Policy White Paper

Guidelines for Non-Personal Data (NPD) Governance

2025

CREDITS AND ACKNOWLEDGEMENTS

Lead organisations:

Indian Institute of Science, Bangalore and Center for Digital Public Goods, Indian Institute of Management Bangalore

Authors:

Anjula Gurtoo

Professor, Indian Institute of Science, Bangalore

Jyotirmoy Dutta

Specialist Scientist, Society for Innovation and Development (SID), Indian Institute of Science, Bangalore

Minnu Malieckal

Project Associate II, Centre for Society and Policy, Indian Institute of Science, Bangalore

Srinivasan R

Professor of Strategy and Chairperson, Center for Digital Public Goods, Indian Institute of Management Bangalore

Chaitrali Bhoi

Research Associate, Center for Digital Public Goods, Indian Institute of Management Bangalore

Other Contributors:

The authors sincerely thank the following individuals for their valuable contributions to this report through discussions and deliberations in the Stakeholder Consultation Meetings held on Governance Framework for Data on Digital Platforms (Date: 18.11.2022) and Deliberations on Data Quality and data Pricing (Date: 03.02.2023)

Sr No	*Name*
1	Ananya Dasgupta
2	Anirban Brahmachari, Head of Delivery Management & Delivery Excellence, Google Cloud
3	Ankita Kapoor, Program Manager, SAFETIPIN
4	Anoop G Prabhu, Co-Founder & CTO, Vehant Technologies
5	Arushi Goel, Specialist, Data Policy and Blockchain, World Economic Forum
6	Ashish Aggarwal, Vice-President & Head, Public Policy, NASSCOM
7	Avik Sarkar, Researcher & Visiting Faculty, Indian School of Business
8	Geeta Menon
9	Gunjan Saini
10	Nalin Agarwal, Product Manager, SAFETIPIN
11	Nandini Chami, Deputy Director, IT for Change+
12	Nishant Chadha, Head of Research, India Development Foundation
13	Prasanna Raghavendra
14	Prof. Inder Gopal, CEO, India Urban Data Exchange
15	PV Rai, Managing Director, Pixcel Softek
16	Rishabh Bezbaruah
17	Sanjeev Narsipur
18	Shefali Girish, Research Analyst, AAPTI
19	Shreevyas H M, Scientist and Project Director, e-Governance, Government of Karnataka
20	Srijoni Sen, Assistant Professor, National Law School of India University
21	Suresh Kumar, Head -IUDX & Data Spaces, India Urban Data Exchange
22	Department of Urban Land Transport

- *Names included with consent of the stakeholders*

Disclaimer

Contents

1

SUMMARY

This report serves as a comprehensive guideline for the governance of Non-Personal Data (NPD) in India, established under the Draft National Data Governance Framework Policy (NDGFP) 2022. Its primary objective is to facilitate seamless data sharing among various stakeholders, including government entities, private companies, and researchers, while ensuring ethical use and adherence to quality standards. Governance Framework Outlines mechanisms for regulating NPD, focusing on **Data Access and Disclosure, Data Quality and Standards, Usage Rights, User Charges and Pricing, Ethical and Fair Use of Data and Policy Monitoring and Enforcement.**

This report is designed for a diverse range of stakeholders involved in the management and governance of Non-Personal Data (NPD) in India. Specifically, it targets:

