centric businesses be unduly burdened by compliance? Even as other countries are prioritising data sharing, they have done so based on the tenet of voluntary sharing, anchored in a robust ecosystem that enhances trust and spurs innovation. India's current approach is unique and potentially problematic, and strongly needs reimagining.

Our research aims to proffer an alternative approach to the NPD Report, anchored in building of a stakeholderdriven ecosystem to support voluntary data sharing, in which the government is an equal participant and not solely an extractor of data. The ecosystem approach is sensitive to the needs of the private sector (both small businesses and Big Tech) as well as the requirements of public-interest datadriven innovations. Additionally, the research also points to the necessity of defining the idea of public value more clearly, such that the first principles on which the report is based are clear, driven by consensus, and reflective of stakeholder-wide consultation. It also highlights the need to engage more sectoral perspectives, recognising that the priorities and concerns of the health sector, for instance, would differ from those of the mobility sector. The research also recommends starting with building blocks and taking a phased, evidence-based approach to regulating NPD so that the new concepts can be tested and evaluated before implementation, and the resultant ecosystem can be made harmminimising, inclusive, and streamlined.

The NPDR, despite its intent, is little more than a premature step in a novel governance arena. Thus, this research strives to interject a moment of reflection in the NPD conversation in India, to ensure that the policy discourse has an opportunity to refer to comprehensive, alternative ideas born of detailed consultation. India's NPDR requires a great deal of nuance, consideration, and revision before implementation of an NPD sharing policy; this work may serve as a vital launch pad.

03

Methodology

This research is a vanguard exploration of the implications of and alternatives to mandatory data sharing as proposed in the NPDR. Given this context, it is worth noting that the question of non-personal data sharing is globally under-researched, and while there is significant theoretical writing both in academia and in policy documents, evidence-based literature is scant. The landscape is determined by ideological leanings on the issue of data sharing, and not empirics. Therefore, this research faced the complex task of wading through polarised views on the subject, and ensuring presentation of a balanced, well-referenced analysis on non-personal data sharing.

To address the issue, the research follows a mixed methods approach that combines detailed secondary research and analysis with qualitative expert interviews. This allows for corroboration of insights derived from the literature and policy review, and anchors the analysis in the experiences of stakeholders.

Crucially, in addition to the extensive writing and video content from India, especially the publicly available comments to the Committee on both its draft reports and the accompanying writing from its members on the conceptual framework, the research has benefited from academic writing from global universities and research by major tech companies. It also refers to legislative and policy analysis from around 15 jurisdictions spread over the European Union, Australia, Japan, and so on. While the literature review strives to be representative, voices from the Global South are not adequately reflected given the absence of relatable policy and academic discourse on the subject. A detailed bibliography is available in Annexe A.

In addition, the research relied on expert interviews in academia and industry, both Big Tech and smaller start-ups that are impacted by the report's proposals. The interviews were designed to be semi-structured and addressed issues of perceived impact from the perspective of first principles of public good and anti-trust and through the lens of compliance. The questions were also aimed at understanding alternatives to mandatory data sharing, and how India could build up to suggested alternatives. Given the nature of the research, i.e., the conversation is ongoing, the interviews are confidential and no attributions are made to any individuals though their views are reflected in the analysis and insights. A list of interviewees is available in Annexe C.

For feedback, Aapti Institute presented a draft of this paper to key stakeholders in a closed-door roundtable and during one-on-one conversations. The roundtable interaction was confidential to ensure attendees were able to comment without reserve.

04

SECTION I: Value of NPD and Data Sharing

Over the past two decades, data has often been referred to as the most valuable resource on Earth. While data has always been valuable, particularly for businesses - enhancements in computational and analytic tools have magnified its value.⁴ This is most obviously apparent in the massive boom in reach, revenue and growth of tech platform companies in recent years. However, the unique nature of data makes it difficult to quantify this value. Unlike oil or gold (both popular but flawed analogies used to describe data's value), data is nonrivalrous – its value does not diminish with use. On the contrary, datasets can be reused multiple times and in various combinations to generate further value. While a natural resource like oil exists in finite amounts, data is seemingly unlimited. There is no demographic or region that does not harbour potential for valuable data mining – for example, a person living in India can provide the same amount of genomic data as someone living in Ghana, the US, or elsewhere. What makes the valuation of data tricky, however, is that its worth is defined by resolution, usage and application. Raw data bunkered into servers, no matter how vast or diverse, would be useless without the means to read into it and

develop insights or algorithms from it. This means that data becomes more or less valuable, depending on whose hands it is in. Algorithms developed to produce analytics also learn and refine themselves as more data is fed in – thus, greater variety, velocity and volume enhances data's value in application. For instance, one of the most prominent value extractions from consumer data is towards targeted advertising.⁵ This is premised on an extensive understanding of consumer behaviour – both past and predictive facilitated by Artificial Intelligence (AI). The more social data is collected through platform traffic the more accurate these predictions become - and, thus, the more value a single advertisement accrues on a given platform. This is one of the many ways in which the private sector has swiftly capitalised on the monetary value that data can generate.

⁴ Enders, Tobias. (2018). Exploring the Value of Data – a Research Agenda. Accessible at <u>https://www.researchgate.net/publication/327246406</u> Exploring_the_Value_of_Data_-_a_Research_Agenda

⁵ Bergemann, Dirk et al (2019), The Economics of Social Data (September 25, 2019). Cowles Foundation Discussion Paper No. 2203R. Accessible at <u>https://ssrn.com/abstract=3459796</u>

The status quo – who retains most of this value?

With the realisation and application of data as not only an immense market power but also a tool of knowledge and insight, there is growing global discourse around who retains the majority of this power. This conversation is driven partly by the fact that large tech companies now hold a seemingly unfair business advantage due to the massive amounts of data they have collected and continue to collect. The nature of data makes it harder for newcomers to enter the market of data-driven businesses⁶, compared to most traditional industries. Related competition concerns have led governments to question major tech companies, and have constituted a paradigm shift in policymaking⁷, toward rebalancing competition in data-driven industries. A second facet of this discourse has been the agency and rights of the individual user and of communities – the primary generators of the data that now drives a massive digital economy. While data principals receive useful services in return for their data, there is a marked power imbalance⁸ not only in the distribution

of value between people and companies, but also in the relationships individuals and communities share with tech platforms. Consent remains a one-way street, a barrier to accessing a service, with limited understanding or visibility afforded to consumers on their data collection and usage. When this information is made available to end-users, it is severely diluted by the fact that most people do not have an adequate understanding of what these terms, data and their use-purpose mean, or the likely consequences. Platform workers similarly have little visibility of the algorithms and decisions that govern their employment, pay, increments or hours.⁹ This power dynamic is further skewed by the fact that the value returned to communities is largely limited to online services they receive, while the immense potential of data to inform solutions for societal good remains locked. The majority of valuable data is siloed with private companies, inaccessible for wider, public-first and community oriented use or governance.

⁶ A test and related findings for measuring data-drivenness and barriers to market entry for the same can be found at <u>https://prufer.net/2021/02/04/</u> mandatory-data-sharing-development-of-a-test-governance-structure/#_ftn1.

Related paper by Prufer, Jens and Graef, Inge (2021), Governance of Data Sharing: a Law & Economics Proposal. TILEC Discussion Paper No. 2021-001, CentER Discussion Paper No. 2021-004. Accessible at SSRN: <u>https://ssrn.com/abstract=3774912</u>

⁷ McDonald, Sean. 2019a. "How Regulations Are Reshaping Digital Companies." Cigionline, April 15. <u>www.cigionline.org/articles/how-regulations-are-</u> reshaping-digital-companies

⁸ Nanda, Amrita (2020), Power Structures in the Data Economy, Aapti Institute. <u>https://thedataeconomylab.com/2020/10/30/power-structures-in-the-data-economy/</u>

⁹ Kapoor, Astha (2021), Collective bargaining on digital platforms and data stewardship, Friedrich Ebert Stiftung, Asia. <u>https://asia.fes.de/news/fow-collective-bargaining</u>

How can we think about public value in the context of data?

While the financial returns from datadrivenness have been recognised, there has also been a shift in how value can be conceived of in the context of data. Specifically, how can data bring community value beyond monetary output - to enhance the online and offline lived experiences of individuals and communities? Numerous open data initiatives and new frameworks for data governance like data stewardship have begun to create a vision for a new data economy - one that puts communities first and affords people greater agency in their data governance. This value is most tangibly seen in public good solutions such as safer cities, citizen-driven scientific research, and predictive solutions for natural disasters, to name a few. Data is also increasingly being used to attempt to rebalance existing inequities in society to drive inclusive medical research that focalises demographics previously left out of primary research, or to rebalance agency in the hands of platform gig workers.¹⁰ Some of these initiatives have been led by governments, while many are the result of collective community activity¹¹, enabled by legislation like the

EU General Data Protection Regulation (GDPR) which facilitates significant agential mechanisms to individuals regarding their data (such as data portability, which has allowed citizens to pool data and shepherd collective decision making on their data).¹²

The move to utilise data as a means to societal value is driven not only by need, but as a means to circularise a presently linear value chain; bringing meaningful value from data to those that generate it. This conversation has also been framed by a notion of commons, with data (particularly non-personal data) being viewed as a community resource, one that communities should ideally have access and agency over.

¹⁰ Driver's Seat presents an example of a data cooperative for gig workers in the ride-sharing industry that works to empower workers with analytic insights into their earnings, and contribute aggregated insights to government agencies toward better-informed policy decisions.

¹¹ Data cooperatives like Salus Coop and MIDATA work to advance citizen agency over their health data, and promote citizen-driven scientific research through collective decision-making.

¹² For more information on Data Stewardship and its instances across the globe, visit <u>https://thedataeconomylab.com/tracking-stewardship/</u>

The value of overcoming barriers to data sharing across societal, academic and economic benefits has been recognised across numerous research metrics, but remains difficult to quantify.^{13 14 15}

Economic

B2G and B2B data sharing and reuse will have significant benefit to the global and national economies.

• ~**USD 3 trillion p.a.** can be unlocked by re-use of public and private sector data globally in up to seven areas of the global economy.

• **Example:** The reuse of Transport for London's open data was generating annual benefits and savings of **~USD 177 million across stakeholders.**

• **Example:** In Japan, data platforms (established to facilitate reuse of private firms' data) added between **~USD 5-13 billion gross value** to the country's economy.

Societal

Data linkages and integration across organisations and sectors can enhance the value of siloed insights and inform policymakers on societal challenges.

• **Example healthcare:** Data linkages have been recognised as crucial to providing efficient and higher quality healthcare. In the UK, linkages have contributed to quicker screening and early recognition of cancer.

• **Example Smart cities:** Smart city initiatives are increasingly prevalent (eg. London, Barcelona), for which data is collected and created by multiple actors. Linkages across players is crucial to integrating different types of applications and enabling synergies.

Academic

In science and research, increased data sharing can reduce cost of conducting research, streamline collaborations across researchers and disciplines, and increase scientific quality.

• **Data scrutiny** is crucial to the scientific process - to verify and validate research thereby enhancing quality. This is best facilitated through data sharing and transparency.

• **Example:** data repositories allow quick access to research data and facilitate interdisciplinary collaborations.

Increased sharing of research data will overall reduce instances of false results, and build trust in the sciences.

¹³ <u>https://thedataeconomylab.com/2020/07/31/data-sharing-for-public-good-theoretical-bases-and-policy-tools/</u>

¹⁴ McKinsey Global Institute (2013), Open data: Unlocking innovation and performance with liquid information. | Deloitte (2017), Assessing the value of TfL's open data and digital partnerships | Cities and Data Sharing, Aapti Institute | The Open Knowledge Foundation: Open Data Means Better Science, Jennifer Molloy

¹⁵ Enhancing Access to and Sharing of Data : Reconciling Risks and Benefits for Data Re-use across Societies, OECD. Aapti analysis

How can data sharing spur this value?