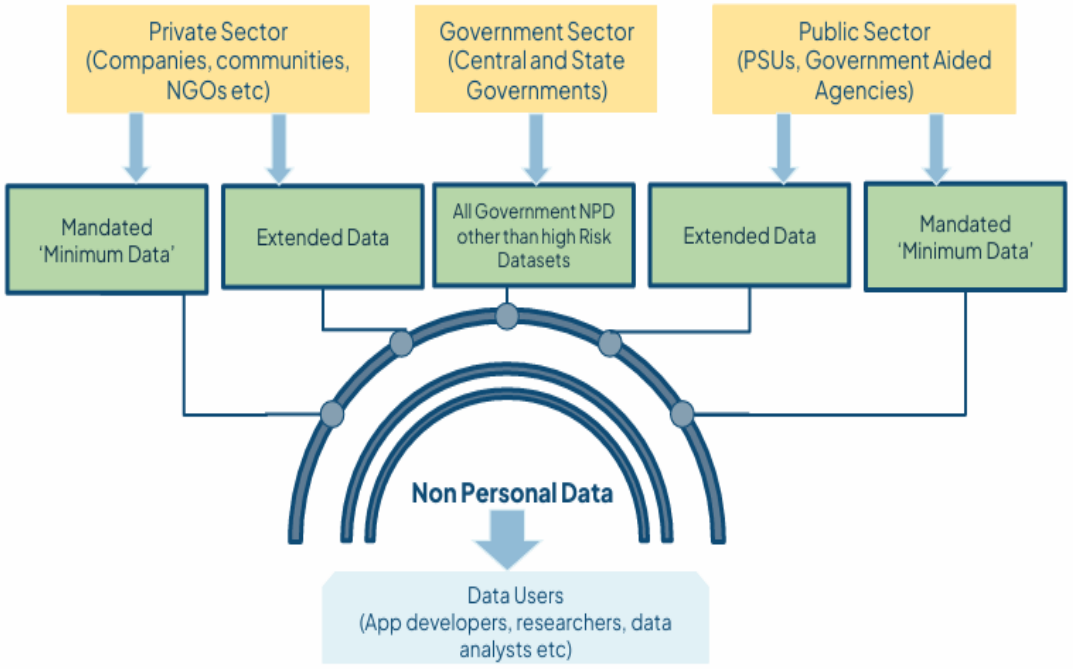
- **Data Principal:** An individual or entity that owns the data and has the right to control its use.
- **Data Collectors:** Private companies, government departments, research projects, and educational institutions that gather data.
- **Data Managers:** Entities such as exchange platforms and data fiduciaries responsible for storing and managing data.
- **Data Users:** Application developers, data analysts, and those involved in machine learning operations who utilize data for various applications.

Overall, the report aims to empower these stakeholders to effectively share and manage data while adhering to established governance frameworks.