A key component of this paradigm shift and the vision of a new and more equitable data economy is data sharing. Existing datasets, particularly within the private sector, hold a number of the aforementioned value metrics for data, i.e., variety, velocity and volume. While there are certainly regions or communities that remain largely offline, opening up existing data can help gear the resultant value toward to data principals. There are some data sharing partnerships that have already demonstrated this value - with the private sector chipping in to innovate as well. Open data initiatives in cities like Rennes, Barcelona and more¹⁸ have also shown data sharing and access to be a powerful foundation for greater innovation toward societal good. For example, sharing datasets on commuter mobility patterns in a city could invite ride-sharing parties to better plan the deployment of their vehicles. In turn, this information may enable cities to plan public transportation better as well - increasing the distribution of commuters between public and private transport, reducing the number of overall vehicles on the road and hence vehicular pollution - which presents

only one application of mobility data sharing. Mobility patterns can help cities determine more useful security measures, mark high-traffic and lowvisibility areas to employ better lighting, and so on. Analyses of a number of data-driven initiatives in cities have shown that multi-stakeholder involvement is another key factor in diversifying and amplifying innovation for societal good (See figure 4.1). While fields of academia or research may benefit purely from sharing (through increasingly verifiable results and interdisciplinary learnings), public good solutions (for eg; smart cities or predictive natural disaster solutions) require a coming together of different abilities and interests.¹⁹

Thus, while data sharing is an important step toward rebalancing the data economy and affording value to individuals and communities, it is pertinent to also create a focus around innovation – and how value is to be accrued from sharing. Without clarity on the mechanisms for value creation, it is difficult to envision how sharing alone may unlock broader societal value.

¹⁸ https://www.aapti.in/blog/the-future-of-cities-data-sharing-stewardship

¹⁹ Josep-Ramon Ferrer, Barcelona's Smart City vision: an opportunity for transformation, Field Actions Science Reports [Online], Special Issue 16 | 2017, Online since 01 June 2017, connection on 20 July 2021. Accessible at http://journals.openedition.org/factsreports/4367

Comparing two data sharing initiatives in Bengaluru & Mysore, India and Metropolis of Rennes, France for their involvement of stakeholders, focus on innovation, and resultant benefits.²⁰ ²¹ ²²





²⁰ Big Data and Urban Transportation in India: A Bengaluru Bus Corporation Case Study (2018), Manchester Centre for Development Informatics. Rakesh, Heeks, et al.

²¹ Open Transport Data Assessment in Mysore. Daniel Rudmark, World Bank Group.

²² Data Insights, Open Data Soft. (Aapti analysis)

What are some of the existing challenges to increased data sharing?

Unlocking data and its consequent value is a challenging task. Data governance remains a novel challenge for most countries, with the fundamentals of data protection and privacy laws still evolving in most regions. In the case of sensitive or personally identifiable data, risks to individuals' right to privacy may be exponentially magnified with data sharing. To mitigate these risks, airtight regulation around re-use, sharing and anonymisation is pertinent. The case of non-personal data is different - it may be characterised as data that was never identifiable (generated from machines or IoT devices, and pertaining to non-human subjects like geospatial data, weather or environmental data, and the like) or as personal data that has been aggregated and scrubbed of potentially identifiable material (through anonymisation and pseudonymisation mechanisms). While non-personal data presents a slightly different, and perhaps more viable frontier for data sharing, data protection and clarity around community data rights must create the basis for NPD sharing as well. Typically, it is through personal data legislation that the contours of NPD - definitions, recourse mechanisms and purposes for sharing - can be most clearly understood. For example, aggregated data about

an online community's activities may not necessarily be traceable to an individual, but can provide often invasive insights into the activities and behaviour of others that may be categorised within said community. This is especially challenging to govern since defining data communities presents numerous overlaps and intersecting demographics, and there may be a seemingly infinite list of data communities that each digital presence or individual fits into. Further, NPD often refers to what may be understood as community resources or data on natural resources such as agricultural or environmental data. It is important here to consider the lens of the commons: it is knowledge that a community may have rights over in the same way that a community can be afforded rights over shared resources such as a lake or public park. While the indigenous data sovereignty movement has exemplified some of this thinking, it is challenging to determine the same for communities that may not classify as indigenous, but may still be theoretically entitled to govern their resources. To outline these rights and the boundaries inherent to them is an important step in protecting communities globally and, more important, in using data to uplift and empower previously

marginalised sections. Moreover, this is a step that must also be constitutionally determined - this requires countries to unpack and shape how groups may have access, ownership or agency over their data.

Beyond risks to individuals and communities, data sharing is also hindered by the immense enabling infrastructure and coherence across sectors that is required.²³ Different regions and sectors differ on how pronounced this challenge is, depending on their existing levels of digitisation, standardisation in data formats, and networks and legislation for safe and delineated sharing. Given the swift rise of data as an industry in itself, and its application across almost every business function, data formats are varied across sectors. This makes it hard for previously unrelated datasets from different actors to speak to one another, or to be combined in usable formats. In the case of AI development, for instance, large amounts of data that is stored in unreadable or otherwise unwieldy formats may be virtually useless as compared to a less varied, smaller dataset that is usefully categorised with clear metadata. Thus, part of the enabling infrastructure required for data sharing is regulation around data formats, harmonisation with existing and evolving standards across industries, and clear policy

pathways to map data sharing agreements to. The following sections in this paper will look more closely at what this enabling infrastructure must look like and why this is necessary – and how policy can pave the way for enabling ecosystems with minimal friction for all stakeholders involved.

Another systemic challenge to data sharing has been a lack of trust and incentives across stakeholders.²⁴ There are certainly perceived risks from regulation that prevent the private sector from moving into robust data sharing, while state actors may not see clear public value in sharing with the private sector. Within the private sector as well, data is a primary market power, and the backbone of massive revenue the competitive disadvantage that could result from data sharing, particularly when competitors are not sharing, acts as a disincentive for sharing. There is a clear need to reimagine incentives in a data-driven context, and foster ecosystems that bring stakeholders together for their unique valueadds towards a more equitable data economy. These exigencies have been recognised by countries across the globe, and we are now seeing steady, varied and unique policy pathways that speak to these challenges. Our next section analyses a number of

these emerging policy make-ups from a principle-first viewpoint to better

²³ OECD (2019), "Risks and challenges of data access and sharing", in Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, OECD Publishing, Paris. <u>https://doi.org/10.1787/15c62f9c-en</u>.

²⁴ Martens, B., Duch-Brown, N., The economics of Business-to-Government data sharing, European Commission, Seville, 2020. JRC119947.

understand how India can frame its data sharing future – what are the motivations driving policy around data sharing, and how do different approaches facilitate various goals and challenges?

05

SECTION II : First Principles and Global Pathways

First principles – an overview of the motivations behind global policy moves toward data sharing

As countries across the globe begin to tackle data governance and are at various points on their journeys toward increased data sharing, top-down approaches can be understood both through enacted legislation and softer policy directives, recommendations or strategic outlines. Some of the more tangible policy pathways are emerging from the European Union, Australia, Finland, Japan, Estonia, Israel, Germany and Barcelona (Spain). While each of these is at a different stage in data governance (See figure 5.1), an analysis of their approaches provides valuable insight into their overarching bases, first principles, problem statements, and roadmaps for implementation.

Global regulations and legal documents analysed in this research are at different stages of implementation.

REGION	REGULATION / DOCUMENT	STATUS
European Union	 Digital Markets Act Digital Services Act Data Governance Act Regulation on Free Flow of Data Open Data Directive Sharing private Sector Data in the European Data Economy European Strategy on Data 	 Legislative Proposal Legislative Proposal Legislative Proposal In Force In Force Expert Group Report / Framework Expert Group Report / Framework
Australia	 Data Availability and Transparency Bill Competition and Consumer Amendment (Motor Vehicle Service and Repair Information Sharing Scheme) Bill 2020 Data Exchange Framework 	 Legislative Proposal Legislative Proposal Expert Group Report / Framework
Finland	 The Act on Transport Services Law on Secondary Use of Health and Social Data Finnish Forest Act 	• In Force • In Force • In Force
Japan	 Act on Special Measures for Productivity Improvement and Amendment to Industrial Competitiveness Enhancement Act Enacted Basic Act on the Advancement of Public and Private Sector Data Utilization 	• In Force • In Force

REGION	REGULATION / DOCUMENT	STATUS	
Estonia	 Public Information Act (Information systems data exchange layer) State Information System Management System 	• In Force • In Force	
Israel	Patient Rights Regulations (Health Data Research use), 2019 Credit Data Law In Force		
Germany	 Patient Data Protection Act German Act against Restraints of Competition "Digitalization Act" (10th Amendment) 	• In Force • In Force	
 Barcelona Legal Regime of the Public Sector Law (Ley 40/2015) Transition Towards Technological Sovereignty city government measure (2016) 		• In Force • In Force	
India • Report by the Expert Committee on Non-Personal Data Governance Framework		• Expert Group Report / Framework	

Figure 5.1

While each region, sector and nationstate has unique challenges regarding data governance, and is at different levels of digitisation and technical capability, the principle bases for a move toward data sharing are markedly similar in their categorisation and weightage. The countries and policy documents listed above all locate their primary motivations around at least one of the following topics:

1. Research and Innovation

Referring primarily to pro-reusability policy and access to training data for

the purposes of innovation, advancing scientific research and spearheading data science itself. Each document analysed makes mention of this as a goal, either as an end in itself or often as a means to enable goals 2 through 5.

2. Competition and Antitrust

Referring to pro-competition policy with a stated purpose of fostering fair business environments and conducive entry points for startups. This is seen outlined in the Digital Markets Act (EU), which establishes a set of narrowly defined objective criteria for qualifying a large online platform as a so-called "gatekeeper".²⁵ This allows the DMA to remain well targeted to the problem that it aims to tackle as regards large, systemic online platforms.

3. Public Good and Citizen Welfare

While the stated purpose of most tie in to citizen welfare (through greater innovation across stakeholders), many also outline transparency, democratic agency over systemic platforms, and new innovation or opportunities in health and social welfare.

4. Sectoral Efficiency

Regulations that target sector-specific contexts are often led by a mapping or selection of which sector may benefit most, and may be best prepared for increased data sharing. For example, Finland's Act on Transport Services²⁶ instils mandatory data sharing from all transport service providers to open up essential data. This includes such information on routes, stops, timetables, prices, availability and accessibility in a machine-readable form via open interfaces. By sharing data, service providers can use their transportation fleet more effectively in moving goods and passengers.

5. Sustainability

As with aforementioned goals, sustainability finds mention both directly and indirectly, through moves toward reduced carbon emissions, smarter waste and pollution management in cities and streamlined traffic management. Sectoral mandates also reflect that many countries are prioritising high-emission industries like transport services to streamline data sharing.

As seen in Figure 5.2 (Venn diagram)²⁷, research and innovation form the basis of most data sharing policy, and notably constitute a metric that lends itself well to other popular first principles such as public good, competition concerns, and sectoral efficiency. In each of these cases, first principles have been the initial leg of policy, and in most countries (such as those of the EU, Australia, Finland) have been delineated thoroughly before the regulatory stage. For example, the European strategy for data²⁸ defines a focus on empowering individuals, investing in skills and SMEs, building data spaces in strategic sector and domains of public interest, and outlines an experimentation-based, agile framework based on 'iteration and differentiation'.

²⁵ REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on contestable and fair markets in the digital sector (Digital Markets Act), 2020. Accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en

²⁶ Act on Transport Services (2018), Ministry of Transport and Communications, Finland. Accessible at <u>https://www.finlex.fi/fi/laki/kaannokset/2017/</u>en20170320_20180731.pdf

²⁷ Aapti analysis of global data sharing legislation and frameworks; complete list in Annexe

²⁸ A European strategy for data (2020), European Commission. Accessible at <u>https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy#documents</u>

Global regulations and legal documents analysed in this research are at different stages of implementation.



Figure 5.2

The focus on first principles can be attributed to the fact that at this stage, all data governance policy is novel and requires an enduring jurisprudence upon which future policy refinements or additions can rely. Further, the staggered timelines and disparity in status across countries is reflective of the iterative approach to data governance - to test well and to tread lightly in building out these frameworks.