2 OVERVIEW

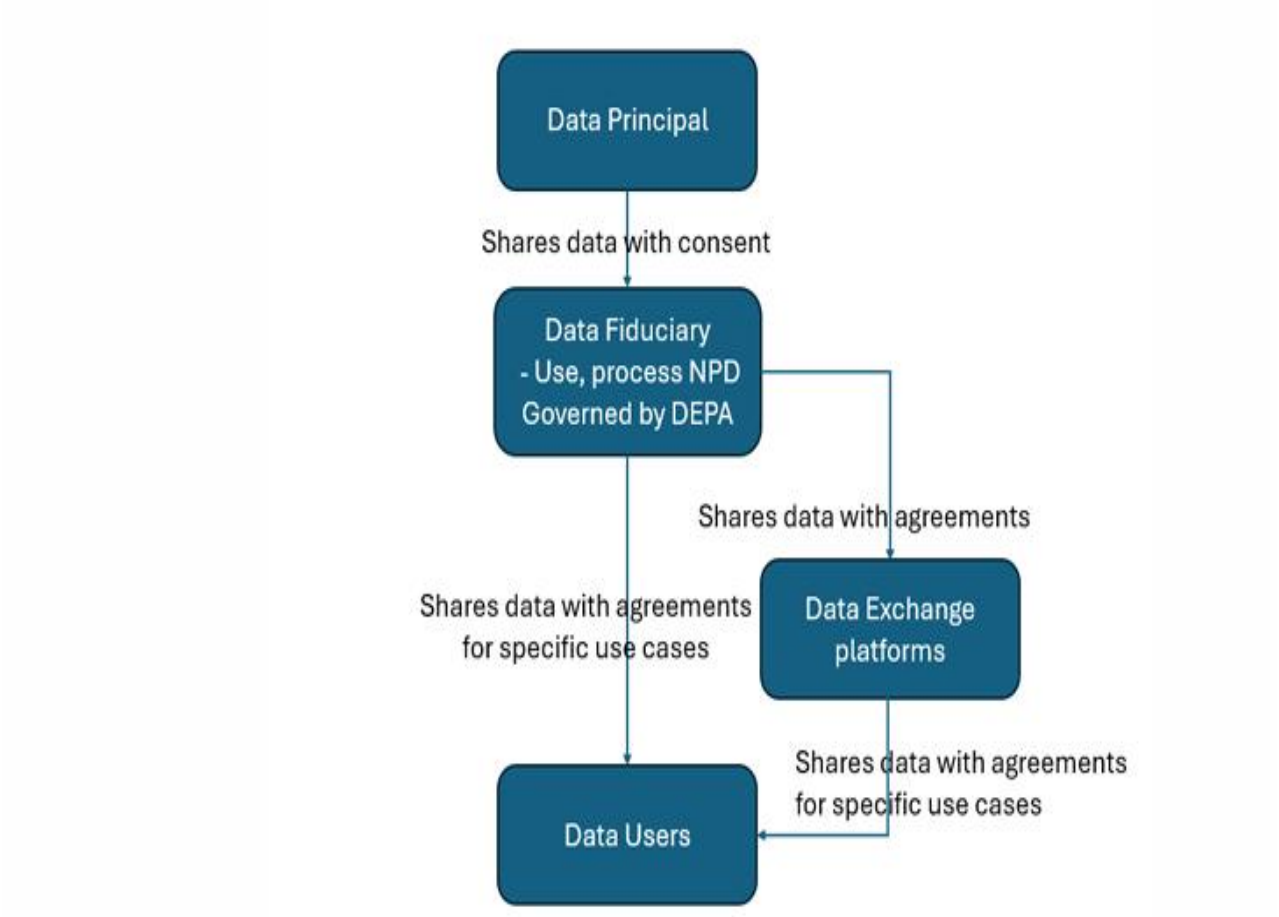
The report establishes guidelines for governance of Non-Personal Data (NPD) in India and details the flow of data as well as the processes involved. Data flow from public, private and government sectors to data users is different data based on their nature. While all government / public NPD other than high risk datasets can be shared with the public in the spirit of transparency, private sector and PSUs would be asked to share minimum data on a regular basis. Private companies would be encouraged to share data beyond the 'minimum data', in a partnership approach. Section 6.1 details the same. The following figure provides an overview of the flow of data discussed in this report.