Understanding the application of mandatory and voluntary approaches to data governance

In considering top-down data governance, there are two broad schools of thought that may guide data sharing policy – mandatory and voluntary. Mandatory approaches employ legislation to necessitate data sharing by stakeholders, while voluntary approaches create outlines for optional data sharing. However, the mandatory versus voluntary binary exists largely at a theoretical level, and in implementation each school is proving to be rather varied. For example, Australia may be categorised as taking a mandatory approach but this does not exist in a purist form as it is currently limited to the automotive sector, with outlines for purpose and data types. Similarly, while the EU's Data Governance Act²⁹ adopts a voluntary basis (termed 'altruistic') for data sharing by facilitating an enabling ecosystem, the Digital Markets Act includes mandates for certain stakeholders above a specified threshold of data-driven economic activity (termed 'gatekeepers'). Thus, the real-world applications we are seeing are combined versions, tailored often by sector, purpose or stakeholder. (See figure 5.3) Since this is an evolving policy landscape, this taxonomy is

not exhaustive but aims to highlight a majority in the implementation structures of data sharing approaches.

A free-form voluntary approach

describes the status quo in most regions, where regulation has not yet come up to ensure data sharing, and there are likely minimal structures for data protection and privacy as well. In these cases, data sharing partnerships occur on an ad hoc basis, with sharing agreements being negotiated by participating entities, and little or no targeted legislation in existence.

Ecosystem-enabled approaches also come under the broader umbrella of voluntary data sharing as the focus of such policy is on mitigating existing challenges in data sharing and investing in infrastructural, legal and technical building blocks that foster a trusted network for voluntary sharing.

Sector-specific approaches are typically mandatory, with legislation identifying key sectors for data sharing, based on maturity, potential or a need identified within first principles. Such mandates may apply to all stakeholders in a sector, or exclude via a threshold, or corral by data type.

²⁹ Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act), 2020. Accessible at <u>https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767</u>

Free - Form Voluntary	 No regulatory intervention Companies and government share/innovate on ad hoc voluntary bases 	 Voluntary sharing for public good Strategic partnerships for business interest Transparency for trust - building 	A pre-regulatory stage (that many countries are in) with few, varied and often unsustainable structures.
Ecosystem - Enabled Voluntary	 Regulation dissolves existing barriers to sharing across stakeholders Companies enabled to share/innovate voluntarily, based on individual capacity 	 Promote voluntary sharing & innovation through a trusted ecosystem Reduce regulatory & capacity burdens on sharing parties 	The EU's Data Governance Act outlines ecosystem enablers such as standards, intermediaries to reduce cost and risk burden of sharing.
Sector - Specific Mandate	 Regulation creates mandates on sharing data within a certain domain All or most stakeholders in the sector made to share data, unless excluded by threshold 	 Spur innovation and research in a particularly feasible and/or pertinent domain Rebalancing competition and access to data in the sector 	Finland's Transport Act consolidates all forms of transport (rail, air, maritime, etc.) & provides for mandatory release of certain information.
Purpose - Specific Mandate	 Regulation mandates sharing or data-related efforts from private sector, to address a specific policy goal/ solution. Companies participate based on time-bound requests from state 	 Targeted sharing to inform policy goals for public value Spur innovation for a particular region or demographic. 	The French Act for a Digital Republic allows the government to request commercial players to provide access to data they hold – to establish public statistics.
Blanket B2G Mandate	 Proposed framework to mandate sharing of metadata from all "data-driven businesses" to the government. Adjudication on purpose assigned to the proposed Non Personal Data Authority. 	 To create public value through increased data sharing for research, innovation & policy. To bolster start ups and small firms by providing access to an otherwise siloed resource – data. 	India's Expert Committee Report on a Non Personal Data Governance Framework for the country.

Recommended

Figure 5.3

Purpose-specific mandates are

varied, defined by legislation-specified goals and are often temporally bound partnerships between sectors or industries, with the aim of meeting public policy goals. These mandates may be used as tests to define future partnerships and regulations.

Each of these structures need not exist in purity - they are often combined with one another as data governance in most regions continues to evolve. As seen in the EU's stance, for example, the Data Governance Act adopts an ecosystem enabled voluntary approach, promoting 'altruistic' data sharing while building an ecosystem of sharing infrastructure, intermediaries and purpose - while the Digital Markets Act creates a mandate on data sharing, defined by stakeholder threshold. Thus, there is no 'one size fits all' approach to data governance and it is important to note that most countries are adopting a staggered or 'layering' policy approach. Beginning with theoretical bases, testing through limited application, and refining as they go forward.

In the case of India's NPDR, the recommended approach appears to be an anomaly in that it adopts a blanket mandatory approach across sectors and players, while also creating a unidirectional framework where only the private sector is mandated to share data. As seen above, even where mandatory approaches have been deployed, they are limited by purpose, sector or stakeholder thresholds - this also serves to stagger data sharing, and tailors mandates to maturity. Mandatory systems may work well for countries such as Finland, whose digitisation journey to legislations on re-use of data has now mapped about two decades³⁰ and has worked to diluted key ecosystem challenges of sharing. However, the ecosystemenabled approach, a version of which we elucidate in Section IV, is a virtually essential step for countries that have not yet been able to mitigate existing issues between stakeholders, and constrains of technical capacity therein.

³⁰ See figure 6.3 in Section III for more.

Challenges and conditions associated with mandatory and voluntary approaches

The two primary approaches to topdown data sharing (mandatory and voluntary) are polarising subjects and views differ primarily from the purpose-lens. Depending on what one is hoping to solve for, different ends of the spectrum or may not work well. In the case of mandatory sharing, many have expressed this as a meaningful way to dilute the power and data silos of big tech companies, and rebalance competition in data-centric spaces. The theoretical basis for this thinking is framed by the fact that data presents the primary market power for such companies, and by mandating sharing, some part of this power will be distributed. This is typically argued as a means to prevent market tipping, and protect environments of innovation. However, it is apparent that most who lean on the mandatory end of this argument also hold that sharing must be equitably deployed for governments as well - particularly to guard against a state monopoly over data. Mandated data sharing would require a governance structure that brings together elements of "economically efficient centralization with legally

necessary decentralization"³¹. This diversity in governance is necessary in order to sustain community welfare and agency. However, mandatory sharing remains largely untested for its impact on markets or on the goal of public value, and most research lives in a primarily hypothetical space. The polarisation on the matter of mandatory sharing can also be attributed in part to this empirical constraint. For the goal of consumer welfare (not necessarily distinct from competition concerns), mandatory data sharing has also been studied as a remedial measure for infringements on consumer welfare. In a remedial context, sharing would have to be tailored to restore competition to the level that had existed at the time the infringement began, instead of the level that would have existed but for the infringement.³² Analysis has shown that remedial mandatory data sharing would in fact result in the latter outcome, and this would be damaging to consumer welfare.

On the other end of the spectrum, voluntary data sharing is better tested, owing largely to the existing prevalence

 ³¹ Prufer, Jens and Graef, Inge (2021), Governance of Data Sharing: a Law & Economics Proposal. TILEC Discussion Paper No. 2021-001, CentER Discussion Paper No. 2021-004. Accessible at SSRN: https://ssrn.com/abstract=3774912. Aapti analysis
 ³² Vikas Kathuria, Jure Globocnik (2020). Exclusionary conduct in data-driven markets: limitations of data sharing remedy, Journal of Antitrust Enforcement, Volume 8, Issue 3, November 2020, Pages 511–534. https://doi.org/10.1093/jaenfo/jnz036. Aapti analysis

of sharing initiatives. While such structures (particularly in the open data movement and in scientific research) have shown considerable achievements toward societal good - they are composed of largely private-led or privately-mediated sharing. This status quo presents a number of concerns the primary being inadequacy of such efforts. Businesses are not inherently structured to negotiate for public good, or partake in public good initiatives in the absence of business incentives. Such a drawback is layered by the fact that private-public data sharing agreements, particularly in India, have been ad-hoc and constitute a significant burden on the sharing party - to negotiate sharing agreements or to develop safe and usable pathways for sharing - in the absence of enabling legislation, standards, or ecosystems. This leads to an inevitable cycle of reinventing the wheel for such partnerships, and creates a marked disincentive to sharing for public good - despite the considerable repetitional advantage that private companies would gain from such initiatives.

On either end, there exists limited but equitable literature to speak to the benefits or downfalls of mandatory and voluntary sharing. For the purpose of this research, and given the proposed NPDR framework, an analysis against first principles and challenges of data sharing was found to be most necessary (See figure 5.4). As mentioned (albeit sparsely articulated) in the NPDR, public good is the primary principle for India's NPD sharing recommendation. In this regard, mandatory sharing without adequate mechanisms to decentralise and ensure equitable distribution does not fare well to solve for public value. This is partly due to the lack of clear pathways to public good from mandatory sharing. Further, the Indian ecosystem faces a problem of maturity - the NPDR has been criticised as overlooking the costs of re-engineering systems for its prescribed data sharing, particularly for startups that lack adequate financial and human resources.³³ To inhibit startups and innovation also inhibits consumer welfare - and may contradict the goal of research and innovation for public good. Industrywide consultation has also reflected that the NPDR framework will act as a disincentive for smaller players to cross over into the threshold of 'data-driven businesses,' thus further limiting the playing field, rather than evening it. Mandatory sharing, where it has been implemented, finds action in limited capacity - through a sectoral or purposeled approach. The NPDR, however, contradicts these evolving standards by creating a sweeping mandate across industries, with little scope for limiting or testing capacity and maturity. Our next section looks at how India is currently placed to implement data sharing - and whether it can withstand the NPDR's mandatory framework.

³³ Impact of Non-personal Data Report (2020), Fifth Elephant, Hasgeek. Accessible at https://hasgeek.com/fifthelephant/impact-of-npd/

An overview of data sharing challenges mapped across the various voluntary and mandatory governance structures.

Recommended					
	Free-form Voluntary	Ecosystem- Enabled Voluntary	Sector - Specific Mandate	Purpose - Specific Mandate	Blanket B2G Mandate (India's NPDR)
Are there clear economic incentives?	Economic benefits may serve as incentives, but difficult to discover	Ecosystem can create opportunities for economic incentives	Sectoral efficiency serves as an economic incentive for stakeholders	Economic incentives complex to establish	Broad mandates are an economic disincentive due to IP and competition concerns.
Are there protections for user privacy?	May compromise user privacy due to opaque sharing agreements	Data privacy and user protection forms the basis of the ecosystem	Mandate focused sharing may make demands of user data without adequate protections		Considers anonymisation the only risk mitigation for user privacy; primary data protection law absent.
ls data sharing infrastructure available?	Voluntary agreements create infrastructure for limited exchange	Ecosystems is made of reliable data sharing infrastructures	Sector and purpose specific mandates may create need-based infrastructure for sharing		Proposed & existing data sharing infrastructure is insufficient, creating burden on sharing parties
Are there standards for sharing?	Ad-hoc agreements narrow the scope of standardisation	Standards are core to ecosystem building	Sectoral standards evolved to ease sharing	Standards likely to be instance-spe cific, may percolate over time	Standards / roadmap for interoperability remains undefined.
Are there processes to execute safe sharing?	Instance - specific contracts likely to define processes	Processes for sharing fundamental for functioning ecosystem	Sectoral processes evolved to ease sharing	Processes likely to be instance and data - type specific	Processes for sharing unestablished, this burden largely rests upon businesses

Likely

Somewhat Likely

Unlikely



SECTION III : India's Status and the NPD Report
As mentioned above, the NPDR is focused on the value of data, and what it means for public good purpose. Touching broadly upon what public good purpose can encompass, the report lists improving public service delivery and achieving broad socioeconomic policy goals in data-driven sectors such as health, agriculture, and mobility. It also includes creating an environment of innovation, especially for start-ups. The social and economic benefits of NPD sharing are aimed at citizen communities and domestic business.

While the focus on public good purpose is important, it also warrants a clearer

understanding of what constitutes public good, especially as businesses will be asked to mandatorily share data pertaining to "public" purpose and, in the context of high-value datasets, value will be defined based on the contribution of that data for achieving this public good. However, data is only valuable, for economic and public purposes, if there is a broader infrastructure, technological and human, that is able to use it, and derive value from it. This section, therefore, reads the NPDR in the broader context of India's need and ability to harness data value which is currently fairly rudimentary and evolving.

Diversity in India's sectoral maturity for data sharing

Data sharing is likely to unfold differently in different sectors: even the NPDR recognises that the diversity in capacity and ability to absorb the recommendations of the Committee hinge on the size of the sector as well as its levels of digitisation (figure 6.1). The NPDR recommends health and mobility as sectors where NPD governance can make a significant impact.

Sectoral mapping for data sharing may be carried out across metrics like digitisation and impact



Figure 6.1

Healthcare clearly lends itself to unlocking the value of data for societal / public impact. For instance, aggregated electronic health records (EHRs) can help understand diseases better, and streamline and improve accessibility through telemedicine especially in rural areas; increased insights from patient data can also lead to quicker diagnosis and make healthcare more efficient and affordable; and evidence-based management of disease improves the quality of care. The benefits of digitisation in healthcare accrue across stakeholders: for patients, making personalised, efficient and affordable care possible; for practitioners, especially in a country like India, where healthcare is overwhelmed, reducing time spent on treatment; for the private sector, opening further avenues for innovation and creating new markets such as for wearable technologies; and finally, for the government, leading to better resource planning and decisionmaking. Despite this clear and obvious impact, India's health landscape is far from ready to unlock the value of data. There are fundamental gaps in the health sector, such as low availability of doctors, poor insurance coverage (70% of all health expenses are out-ofpocket³⁴) and the Government of India's overall investment in health is limited. These issues compound the problems

of digitisation; for instance, only 20% of all immunisation records are accessible digitally³⁵ and information sharing systems between hospitals are limited.³⁶ It is well known that healthcare data, even when digitised, exists in multiple forms and may not be entirely operable. Collation of data is another challenge, as data collected from various sources may exist in siloes and cannot be pooled to derive insights. The government has proposed a National Digital Health Mission (NDHM) to address issues around digitisation and data sharing but the policy remains mired in the draft stage, awaiting deliberations. It is also likely to come into conflict with the Committee recommendations, and will need further consideration prior to any implementation at scale. So, while enormous potential exists for impact and there is significant public value at stake, health data sharing policy, infrastructure and capacity do not offer adequate support.

The case of mobility is more robust, there is vast business potential and public opportunity in the sector where just revenue from ride hailing services is likely to be \$43 billion³⁷ by 2025. The government has invested significant resources in Smart Cities (~\$800 million³⁸) and it is a highly data-driven sector with tech-enabled platforms

³⁶ ibid

³⁴ <u>http://www.mohfw.nic.in/sites/default/files/38300411751489562625.pdf</u>

³⁵ <u>https://isbinsight.isb.edu/digitising-indian-healthcare-records/</u>

³⁷ https://www.business-standard.com/article/companies/india-s-new-mobility-market-expected-to-touch-90-bn-by-2030-data-119101500156_1.html ³⁸ https://www.ibef.org/blogs/revitalising-india-s-smart-city-mission#:~:text=Budget%20Allocation%20and%20Recent%20Initiatives&text=Out%20 of%20the%20total%2C%20the

like Uber and Ola providing mobility solutions across urban India. There is also greater data sharing in the sector – for instance, Uber's Movement³⁹ platform provides aggregated citylevel data and tools to cities to better understand urban transportation and address issues of road safety, congestion, emissions, and so on. The Government of India, recognising the value of mobility data in cities, has established the Indian Urban Data Exchange (IUDX) to facilitate secure and authenticated exchange of data between data producers (such as city governments) and data consumers (start-ups, researchers).

This investment in both private and public infrastructure to collect, collate, and share data is integral to establishing a culture and environment in which the value of NPD is recognised and there is a supporting architecture that enables unlocking.

³⁹ https://movement.uber.com/?lang=hi-IN

What is needed to prime the Indian ecosystem for data sharing?

While India has a number of data sharing initiatives, some governmentled, some public-private partnerships and a few open data initiatives - these remain disjointed from one another. There is a need to coalesce learnings from each of these, double down on initiatives in mature sectors for data sharing, and create new pilots. Initiatives like Open Data Telangana, or the Hyderabad Open Transit movement, or data sharing and digitisation as the aforementioned BMTC initiatives are still nascent, and hardly comparable to global best practices or roadmaps for implementation. In our comparative

study of BMTCs data sharing initiatives in Bengaluru and Mysore with that of Metropolis of Rennes, France - key factors constrained the BMTC project from creating meaningful value (See figure 6.2). Such initiatives, and their levels of maturity indicate that India must first develop fresh and refine existing pilots to identify what structures may streamline sharing in various sectors, and for various stakeholders. Mandatory sharing without key infrastructural reforms may result in chaotic, and value-inhibiting sharing.

A closer look at the comparative study of Bengaluru and Rennes' data sharing initiatives in mobility

Case	Description	Implementation	Analysis
Bengaluru & Mysore, India	 Transport type: <i>Public buses</i> under BMTC government agency Main actors: <i>Public & Private</i> actors BMTC, Trimax, PwC, Mapunity Features: Vehicle tracking, Electronic ticketing and passenger information systems, Mobile application and ITS control rooms. 	 Data used for vehicle, ticketing and bus crew tracking. Commuters receive information on ETA, routes. Control room verifies revenue generation, accurate EOD collection. Management sees reduced discrepancies in 'ticket pilferage,' bus crew workflow. 	 Citizen centricity missing: OS not user-friendly, led to a decline of users Data underused: non-predictive, limited to tracking. Technical flaws: service shortfalls at kiosks, bus crew usage. Organisational potential underused: fleet utilistion declined Dark data: A lot of the data generated was not put to use immediately, creating critical time lags, or its use suffered due to limited actors.
Metropolis of Rennes, France	 Transport type: Buses, Metro, Cycles – holistic mobility network. Main actors: Public & Private (Data2B with Keolis, STAR, Cycleo) Features: Data sharing facilitated through the city's open data portal. Data2B is one of many st data products in Rennes. 	 Combining weather, events, mobility and historic ticketing data. Predictive software 1: Estimating bus capacities, informing users whether to wait for the next. Predictive software 2: Estimating demand at bike stations, high demand bike routes. 	 Multiple citizen-facing instances: Various services, applications developed with high user-adoption. Data utilised: Multiple stakeholders enter to facilitate different uses Technical efficiency: entry of data product & service providers; creating variety & competition in analytics, Ul's,. Usable data: Open availability allows data to be used swiftly and diversely, often in real time due to multiple actors' access.

Figure 6.2

Global case studies across regulation also present a nuanced and lengthy timeline for robust data sharing in key sectors. For example, Finland's journey to re-use of healthcare data began with the goal of nationwide interoperable EHR (electronic health record) systems by the end of 2007 - the goal having been outlined in 2002. An agreement on the national archive for health information (Kanta) was reached by 2005, the building project for which was defined in 2007 - simultaneous with a law on e-prescriptions. The introduction of electronic prescriptions progressed first so that all Finnish pharmacies were able to receive electronic prescriptions in 2012, and ten years after the law came into force, the electronic prescription became mandatory in Finland in 2017. E-prescribing services are connected to the pharmaceutical database that provides professionals unified and up-to-date information about medicines and medicinal substances to prescribe and dispense medicines.⁴⁰ Kanta services (producing digital services for social welfare and the healthcare sector) launched years later in 2010, and is accessible to any region in Finland. This is also continuously in development and being refined to become more comprehensive. After a patient data repository was set up, and previous pilots aligned, a new act on secondary use of health data was passed in 2019.⁴¹ In summary,

Finland passed a number of regulations spanning nearly 20 years to mandate digitisation and create a database for health data prior to legislation for secondary use of healthcare data (See figure 6.3). In Israel as well, healthcare providers began digitising their patient records over 25 years ago; today, ~98% of the country's medical records are digitally stored as EMRs, available at various stages of care.⁴²

Learnings from such case studies have shown that the task of mapping, preparing and ushering a sector toward data sharing is not only specific to each industry, but also must be staggered in order to mitigate risks, test well for viable sharing structures - if the resultant legislation and ecosystem is to be sufficiently robust and sustainable. While India's challenges are unique and its journey may not map directly to those of other countries, there is a need to take a step back and assess which sectors may be prepared for data sharing, how we can amplify digitisation and infrastructural capacity, and adopt an iterative, layered approach to data governance. To do this, India must adopt the ecosystem-enabled approach - creating a focus on capacity, stakeholder involvement, enabling legislation, and rigorous testing. To this end, our next section details a set of recommendations to the committee for India's NPD sharing approach.

⁴⁰ https://recibus.com/?smd_process_download=1&download_id=340

⁴¹ https://stm.fi/en/secondary-use-of-health-and-social-data

⁴² https://blog.startupnationcentral.org/digital-health/while-helping-save-lives-the-digitization-of-patient-records-poses-unprecedented-privacychallenges/

An overview of Finland's timeline toward the eventual legislation on data sharing for re-use of healthcare data



Figure 6.3



SECTION IV : Recommendations to the Committee

1. Establish an ecosystem-based voluntary approach

The ecosystem approach focusses on enabling infrastructure, incentivising sharing, and adopts a voluntary structure at its core.





Maturity

An ecosystem-based approach to data sharing and related legislation is a promising way to solve for the prematurity of the current Indian ecosystem. While India has a number of data sharing and data-driven initiatives for public good (see Section III), they are not yet prepared or cohesive enough for the burden of widespread adoption of NPD sharing. Moreover, instances of mandatory sharing across the globe (whether in Finland or Australia) have been found to be limited in scope - either by sector, purpose or data type. The NPDR, on the other hand, does not account for the benefits of testing such an approach before implementation. Further, India's current levels of digitisation reflect a number of offline demographics. Pending a move toward onboarding such offline sections of society, NPD sharing threatens to further marginalise and exclude such communities in India's move toward greater data sharing. In the absence of fundamental data privacy legislation too, the contours of NPD remain unclear, as do recourse mechanisms for those possibly harmed during anonymisation and sharing processes. India's Personal Data Protection Bill remains untitled and pending action. Without codifying these foundational elements to protect India's communities, NPD sharing appears untimely, and in need of primary groundwork. An ecosystem-enabled approach would allow the time and effort required to spur this groundwork,

while also progressing toward greater data sharing. The ecosystem approach also allows stakeholders a chance to divide the burden of conceptualising and testing data sharing in India, which would also allow resultant policy to be reflective of varied needs and concerns across the board.

Innovation

In order to catalyse innovation in the context of data, focus must be placed on stakeholders who are most able to create such value. Analysis has shown (see Section I) that private sector entities (particularly larger companies) lead data-driven innovation and progress toward data science applications. Given this, it is pertinent to involve the private sector in formulating data-driven solutions for societal impact. Mandatorily sharing data with the government does not necessarily ensure this, and the NPDR itself does not make clear the processes of value creation. High-value datasets alone, without adequate action to extract value from them, leave unanswered the question of how this will be used and by whom to create broader public value. Further, the report highlights innovation by entrepreneurship and start-ups as a purpose for such sharing. However, from conversations with data-driven small businesses and startups, it is clear that smaller players are apprehensive about the structure of mandatory sharing. Contradictory to the NPDR's goal, responses to the

mandatory approach have been marred by uncertainty on the part of small businesses. Lack of clarity around the processes of such regulation and the burden of compliance constitute concerns for start-ups that require key regulatory stability in order to glean investments. There is also a marked concern around growth of 'data driven businesses' as described by the NPDR. Thus, rather than contributing to the growth of data-driven entrepreneurship, a mandatory approach may well disincentivise data-drivenness across industries. It is important, therefore, to revisit the recommended framework through a consultative and stakeholder-wide lens⁴³ – involving the ecosystem must be the key basis of building the ecosystem.

While global best practices reflect multiple, and often simultaneous steps toward data sharing, India's position is premature to institute a mandate. An ecosystem enabled, voluntary approach will allow opportunities to build upon existing sharing, and inform a sustainable data governance structure.



⁴³ See figure 7.4 for a possible distribution of stakeholder roles

2. Clarify and delineate first principles

As the global move toward increased data sharing for public value gains traction, most nations are untangling what theoretical bases should drive this move for them. India's challenges are unique, dynamic and framed by the context of an immensely diverse population with varied levels of digitisation and social inequities. Data sciences and their applications are also a dynamic field, swiftly changing and entering new avenues. This makes the task of data governance not only a lofty one, but one that must prepare for this dynamism – by building enduring legal frameworks for future governance to rely on. One of the most important pillars in creating such jurisprudence is clarifying the core motivations for data governance. For this, there is a need to take a step back and ask - what are the first principles for India's data sharing journey? As seen in Section II, mitigating antitrust practices, rebalancing competition and creating wider societal good are key motivations for numerous countries. While the NPDR also highlights these, along with a nod to sovereign and business purposes, they need to be fleshed out in order to act as durable grounds for the mechanisms outlined in the NPDR, and for future legislation on data sharing in India. Currently, the first principles outlined by the NPDR go only so far as

to list broader socio-economic policy goals. While these are certainly key starting points, it is unclear how these principles of public value or competition rebalancing are to be adjudicated on. Particularly if the NPDA is to be the sole authority defining such purposes, businesses and communities alike must be able to look to a more detailed outline to understand the contours of this purpose, and the bases for sharing. As a preliminary exploration, see Figure 7.3 for a possible framework to envision the bases of public value and antitrust as first principles in the context of data. This can be understood through the lens of agency, transparency, and tangible benefits for lived experiences (both offline and online).