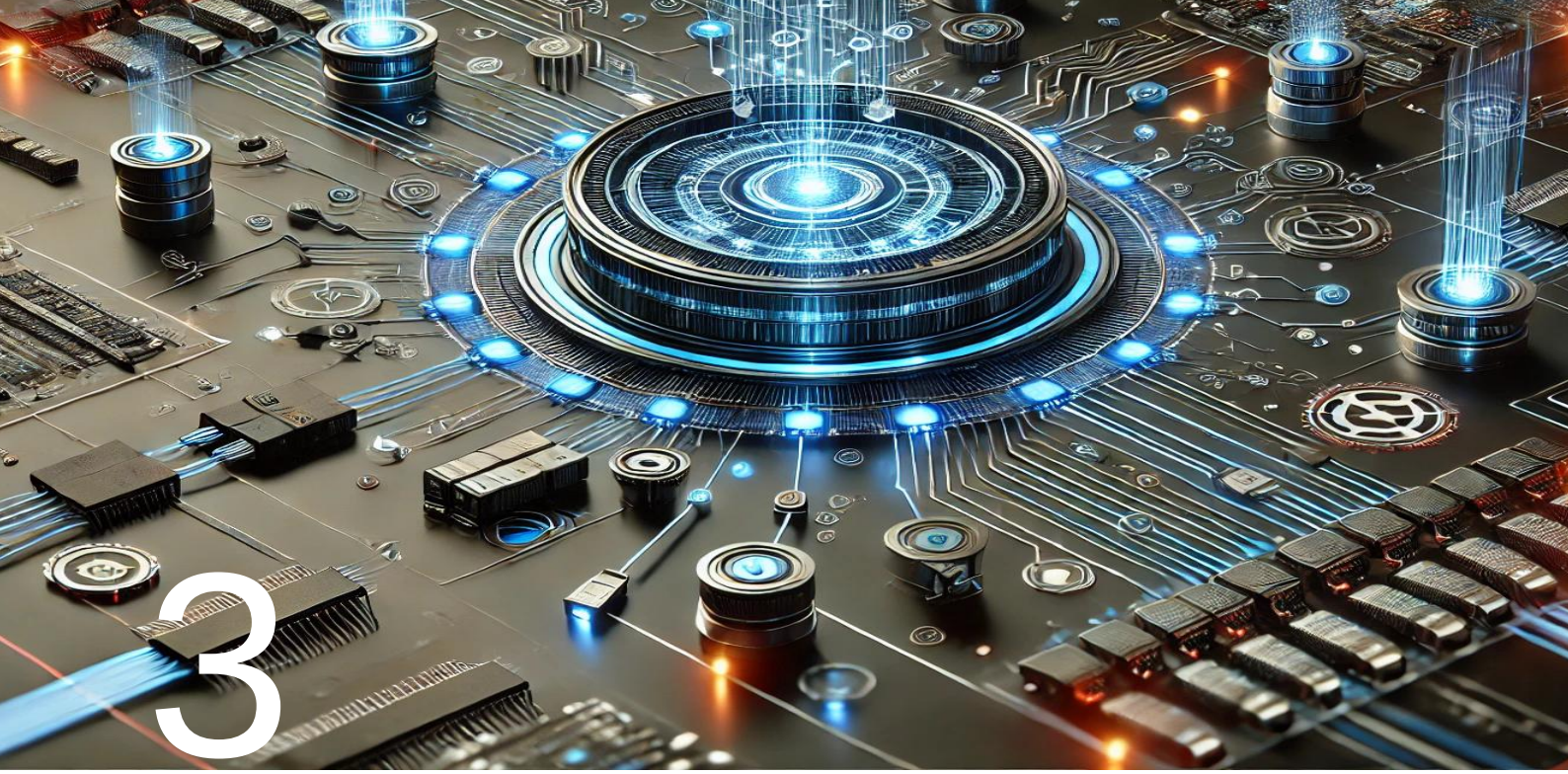
Figure1: Data Flow



Data is shared by Data Principal with consent. Data fiduciary signs agreements with the Data Principal to get consent and for the processing of NPD for the specified purposes following the DEPA architecture. They also sign agreements with data exchanges where the data fiduciary is not acting as the exchange. The agreements include conditions for providing data for exchange, as well as terms and conditions for processing for specified purposes. Data exchanges can also be data fiduciaries. However, not all data fiduciaries are data exchanges. Thus, in other words, the data fiduciary, as the custodian of the data, will sign separate agreements with the specific stakeholders directly involved with the data sharing. Processes involved in data sharing are detailed in section 6.3 and figure 2.

Figure 2: Process Flow





3

INTRODUCTION

The Draft National Data Governance Framework Policy (NDGFP) 2022[1] by the Ministry of Electronics and Information Technology (MeitY) emphasises sharing of Non-Personal Data (NPD) for building a repository of India-specific datasets. The NDGFP 2022 outlines a governance framework that addresses mechanisms for regulating NPD. The framework puts forth the main governance domains, namely, Data Access and Disclosure, Data Quality and Standards, Usage Rights, User Charges and Pricing, Ethical and Fair Use of Data, Policy Monitoring and Enforcement, Data Security and Privacy, and Redressal Mechanisms.

This document elaborates the processes and procedures for each of the governance domains as identified by NDGFP 2022. This document refrains from addressing the issues of data security and privacy as it mainly entails a technological approach that must be investigated through a techno-legal lens. Moreover, Digital Personal Data Protection Act (DPDP), 2023 already has detailed these issues effectively in case of personal data. The guidelines on these issues could be borrowed from DPDP Act, 2023 and can be adapted to suit non-personal data.