A framework for unpacking public value and antitrust concerns in the context of data sharing.

Principles		Framework	Key Questions	
PUBLIC VALUE	Participation	• Ability of individuals and communities to exercise decision-making on data governance		
	Protection	• Rights to safeguard against any harms from data collection or processing	 Are citizens empowered to make better decisions based on insights from NPD? Is there overall improvement in health & well-being, environment & transparent 	
	Agency	• Agential rights of rectification, erasure and ability to appoint intermediaries to safe guard interests	governance from the data sharing? • Can the data can be used easily and safely by research and public institutions?	
	Transparency	• Visibility into data collection and processing across stakeholders	 Does it safeguard individual and community rights when sharing data? Does it improve efficiency or increase access to public (up for a comisso?) 	
	Equitable Benefit	• Ability to derive value from data and democratic distribution of value	public/wellare services?	
ANTI TRUST	Access to Markets	• New players access markets with ease, and have the ability to compete		
	Consumer Choice	• Consumers have access to diverse goods and services and are not subjected to monopolistic prices	 Does sharing ensure beneficial use of the data for and by small/medium enterprises? Does it prioritise or spur 	
	Symbiotic Collaboration	• There is two-way collaboration between businesses for innovation and public value	innovation across various stakeholders? • Does it provide greater consumer choice, and sectoral efficiency?	
	Diversified Economy	• Multitude of goods and services available – built by businesses	• Does it reduce barriers to market entry, self-preferencing, and abuse of dominance by gate-keeping platforms?	
	Innovation	• New solutions build through collaboration and improved data		

Figure 7.3

It is also pertinent to ensure that any resultant data sharing structure affords meaningful and agential recourse to individuals and communities involved. In a sense, data and data governance are novel enough to call for a move to the drawing board – to unpack what is needed at the theoretical, constitutional level, before laying down (and informing) what must be done at the implementation level. Without clear notions of data sharing principles for India, uncertainty and apprehension may be inevitable fallouts of legislation. Clarifying these principles will not only serve to build trust and agency for the public sector, but allow private sector entities to have a clearer understanding, and adequate preparation for their involvement in data sharing for societal value. As part of this research, we also recognise that the burden of developing these first principles is complex, to say the least. Given this, our ecosystem approach recommends that relevant stakeholders be involved in the

delineation of first principles. There has been significant academic interest and will in the study of fundamentals for data sharing, and avenues for public value creation. Besides researchers, civil society organisations have also begun to unpack various facets of data governance, be it the role of the intermediary, protection and empowerment of the community, or the need for regulators in the space. A collaborative approach to define India's first principles must identify such stakeholders, and involve them proactively in shaping this basis. Not only would this be welcomed by such stakeholders, and increase confidence in governance processes, and equitably divide the burden of formulation, it would also create a conduit by which community needs, concerns and interests may be represented in regulation through civil society organisations and NGOs rooted in India's communities.

A distribution of responsibilities across stakeholders, commensurate with their abilities is an important element of the ecosystem approach, and reduces both the burden and timeline for data sharing in India.

STAKEHOLDER	DESCRIPTION	EXAMPLE	ROLES & RESPONSIBILITIES	
Government	Government agencies not including regulatory bodies	MoHFW, BMTC	 Safeguard rights and ensuring responsible data use Outline policy goals 	
Consumer - Facing Product Providers	Data businesses that offer end products or services to consumers.	Google, Facebook, Microsoft	 Innovate data-driven solutions Identify key data uses across industries / data types Collaborate with the government to derive solutions 	
Tech Solutions Providers	Businesses offering data-related tech services - analytics, storage, etc.	Scribble Data, Microsoft, Nilenso	 Build sharing & storage architectures Outline data quality & sharing standards 	
Individuals & Communities	Members of society that act as end-users and/or data generators.	Farmer collectives, Migrant labour, Individuals	 Primary role is that of data generators Represent individual / community rights, needs and interests through collective action 	
NGOs & CSOs	Public sector organisations working in the interest of individuals & communities	Jal Jeevika, Jan Sahas	 Outline & support community needs Facilitate communities' involvement in their data governance 	
Researchers	Researchers / academics / organisations in the pursuit of knowledge (any discipline)	Aapti, CIS, ODI	 Identify key sectors and avenues for public value from data sharing Identify & create scientific value from data 	
Philanthropies	Private initiatives providing non-extractive funding toward public good	Omidyar Network, Imaginable Futures	 Connect stakeholders across the ecosystem Support valuable initiatives for data-driven innovation 	

Figure 7.4

India already has a number of sectoral regulators, most of which are wellplaced to speak to the needs of each sector. In the context of data, these differences in sectoral organisation, size, challenges and needs become magnified or, at the very least, much more complicated. The value of data is defined primarily by how it is used; but it is also defined greatly by data type and existing sectoral capacity. This value can also be amplified by sharing data across sectors - something the NPDR recognises. However, upon recognising this, it is pertinent to map clearly how each sector may or may not be able to take on a framework for data sharing. For example, the implications of sharing health data are vastly different from those of sharing agricultural data. While the latter has potential associated harms for farmer communities in India, the former comes with risks to personal privacy as well. While neither issue can be resolved without fundamental data protection legislation in place, they also cannot be resolved without robust mapping of sectoral needs. It may be prudent to empower existing regulators to understand their respective landscapes with the lens of data sharing – ensuring that resultant structures and ecosystems are reflective of their unique needs. Further, the NPDR maintains that sectoral

regulators may specify additional data regulations along with those formulated by the NPDA. This creates a significant risk of overregulation across sectors, increasing the regulatory burden and uncertainty faced by datadriven businesses. There is a need to address such potential overlaps, and we recommend that a sector-first approach be adopted in doing so. This speaks also to the requisite role of an NPD regulator such as the NPDA. In order to avoid regulatory overlaps as well as reduce burden, it is important to empower existing bodies to prepare for data sharing – building their technical expertise, ability to identify data-driven solutions and to run pilots that may communicate with other sectors. Having done this, a regulator like the NPDA may then be streamlined to solve for challenges faced by various sectors, and work toward building a durable and robust ecosystem by amplifying technical and infrastructural capacity.

We also recommend that the sectoral approach to data sharing incorporate a phased implementation. While the eventual goal may be to share data across sectors, it is important to begin by testing and building an ecosystem for especially mature or impactful sectors.

4. Invest in building blocks for technical and infrastructural capacity

Globally, the rise of data-driven economies has been led by the private sector. As a result, the technical structures supporting the data economy have been built in siloes, reflective primarily of market-first interests and hardly oriented towards data sharing. The fallout of this has been a lack of technical and infrastructural pathways to safely and effectively share data across parties and sectors. This has been evidenced in our research of the challenges associated with sharing, and the burden of negotiating agreements and technical safeguards amongst parties that have shared data for public value in the past. In order to mitigate this, and certainly in order to build a nationwide ecosystem for data sharing, there is a primary need to invest in major, collaborationoriented infrastructure. The ecosystemenabled voluntary approach focalises these building blocks, in order to streamline sharing and build trusted networks for sharing and innovation to move through. Interoperability, standardisation and digitisation form the basis of such an approach.

Interoperability and standardisation

Both government and private sector data currently exists in siloed forms, with varied formats, data resolutions and metadata classifications. Consultation has reflected that while access to data poses a problem, a more fundamental challenge emerges from the lack of interoperability across players - even where access is achieved, a large amount of investment and labour goes into cleaning datasets, ensuring that they are usable, and in the taxonomisation of metadata. While the NPDR recognises technical concerns, and proposes that the NPDA regulate on metrics like interoperability, it is unclear how this will take place simultaneously with data sharing without placing an immense burden upon sharing parties and creating risk for individuals and communities whose data is being shared. There is a need to first legislate on data standards and specify levels of anonymity and transparency, buttressed by standardisation on metadata taxonomies.

The EU's data strategy reflects this approach well – creating a pronounced

focus on efficient networks for data sharing, and delineating terms of sharing and detailed mechanisms for data sharing. In order for India to successfully create a lasting ecosystem of interoperable and standardised data sharing, these initial efforts are necessary. India may look to efforts such as Gaia X or X-road (see figure 7.5), in order to successfully institute interoperability and common data infrastructures to streamline data sharing. Further, interoperability is the pillar supporting innovation. Particularly in the case of AI or machine learning, data quality is increasingly important in order to train algorithms and build upon them – sharing alone cannot ensure usability of datasets. For dataholding businesses to move towards new standards and taxonomies would require a paradigm shift, one that must happen as a parent process to NPD sharing, not in conjunction.

In order for nationwide data sharing to become a reality, investment in key technical first efforts are necessary, X-Road and Gaia X present useful case studies for this process.

	GAIA X	X-ROAD
REGION	EU	MULTIPLE: Estonia, Finland, Iceland, Japan, Argentina, Germany, El Salvador
PURPOSE	To develop common requirements for a European data infrastructure and facilitate work and collaboration within the GAIA-X Community.	An open-source software and ecosystem solution that provides unified and secure data exchange between organisations.
PARTICIPATING	The GAIA-X community consists of companies and organisations in and outside of Europe.	An X-Road ecosystem is a community of organizations using the same instance of the X-Road software for producing and consuming services. The ecosystem may be nationwide, or it may be limited to organizations meeting specific criteria.
INTEROPERABILITY STANDARDS / MEASURES	 GAIA X identifies minimum technical requirements and services necessary to operate the federated GAIA-X Ecosystem. The development of these services follow the principles of Security by Design and Privacy by Design. Gaia X addresses issues in aligning network and interconnection providers, Cloud Solution Providers (CSP), High Performance Computing (HPC) as well as sector specific clouds and edge systems. 	 X-Road is a centrally managed distributed data exchange layer between information systems that provides a standardized and secure way to produce and consume services. The basic idea of X-Road is that members of an ecosystem exchange data through access points (Security Servers) that implement the same technical specifications. X-Road implements a set of standard features to support and facilitate data exchange and ensures confidentiality, integrity, and interoperability between data exchange parties.

Figure 7.5

Digitisation and capacity building

As mentioned throughout this research, digitisation remains a challenge for many sections of Indian society. Given that the motivation for data sharing is rooted in societal value, it is important not to leave these communities behind on the journey. The NPDR, while progressing towards increased sharing, does not chart any clear timelines or trajectories for India's digitisation journey. While digitisation is an important step in uplifting sections of Indian society ceteris paribus, it is even more important in the context of NPD sharing. Some of the sectors highlighted in the NPDR include healthcare and agriculture. India's agricultural workers and farmers remain extremely vulnerable due to a lack of digital literacy which often results in an absence of agency regarding datadriven solutions proposed for them. Further, data sharing in healthcare without digitisation as a building block runs the risk of further marginalisation of offline demographics. Most of Indian healthcare, particularly in rural areas - is mediated in analog and rarely are records or physician responses digitised to the extent that they might contribute meaningfully to a larger move toward secondary use. A sister concern with digitisation is India's core infrastructural capacity to carry out data sharing. Per capita availability of physicians and

doctors is still incredibly low, with only 1 practitioner per 1457 people of India's population.⁴⁴ Moreover, governance bodies within these areas must also be provided the capacity and digital literacy to make sense of their data sharing, particularly if they are to act as conduits for recourse as proposed by the NPDA. Without efforts to bring all sections of Indian society along with the journey of data sharing, there is a justified concern that existing inequities will be magnified, and numerous communities will be left with tech-solutionism that they can neither make sense of or have agency over. This outcome would be antithetical to the principles of public value.