[1]https://www.meity.gov.in/writereaddata/files/National-Data-Governance-Framework-Policy.pdf

4

AIM AND OBJECTIVES

The aims and objectives of the guidelines are:

- **Elaborate** on the existing data governance framework for India
- **Enable** seamless data sharing and ethical use of data across sectors and stakeholders
- **Empowe**r Data Principal, data fiduciary, data processors and data consumers to share and use data

APPLICABILITY OF THE GUIDELINES

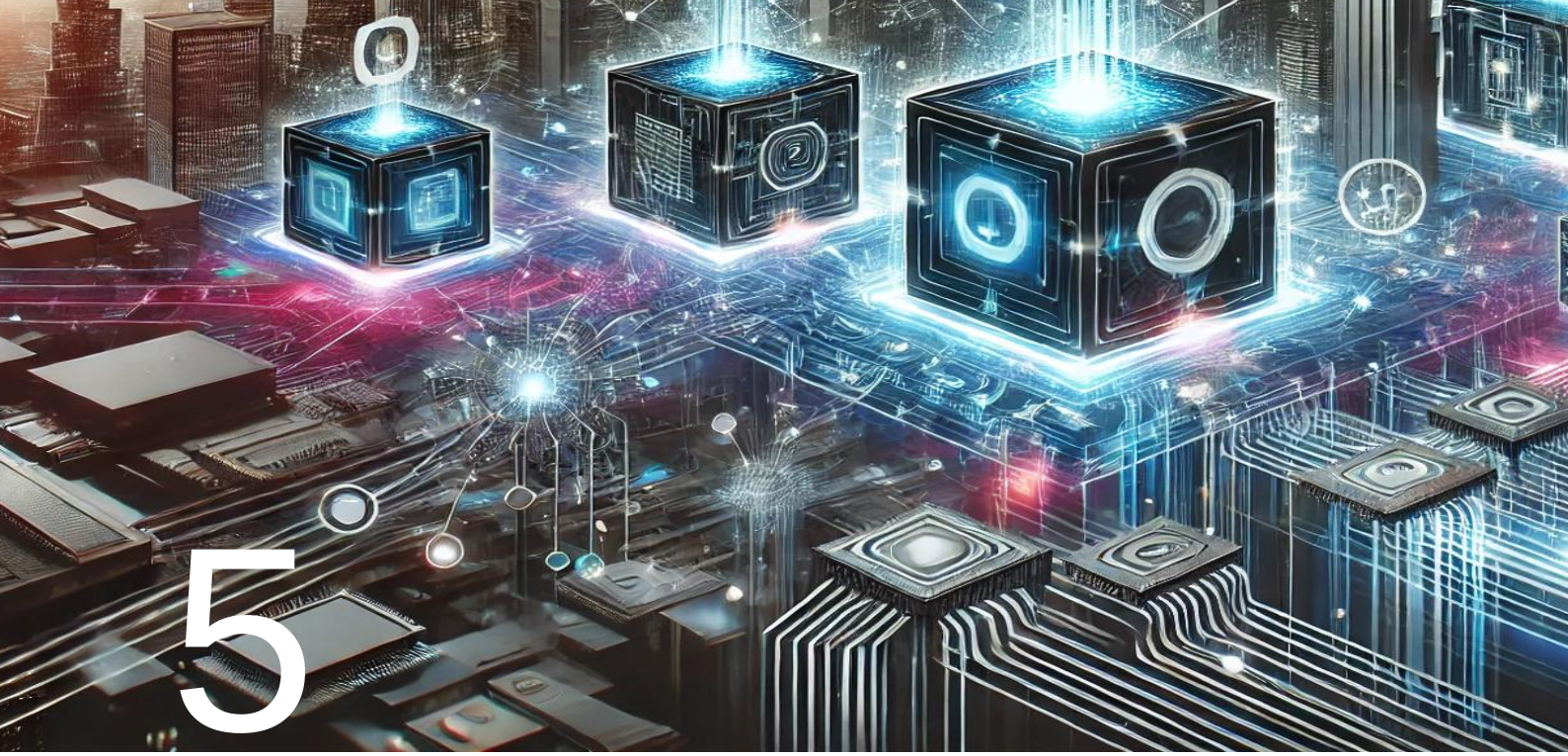
The guidelines are applicable to all the entities and individuals involved in:

Data Collection: Applies to private companies, individuals, government departments, research projects, policy think tanks, and educational institutions.

Data Management: Relevant for exchange platforms, data fiduciaries, government entities consuming public data, data storage and service providers

Data Sharing: Involves exchange platforms, data principals, and data fiduciaries engaged in sharing data.

Data Application: Pertains to application developers, data processors, data analysts, and use of NPD to train models.



5

GUIDING PRINCIPLES

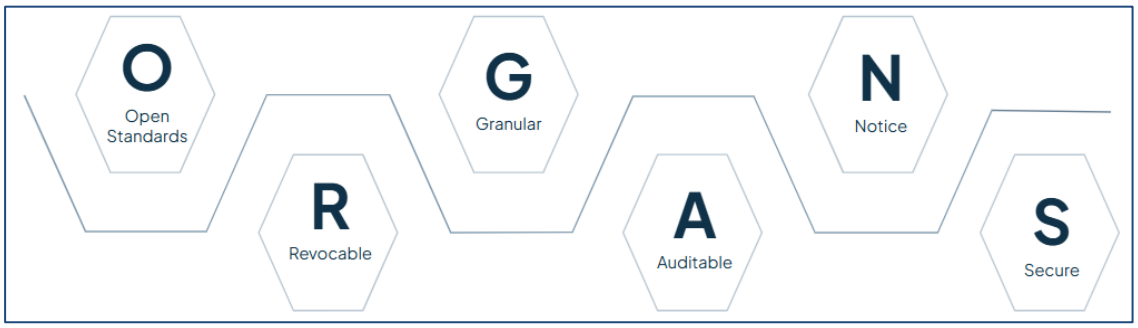
Digital public infrastructure (DPI) is a major enabler of digital transformation and contributes to better public service delivery at scale. It is an approach to solving socio-economic problems at scale, by combining minimalist technology interventions, public-private governance, and vibrant market innovation. DPI refers to blocks or platforms such as digital identification, payment infrastructure and data exchange solutions that help countries deliver essential services to their people, empowering citizens and improving lives by enabling digital inclusion[2]. India, through India Stack, became the first country to develop all three foundational DPIs, Digital identity (Aadhar), Real-time fast payment (UPI) and Account Aggregator built on the Data Empowerment Protection Architecture (DEPA). DEPA implements a digital consent artefact through which data principals can provide their consent to individual data transfer requests. Open standards - the consent standard is published and designed to operate as an open standard ensuring that all institutions have the same approach to consent and use it interoperability. Each such consent request must be informed, specific, granular, and Revocable by the data subject providing it. DEPA uses an electronic consent artefact that implements the ORGANS principles. This framework follows DEPA's ORGANS principles:

- Revocable - the consent is designed to be revocable at any point in time by the data subject who provided it.
- Granular - consent needs to be provided each time data is shared data as it specifies what data has been requested, how long it will be retained and who will process it.
- Auditable - records of all consent provided by a data subject can be retained in machine readable logs.

[2] https://economictimes.indiatimes.com/tech/technology/indias-digital-public-infra-central-to-its-goal-of-becoming-5-trillion-economy-emphasis-ventures-ceo/articleshow/99115056.cms?from=mdr