⁴⁴ World health statistics, World Health Organization

5. Provide incentives for stakeholders

In order to spur voluntary sharing and build upon existing efforts, it is necessary to instil meaningful incentives for data sharing – particularly to achieve participation and value creation from the private sector. While current thinking around incentives for data sharing is nascent and evolving, there are some that can be considered viable at this stage and may be tested for efficacy.

Possible incentives to data sharing, to foster trust, capacity and policy learnings.

		SYMBIOTIC SHARING	SOLUTION COLLABORATIONS	REGULATORY SANDBOXES	DATA MARKETPLACES
ES	GOVERNMENT ACTION	 Increase G2B, G2C data sharing Provide better data quality and accessibility Create industry standards for sharing through consistency in government data 	 Formulate clear, micro-level policy goals Collaboratively identify data businesses with the ability to solve for stated goals Mandate purpose specific commitments of expertise and effort from identified businesses 	 Create an isolated testing environment for new data sharing structures Invite ecosystem stakeholders to test, provide feedback, and increase visibility on standards Evolve regulations, standards, infrastructure based on learnings 	 Explore platforms where data holders (or users with personal data) can sell datasets / subscriptions of datasets to consumers Develop clear & standardized pricing models through extensive consultation and research
INCENTIV	SOLVING FOR	TRUST Demonstrates government's intention as a player in the data economy willing to share	CAPACITY Allows private sector to solve for societal goals based on ability	STAKEHOLDER INVOLVEMENT Allows holistic ecosystem participation in key policy-making	ECONOMIC INCENTIVES Allows data sharers to be compensated for their datasets
	BENEFITS	 Increased data flow between sectors New avenues for innovation Opportunity to refine and broaden consumer experience 	 Targeted creation of societal value Reputational advantage for private sector Opportunity to test private - public partnerships, scaled per need and feasibility 	 Stakeholder insight into evolving regulation enables the ecosystem to prepare Resultant regulations and infrastructure befitting of sectoral / industry-wise capacity and needs 	 Opportunity to build valuation for data and compensate data generators & holders However, myriad concerns around ethics, monopolistic pricing and difficulty in valuating data are seen in current models.

Figure 7.6

Symbiotic sharing

In order to create greater trust and demonstrate the government's intentions as a player in the data economy willing to share data, data sharing must be equitable and symbiotic. This can be done by increasing G2C and G2B data sharing, creating industry standards for sharing through increased consistency in government data and by providing better data quality and accessibility. Not only does this solve for trust barriers, but can also serve to increase data flow across sectors, set a top-down precedent for willingness to share data, and open up new avenues for innovation. The private sector is greatly incentivised by this, as is the consumer who will benefit from refinements and broader consumer experience as a result of increased data sharing. Building this trust is key to increased data sharing, and can amplify existing data-driven innovations that are missing information or understanding - for example, of demographics that may be offline, or environmental data from the government that will refine such efforts. The existing lack of trust between parties would only be amplified by a mandatory, unidirectional structure of sharing, leading to resistance from the private sector and possible futility of the endeavour. Adopting a staggered, voluntary approach that builds trust before regulation may be an effective way to mitigate these challenges and, in the long term, allow greater pathways for public value and innovation.

Solution collaborations

As with clarifying the first principles of data sharing, outlining solutiondriven collaborations is an important method to solve for uncertainty and for maturity. The process of such collaborations may involve a formulation of clear, micro-level policy goals (for example, districtlevel problem statements) that have potential for data-driven solution. Key areas for such collaboration include air pollution, traffic management, and energy conservation. These goals can then be used to collaboratively identify data businesses with commensurate ability and expertise to be innovating. The commitment sought of businesses in this case is centred around expertise, effort and labour. This allows the private sector to solve for societal goals based on ability, and in the proposed voluntary structure only those parties that are capable of sharing and innovating would participate mitigating risks arising from prematurity. Further, this leads to a targeted creation of societal value with clear outcomes, and presents a testing ground for public-private partnerships: these may then be further scaled to similar challenges, refined for various regions and layered with other similar collaborations. Such incentives create valuable learnings for policymakers, tangible benefits for public value, and a significant reputational advantage for the private sector – a leg-up that is increasingly being recognised by private tech companies across the globe.

Regulatory sandboxes

Regulatory sandboxes present a powerful opportunity both to incentivise data sharing and to build meaningful policy that may be more reflective of stakeholder-wide concerns. A limited space where practitioners and innovators may experiment with designs and prototypes either on the edge of or beyond existing regulatory framework, sandboxes for NPD sharing in India can go beyond the ambit of technical structures. They may also be used to test frameworks of public good purpose, mechanisms for recourse and increased agency, and more. Within companies as well for the larger ecosystem, sandbox testing environments are a feasible means to understand the value of data sharing, the processes behind such value creation, and spur novel ideas and designs while maintaining a temporally bound environment. Given that sandboxes are a tested concept, implemented in various forms across the globe and in India most recently through the NDHM, they can serve as not only a test but an incentive to data sharing for the private sector. Creating an isolated testing environment for potential data sharing structures will not only invite ecosystem stakeholders to experiment and provide feedback but also increase visibility of existing standards, processes and challenges within the private sector. This can help evolve regulations, interoperability standards and capacitybuilding infrastructure based on learnings. Further, allowing holistic ecosystem participation in policymaking lets stakeholders prepare and shift

organisational standards and procedures commensurate with resultant regulations. In turn, the consequent infrastructure will be durable and befitting of sectoral and industry-specific capacity and needs.

Data marketplaces

There is scope to explore (as mentioned in the NPDR itself) compensation plans for data sharing. However, this notion is rather nascent and current instances of data marketplaces point to a myriad of ethical and economical concerns. As platforms where data holders (or users of personal data) can sell datasets to consumers, data marketplaces allow consumers to purchase or subscribe to these datasets, depending on the model. There are several data marketplaces functioning, and they vary in governance structures, pricing models, data-sourcing methods and subscription models. Community-driven open-source models like Ocean Protocol allow anyone to buy and sell data, using smart contracts and trial periods for monthly data subscriptions. Platforms like Dawex cater to organisations that want to buy or sell data and operate as global marketplaces of >7,000 organisations across industries. However, pricing models for data are not standardised - it is hard to valuate datasets as this value changes with purpose and use. Data monetisation also presents ethical concerns around the reduction of the 'value' of an individual or community's data to economic value - and how this may harmfully incentivise individuals and communities to sell data. Marketplaces

can also result in monopolistic pricing, thus icing out smaller players or newcomers who cannot afford key datasets for innovation. Thus, in order for this to be a viable model for data sharing under law, there is a need to develop clear and standardised pricing models, through extensive consultation and research, determine what data types and parties may fit these models, and solve for the aforementioned ethical concerns around data monetisation.

6. Test NPD sharing through sandboxes and pilots

The overarching concerns associated with the current form of the NPDR are based in prematurity and uncertainty. The primary solution and first step for India to begin to understand and govern NPD is robust and simultaneous testing of such sharing. It is equally necessary to build upon existing pilots and initiatives (see section III), and align them with new learnings and sectoral needs. Regulatory sandboxes, as mentioned above, provide a particularly viable means to understand structural needs, involve key stakeholders and distribute the burden of testing providing valuable insights for eventual

governance. Sandboxes also create a focus on innovation, a primary goal for the NPDR, by allowing companies to test new ideas in a live, time-bound manner. With testing oversight, regulators are then enabled to better understand data sharing in the private sector. Beyond sandboxes as well, existing tests and pilots within India may be amplified, refined and renewed toward a robust and durable ecosystem. We propose a roadmap for India's data sharing journey (see figure 7.7) that incorporates sandboxes, pilots and testing with enabling legislation.





Figure 7.7

There are multiple ways to do this, and early exploration presents a number of benefits and approaches to such testing. As tech-enabled disruption and innovation rises, it is important to create an agile and conducive environment not only for responsible testing to take place, but for diverse stakeholders to participate in and understand datadriven innovation and societal value that may be derived from it.

Benefits:

Evidence

- Create reliable evidence on efficacy and impact on the ground
- Test consequences of policy at scale, especially to understand policy interface with marginalised/otherwise vulnerable communities
- Opportunity to mitigate future systemic failures through learnings from pilots

Reputation

- Enhance the government's reputation as consultative and evidence driven
- · Build trust between government and community, private sector

Transparency & Accountability

- Ensure transparency of results and the openness on decisions to proceed or withdraw policy
- Build feasible, codified mechanisms for accountability through pilots and learnings

Possible approaches:

Randomised Control Trials

• Randomised experiments aim to test a policy idea or innovation by investigating what difference it has made for the people it is aiming to help

Quasi-Experimental Designs

• Quasi-experimental designs (QEDs) use statistical methods to create a comparison group, allowing us to learn about how an innovation works and what impact it has had

Pre-Experiment

• Pre-experiments compare one group of participants before and after an intervention, to see what's changed.

A resultant ecosystem, as recommended in this research, must incorporate regulation, infrastructure, standardisation and intermediaries in order to amplify existing relationships across stakeholders, and create a supporting environment to layer further data sharing initiatives.



- · Outline goals for data efforts in a consultative and
- inclusive manner • Establish sectoral regulators

Infrastructure

- Ensure data resolution and interoperability
- Build technical pathways across disciplines/industries
- Create data repositories, commons, open data hubs
- · Building standards on best practices for data sharing for
- sector specific data sharing
- · Developing central structures for open data

Intermediaries

- Identifying and facilitating valuable data sharing partnerships
- across stakeholders through data trusts, cooperatives, collaboratives etc.
- Protecting individual/community rights by layering accountability and
- representing rights of data generators
- · Managing regulation-resultant central structures/platforms for data sharing

Figure 7.8



CONCLUSION

This research has found that nonpersonal data sharing, as framed by the NPDR's mandatory approach, poses a premature step in an extremely nascent Indian data ecosystem. However, efforts required to prepare for wider data sharing in India are tedious, and the recommendations detailed in this work must be viewed as an ecosystem effort. The state must serve to bring stakeholders together, adopting a consultative and iterative approach to data governance in order to harness the immense and latent potential of non-personal data. Further, this research has been limited in its scope to the proposal of the expert committee, and many open questions remain. There remains a lot to be unpacked for both global and Indian

data governance - community data rights, the role of data intermediaries and stewards, reimagining incentives in a data driven context, understanding the economics of data markets, and more. It is hoped that the existing network of academic, policymakers, civil society and industry experts will build upon such research. Before moving to mandate nationwide data sharing, it is important for the Government of India to not only invest in key building blocks to ensure a capable data sharing infrastructure, but to do so by ramping up collaboration, testing and communication with stakeholders; and to arrive at the core principles and fundamental jurisprudence for Indian data governance.



ANNEXE

A: Bibliography

Includes works cited in-line and other works referenced

- 1. A European strategy for data (2020), European Commission. Accessible at <u>https://ec.europa.eu/info/strategy/</u> <u>priorities-2019-2024/europe-fit-digital-age/</u> <u>european-data-strategy#documents</u>
- 2. Act on Special Measures for Productivity Improvement (2018), Japan. Translation accessible at <u>http://www.</u> japaneselawtranslation.go.jp/law/detail_ main?re=02&vm=02&id=3420
- 3. Act on Transport Services (2018), Ministry of Transport and Communications, Finland. Accessible at <u>https://www.finlex.fi/fi/laki/</u> <u>kaannokset/2017/en20170320_20180731.</u> <u>pdf</u>
- 4. Adem Koyuncu, Valerie Mei (2020), Germany Prepares New Law for Patient Data Protection and Increased Digitalisation in Healthcare and for "Data Donations" for Research Purposes. https://www.covingtondigitalhealth. com/2020/08/germany-prepares-new-lawfor-patient-data-protection-and-increaseddigitalisation-in-healthcare-and-for-datadonations-for-research-purposes/
- 5. AIG (2017), The Data Sharing Economy: Quantifying Tradeoffs that Power New Business Models, American International Group, Inc. Retrieved from <u>https://www.aig.</u> <u>com/content/dam/aig/america-canada/us/</u> <u>documents/insights/aig-the-data-sharing-</u> <u>economy.pdf</u>
- Bergemann, Dirk et al (2019), The Economics of Social Data (September 25, 2019). Cowles Foundation Discussion Paper No. 2203R. <u>https://ssrn.com/</u> <u>abstract=3459796</u>

- 7. Credit Data Law, Bank of Israel (2017). https://www.boi.org.il/en/CreditRegister/ Pages/Default.aspx
- 8. Data Sharing and Release Bill New Australian Government Data Sharing and Release Legislation: Issues paper for consultation (2018), Department of the Prime Minister and Cabinet, Australia. https://pmc.gov.au/resource-centre/publicdata/issues-paper-data-sharing-releaselegislation
- 9. Dr. Hans-Geog Kamann et al (2021), German Antitrust Law Amendment Entered Into Force: "Digitalization Act" Establishes New Antitrust Rules For The Digital Economy. <u>https://</u> www.mondaq.com/germany/ antitrust-eu-competition-/1028988/ german-antitrust-law-amendmententered-into-force-digitalization-actestablishes-new-antitrust-rules-for-thedigital-economy
- Enders, Tobias. (2018). Exploring the Value of Data – a Research Agenda. <u>https://www. researchgate.net/publication/327246406_</u> <u>Exploring_the_Value_of_Data_-a_</u> <u>Research_Agenda</u>
- 11. Establishing incentives and changing cultures to support data access (2014), EAGDA (Expert Advisory Group on Data Access) under Cancer Research UK, Wellcome Trust et al. <u>https://wellcome.org/ sites/default/files/establishing-incentivesand-changing-cultures-to-support-dataaccess-eagda-may14.pdf</u>
- 12. European Commission (2021), Good practices on B2G data sharing: Finnish

forest data ecosystem. <u>https://digital-</u> strategy.ec.europa.eu/en/node/1121/ printable/pdf

- 13. European Political Strategy Centre (2017), Enter the Data Economy: EU Policies for a Thriving Data Ecosystem, Brussels.
- 14. Fia, Tommaso (2021), An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons, Global Jurist, vol. 21, no. 1, 2021, pp. 181-210. <u>https://doi.org/10.1515/gj-2020-0034</u>
- 15. Gal Omer (2020), Israel draft regulations protect the use of health data for research purposes, IAPP. Link
- 16. Graef, Inge and Gellert, Raphael and Husovec, Martin, Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation (September 27, 2018). TILEC Discussion Paper No. 2018-029, Available at SSRN: <u>https://ssrn.com/</u> <u>abstract=3256189</u>
- 17. Graef, Inge and Prufer, Jens (2018), Mandated Data Sharing Is a Necessity in Specific Sectors, Economisch Statistische Berichten, vol. 103 (4763) 5 July 2018, p. 298-301, Tilburg Law School Research Paper. <u>https://ssrn.com/abstract=3206685</u>
- Graef, Inge and Prufer, Jens, Mandated Data Sharing Is a Necessity in Specific Sectors (2018). Economisch Statistische Berichten, vol. 103 (4763) 5 July 2018, p. 298-301., Tilburg Law School Research Paper, Available at SSRN: <u>https://ssrn.com/ abstract=3206685</u>
- IDC and Lisbon Council, Story 1: Opening Up Private Data for Public Interest. Retrieved from <u>https://datalandscape.</u> <u>eu/data-driven-stories/story-1-openingprivate-data-public-interest</u>

- 20. Impact of Non-personal Data Report (2020), Fifth Elephant, Hasgeek. Accessible at <u>https://hasgeek.com/fifthelephant/</u> <u>impact-of-npd/</u>
- India's Non-Personal Data (NPD) framework: Knowledge repo, archives and collaborations, Hasgeek (2021). Submissions accessible at <u>https://hasgeek. com/PrivacyMode/non-personal-data/sub</u>
- 22. Jens Prufer (2021), Mandatory Data Sharing: Development of a Test & Governance Structure. <u>https://prufer.</u> <u>net/2021/02/04/mandatory-data-sharingdevelopment-of-a-test-governancestructure/#_ftn1</u>
- Josep-Ramon Ferrer, Barcelona's Smart City vision: an opportunity for transformation, Field Actions Science Reports [Online], Special Issue 16 | 2017, Online since 01 June 2017, connection on 20 July 2021. Accessible at <u>http://journals. openedition.org/factsreports/4367</u>
- 24. Josh D'Addario et al; ODI, (2020), Sharing data to create value in the private sector, ODI. Retrieved from <u>https://theodi.org/</u> <u>article/report-sharing-data-to-create-value-</u> <u>in-the-private-sector/.</u>
- 25. Kapoor, Astha (2021), Collective bargaining on digital platforms and data stewardship, Friedrich Ebert Stiftung, Asia. <u>https://asia.</u> <u>fes.de/news/fow-collective-bargaining</u>
- 26. KPMG International (2018), What to Tax?, KPMG. <u>https://assets.kpmg/content/dam/</u> kpmg/it/pdf/2018/09/What-to-TAX.pdf
- 27. Lisbon Council, Deloitte (2020). Study on public sector data strategies, policies and governance: Data analytics for Member States and Citizens. <u>https://joinup.</u> <u>ec.europa.eu/sites/default/files/custompage/attachment/2020-06/DIGIT%20-%20</u>

D01%20-%20Study%20on%20public%20 sector%20data%20strategies%2C%20 policies%20and%20governance%20 v3annexes.pdf

- 28. Manohar, S., Kapoor, A., Ramesh A. (2019), Understanding data stewardship: taxonomy and use cases, The Data Economy Lab, Aapti Institute. <u>https://uploads.strikinglycdn.</u> <u>com/files/64aa4010-6c11-4d6f-8463-</u> <u>efaed964d7d9/Understanding%20Data%20</u> <u>Stewardship%20-%20Aapti%20Institute.pdf</u>
- 29. Martens, B., Duch-Brown, N., The economics of Business-to-Government data sharing, European Commission, Seville, 2020. JRC119947.
- 30. Mayer-Schönberger V., and Kenneth. C. (2013), Big Data : A Revolution That Will Transform How We Live, Work and Think, John Murray Press.
- 31. McDonald, Sean (2019), The Fiduciary Supply Chain, Cigionline. Accessible at <u>https://www.cigionline.org/articles/</u> <u>fiduciary-supply-chain/.</u>
- 32. McDonald, Sean. 2019a. "How Regulations Are Reshaping Digital Companies." Cigionline, April 15. <u>www.cigionline.org/</u> <u>articles/how-regulations-are-reshaping-</u> <u>digital-companies</u>
- 33. Mills, Stuart, Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership (September 24, 2019). Available at SSRN: <u>https://ssrn.</u> <u>com/abstract=3437936</u> or <u>http://dx.doi.</u> <u>org/10.2139/ssrn.3437936</u>
- 34. Nanda, Amrita (2020), Power Structures in the Data Economy, Aapti Institute. <u>https:// thedataeconomylab.com/2020/10/30/</u> power-structures-in-the-data-economy/
- 35. OECD (2019), "Risks and challenges of

data access and sharing", in Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, OECD Publishing, Paris. <u>https:// doi.org/10.1787/15c62f9c-en</u>.

- 36. OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, OECD Publishing, Paris. <u>https://doi.org/10.1787/276aaca8-en</u>.
- 37. Osaka Track: Ministerial Statement on Trade and Digital Economy (2019). <u>https://www.meti.go.jp/ press/2019/06/20190610010/</u> 20190610010-1.pdf
- 38. Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. Data Science Journal, 16, 8. DOI: <u>http://doi.org/10.5334/dsj-2017-008</u>
- Pasquetto, I.V., Randles, B.M. and Borgman, C.L., 2017. On the Reuse of Scientific Data, Data Science Journal, 16, p.8. DOI: <u>http://doi.org/10.5334/dsj-2017-008</u>
- 40. Paul Romer (2020), Progressive, Pigouvian taxes can protect competition and democracy.
- 41. Peppin, Aidan (2020), Doing good with data: what does good look like when it comes to data stewardship?Ada Lovelace Institute. <u>https://www.adalovelaceinstitute.</u> <u>org/blog/what-does-good-look-like-datastewardship/</u>
- 42. Personal Information Protection Act 2004, Tasmania (2004). <u>https://www.legislation.</u> <u>tas.gov.au/view/html/inforce/current/act-</u> <u>2004-046#GS1@EN</u>
- 43. Phillips, M (2018), International datasharing norms: from the OECD to the General Data Protection Regulation

(GDPR). Hum Genet 137, 575–582 (2018). https://doi.org/10.1007/s00439-018-1919-7

- 44. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act), 2020. Accessible at https://eur-lex.europa.eu/legal-content/EN/ TXT/?uri=CELEX:52020PC0767
- 45. Prufer, Jens and Graef, Inge (2021), Governance of Data Sharing: a Law & Economics Proposal. TILEC Discussion Paper No. 2021-001, CentER Discussion Paper No. 2021-004. Accessible at SSRN: https://ssrn.com/abstract=3774912
- 46. Prüfer, J. (2020). Competition Policy and Data Sharing on Data-driven Markets: Steps Towards Legal Implementation.
 Friedrich Ebert Stiftung. Retrieved from <u>http://library.fes.de/pdf-files/fes/15999.pdf</u>
- 47. REGULATION (EU) 2018/1807 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 November 2018: On a framework for the free flow of nonpersonal data in the European Union. https://eur-lex.europa.eu/legal-content/EN/ TXT/PDF/?uri=CELEX:32018R1807&from=EN
- 48. REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on contestable and fair markets in the digital sector (Digital Markets Act), 2020. Accessible at <u>https://eur-lex.</u> <u>europa.eu/legal-content/EN/TXT/</u> <u>PDF/?uri=CELEX:52020PC0842&from=en</u>
- 49. Report by the Committee of Experts on Non-Personal Data Governance Framework (July 2020), Kris Gopalakrishnan et al. Ministry of Electronics and Information Technology, GOI. https:// static.mygov.in/rest/s3fs-public/ mygov_159453381955063671.pdf

- 50. (Revised) Report by the Committee of Experts on Non-Personal Data Governance Framework (December 2020), Kris Gopalakrishnan et al. Ministry of Electronics and Information Technology, GOI. <u>https://static.mygov.in/rest/s3fs-</u> <u>public/mygov_159453381955063671.pdf</u>
- 51. Richter, H., Slowinski, P.R. The Data Sharing Economy: On the Emergence of New Intermediaries. IIC 50, 4–29 (2019). <u>https://doi.org/10.1007/s40319-018-00777-</u> Z
- 52. Rowhani-Farid, A., Allen, M. & Barnett, A.G. What incentives increase data sharing in health and medical research? A systematic review. Res Integr Peer Rev 2, 4 (2017). Accessible at https:// researchintegrityjournal.biomedcentral. com/articles/10.1186/s41073-017-0028-9
- 53. Setu Bandh Upadhyay (2021), Regulatory Overlap Issues in the NPD Committee Report, Hasgeek. <u>https://hasgeek.com/</u> <u>PrivacyMode/non-personal-data/sub/</u> <u>regulatory-overlap-issues-in-the-npd-</u> <u>committee-rep-3HQVNRki8tyQdkhGS6x9fn</u>
- 54. Sonja, Marjanovic et al., Understanding value in health data ecosystems: A review of current evidence and ways forward, Rand health quarterly Vol. 7,2 3. 29 (Jan. 2018). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5798965/
- 55. Stahl, Florian & Schomm, Fabian & Vossen, Gottfried & Vomfell, Lara.
 (2016). A classification framework for data marketplaces. Vietnam Journal of Computer Science. 3. 10.1007/s40595-016-0064-2
- 56. Statista (2020), The 100 largest companies in the world by market capitalization in 2020. https://www.statista.com/ statistics/263264/top-companies-in-theworld-by-market-capitalization/

- 57. The Data Exchange Framework (2019), Australian Government. <u>https://</u> <u>dex.dss.gov.au/sites/default/files/</u> <u>documents/2019-09/data-exchange-</u> <u>framework-2019.pdf</u>
- 58. The Department of Premier and Cabinet, Tasmania, Administrative Data Exchange Protocol for Tasmania (ADEPT). <u>http://</u> www.dpac.tas.gov.au/divisions/digital_ strategy_and_services/policies/digital_data_ privacy/adept#legal-issues
- 59. The Personal Data Protection Bill (2019), as introduced in Lok Sabha, GOI. <u>http://164.100.47.4/BillsTexts/LSBillTexts/</u> <u>Asintroduced/373_2019_LS_Eng.pdf</u>
- 60. Vikas Kathuria, Jure Globocnik (2020). Exclusionary conduct in data-driven markets: limitations of data sharing remedy, Journal of Antitrust Enforcement, Volume 8, Issue 3, November 2020, Pages 511–534. <u>https://doi.org/10.1093/jaenfo/jnz036</u>
B : Further analysis of policy pathways

Analysis of data sharing regulations across primary goals (1/4)

	1. EUROPEAN UNION
European Strategy for Data	 To empower citizens to make better decisions based on insights gleaned from non-personal data. Enabling data availability to all – whether public or private, big or small, start-up or giant, to help society to get the most out of innovation and competition and ensure that everyone benefits from a digital dividend. Europe aims to capture the benefits of better use of data, including greater productivity and competitive markets, but also improvements in health and well-being, environment, transparent governance and convenient public services.
Data Governance Act	 Data-driven innovation that bring enormous benefits for citizens, for example through improved personalised medicine, new mobility, and its contribution to the European Green Deal. Addressing the barriers to a well-functioning data-driven economy and to create a Union-wide governance framework for data access and use, in particular regarding the re-use of certain types of data held by the public sector, the provision of services by data sharing providers to business users and to data subjects, as well as the collection and processing of data made available for altruistic purposes by natural and legal persons. Application of privacy-enhancing technologies, together with comprehensive data protection approaches to ensure the safe re-use of personal data and commercially confidential business data for research, innovation and statistical purposes.
Digital Markets Act	 Creating a fairer business environment Enabling Innovators and technology start-ups to have new opportunities to compete and innovate in the online platform environment without having to comply with unfair terms and conditions limiting their development. Enhancing Consumer Choice in terms of services and pricing.

Analysis of data sharing regulations across primary goals (2/4)

	2. AUSTRALIA
Data Availability and Transparency Bill	 Serve the public interest by- Promoting better availability of public sector data; and enable consistent safeguards for sharing public sector data; and enhance integrity and transparency in sharing public sector data; and build confidence in the use of public sector data; and establish institutional arrangements for sharing public sector data.
Competition and Consumer Amendment (Motor Vehicle Service and Repair Information Sharing Scheme) Bill 2020	 Promote competition between Australian motor vehicle repairers and establish a fair playing field by mandating access to diagnostic, repair and servicing information for motor vehicles covered by the Bill to all repairers and Registered Training Organisations on fair and reasonable commercial terms. Enhanced sharing of repair information by motor vehicle manufacturers to enable repairers to effectively remedy issues. This measure improves overall choice and price competition.
	3. FINLAND
The Act on Transport Services	 To promote digitalisation of transport services and enables efficient use of data in the Transport Sector. To enable interoperability of information and ticketing systems and openness of interfaces.
Law on Secondary Use of Health and Social Data	 The aim is to enable the efficient and secure processing of personal data stored in social and health care activities and for social, health, guidance, research and statistical purposes, and to combine them with the personal data of the Social Insurance Institution, Population Register Center, Statistics Finland and the Finnish Center for Pensions. The law also aims to safeguard the protection of the individual's trust and the rights and freedoms when processing personal data.

Analysis of data sharing regulations across primary goals (3/4)

	4. ESTONIA
Information systems data exchange layer	 Builds a secure technical infrastructure and environment between X-Road members that enables secure and evidence-based Internet data exchange
	5. ISRAEL
Patient Rights Regulations (Health Data Research use), 2019 Credit Data Law	 To set the legal framework for the use of health data for research purposes while balancing research in respect of health data with privacy and confidentiality of health data. To enhance ability to provide quality medical care through research. To help increase competition in the retail credit market, expand access to credit, enable an expansion of the
	information available to credit providers wishing to evaluate a customer's credit risk level, and constitute an anonymous information base that will serve the Bank of Israel in carrying out its functions, including for macroeconomic research purposes and for the development of effective policy tools.
	6. GERMANY
Patient Data Protection Act	 To enhance Digitization, and provide protections for sensitive health data. Creating a framework wherein patients may be able to donate their data for secondary use by researchers.
German Act against Restraints of Competition (ARC) "Digitalization Act"(10th Amendment)	 To restrict the market power of large digital companies through creating obligations and prohibitions.

Analysis of data sharing regulations across primary goals (4/4)

	7. JAPAN
Act on Special Measures for Productivity Improvement and Amendment to Industrial Competitiveness Enhancement Act Enacted	• To maintain and strengthen the international competitiveness of Japanese industries, amid the dramatic changes in the industrial structure and conditions for international competition due to the rapid progress in technological innovation in recent years in the field of information technology, to achieve productivity in a short period by carrying out, initiatives for productivity improvement in a short period through innovative business activities such as promoting the testing of new technology, etc. and promoting innovative use of data for industrial activities, and thereby contributing to the enhancement of the lives of the people and the sound development of the national economy.

FINDABILITY	This is a step higher than access, where provisions are made efficient- for eg - making data vailable in digital or machine readable format.		Secondary Focus	Absent	Primary Focus	Secondary Focus
RE - USABILITY	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection		Absent	Secondary Focus	Primary Focus	Secondary Focus
ACCESSIBILITY	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access		Primary Focus	Secondary Focus	Primary Focus	Primary Focus
CROSS - SECTOR SHARING	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	EAN UNION	Absent	Absent	Secondary Focus	Secondary Focus
TRANSPARENCY	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	1. EUROPI	Primary Focus	Primary Focus	Primary Focus	Primary Focus
INTERO- TPERABILITY	Regulation mandates / provides for the need to ensure communication between multiple systems where data is stored		Secondary Focus	Secondary Focus	Primary Focus	Primary Focus
DATA PROTECTION	Regulation mandates compliance to security standards / creates bodies for ensuring compliance ensures data is transported through secure channels		Secondary Focus	Absent	Secondary Focus	Secondary Focus
Feature	Description		Digital Markets Act	Digital Services Act	Data Governance Act	Regulation on Free Flow of Data

ABSENT : No mention

SECONDARY : If document addresses it

FINDABILITY	This is a step higher than access, where provisions are made efficient-for eg - making data vailable in digital or machine readable format.		Secondary Focus	Secondary Focus
RE - USABILITY	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection		Primary Focus	Absent
ACCESSIBILITY	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access		Primary Focus	Primary Focus
CROSS - SECTOR SHARING	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	TRALIA	Absent	Absent
TRANSPARENCY	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	2. AUS	Primary Focus	Secondary Focus
INTERO- TPERABILITY	Regulation mandates / provides for the need to ensure ensure between multiple systems where data is stored		Absent	Absent
DATA PROTECTION	Regulation mandates compliance to security standards / creates bodies for ensuring compliance ensures data is transported through secure channels		Primary Focus	Primary Focus
Feature	Description		Data Availability and Transparency Bill	Competition and Consumer Amendment (Motor Vehicle Service and Repair Information Sharing Scheme) Bill 2020

ABSENT : No mention

SECONDARY : If document addresses it

Feature	DATA PROTECTION	INTERO- TPERABILITY	TRANSPARENCY	CROSS - SECTOR SHARING	ACCESSIBILITY	RE - USABILITY	FINDABILITY
Description	Regulation mandates compliance to security standards / creates bodies for ensuring compliance ensures data is transported through secure channels	Regulation mandates / provides for the need to ensure communication between multiple systems where data is stored	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection	This is a step higher than access, where provisions are made efficient-for eg - making data available in digital or machine readable format.
			3. FIN	VLAND			
The Act on Transport Services	Secondary Focus	Primary Focus	Secondary Focus	Secondary Focus	Primary Focus	Secondary Focus	Primary Focus
Law on Secondary Use of Health and Social Data	Primary Focus	Absent	Primary Focus	Secondary Focus	Primary Focus	Primary Focus	Secondary Focus
Finnish Forest Act	Secondary Focus	Absent	Primary Focus	Secondary Focus	Primary Focus	Primary Focus	Primary Focus

ABSENT : No mention

SECONDARY : If document addresses it

FINDABILITY	This is a step higher than access, where provisions are made efficient- for eg - making data available in digital or machine readable format.		Primary Focus		Absent	Absent
RE - USABILITY	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection		Secondary Focus		Primary Focus	Primary Focus
ACCESSIBILITY	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access		Primary Focus		Primary Focus	Primary Focus
CROSS - SECTOR SHARING	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	TONIA	Primary Focus	SRAEL	Absent	Absent
TRANSPARENCY	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	4. ES	Primary Focus	5. 15	Primary Focus	Secondary Focus
INTERO- TPERABILITY	Regulation mandates / provides for the need to ensure communication between multiple systems where data is stored		Primary Focus		Absent	Absent
DATA PROTECTION	Regulation mandates compliance to security standards / creates bodies for ensuring compliance ensures data is transported through secure channels		Primary Focus		Primary Focus	Primary Focus
Feature	Description		Information systems data exchange layer		Patient Rights Regulations (Health Data Research use), 2019	Credit Data Law

ABSENT : No mention

SECONDARY : If document addresses it

PRIMARY: If mentioned in object

80

FINDABILITY	This is a step higher than access, where provisions are made efficient- for eg - making data available in digital or machine readable format.		Absent	Absent
RE - USABILITY	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection		Primary Focus	Absent
ACCESSIBILITY	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access	1	Primary Focus	Primary Focus
CROSS - SECTOR SHARING	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	RMANY	Absent	Absent
TRANSPARENCY	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	6. GER	Primary Focus	Primary Focus
INTERO- TPERABILITY	Regulation mandates / provides for the need to ensure ensure communication between multiple systems where data is stored		Absent	Secondary Focus
DATA PROTECTION	Regulation mandates compliance to security standards / creates bodies for ensuring compliance ensures data is transported through secure channels		Primary Focus	Secondary Focus
Feature	Description		Patient Data Protection Act	German Act against Restraints of Competition (ARC) "Digitalization Act"(10th Amendment)

ABSENT : No mention

SECONDARY : If document addresses it

Feature	DATA PROTECTION	INTERO- TPERABILITY	TRANSPARENCY	CROSS - SECTOR SHARING	ACCESSIBILITY	RE - USABILITY	FINDABILITY
Description	Regulation mandates compliance to security standards / creates bodies for ensures data is transported through secure channels	Regulation mandates / provides for the need to ensure communication between multiple systems where data is stored	Regulation provides a clear mechanism for sharing data, mandating release of information on use of data	Regulation provides for sharing of data between sectors like health, education, mobility, energy, etc.	Regulation mandates or aids third parties including government agencies, businesses, citizens, organisations etc., in data access	Regulation allows or aids the use of data for a purpose other than the primary purpose of collection	This is a step higher than access, where provisions are made efficient- for eg - making data available in digital or machine readable format.
			7. J	APAN			
Act on Special Measures for Productivity Improvement and Amendment to Industrial Competitiveness Enhancement Act Enacted	Secondary Focus	Absent	Primary Focus	Secondary Focus	Primary Focus	Primary Focus	Absent
			8. II	NDIA			
Report by the Expert Committee on Non-Personal Data Governance Framework	Secondary Focus	Secondary Focus	Secondary Focus	Secondary Focus	Primary Focus	Primary Focus	Secondary Focus

(6/6)

ABSENT : No mention

SECONDARY : If document addresses it

PRIMARY: If mentioned in object

Understanding NPD Sharing - A Principle First Approach

C: List of experts interviewed

Viktor Mayer-Schönberger Oxford University

Jens Prüfer Tilburg University

Maria R. Coduti European Commission

Tommaso Fia European University Institute

Alexsis Wintour Lapin Institute

Nehaa Chaudhri Ikigai Law

Sreenidhi Srinivasan Ikigai Law

Nikhil Narendran Trilegal Rahul Singh Khaitan & Co

Gowree Gokhale Nishith Desai Associates

Udbhav Tiwari Mozilla

Keshav Dhakkad Microsoft

Rohini Srivathsa Microsoft

Venkata Pingali Scribble Data

Deepa Venkatraman Nilenso

Members of the Google Maps, Google AI and Google Research teams

aapti institute 2021

Aapti is a public research institute that works at the intersection of technology and society. It examines the ways in which people interact and negotiate with technology both offline and online.





This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 India License.

To view a copy of this license, visit <u>http://creativecommons.org/licenses/by-nc-sa/2.5/in/</u>