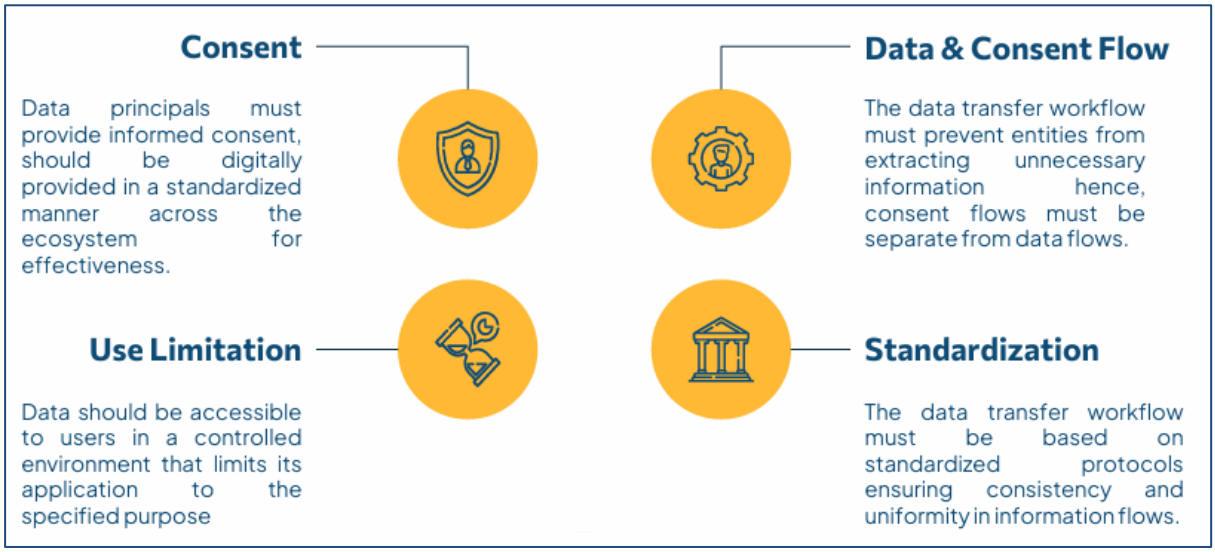
- Notice - the consent will provide data subjects with notice of the purpose to which it will be put, the parties who will process it and the duration for which it will be retained.
- Secure - the digital consent artefact is secure by design.

Figure 3: DEPA's ORGANS principles



DEPA creates a digital framework that allows users to share their data on their own terms through a third-party entity, who are known as Consent Managers. With the success of Aadhar and UPI, the lessons learned from these two pillars can be brought to develop and improve the third pillar. IndiaStack which is a set of APIs allows governments, businesses, startups, and developers to utilise a unique digital Infrastructure to solve India's hard problems towards presence-less, paperless, and cashless service delivery. Figure 4 details the key elements of DEPA implementation

Figure 4: Key elements of DEPA implementation

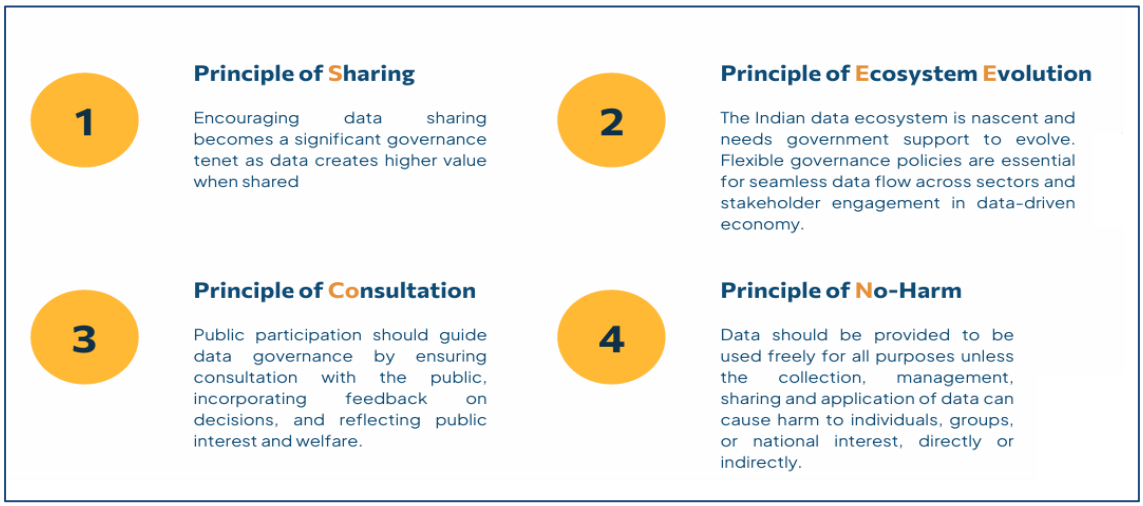


Governments and public entities all over the world are looking to increase their governance capacity and effectiveness through seamless and secure access to data. Private entities are looking to use the same in the commercial sphere. For seamless and secure flow of data throughout the data life cycle, the processes to be followed and the technology to be used should be well defined. Hence, the guidelines abide by the following principles (SEECoN):

- **Principle of Sharing**
- **Principle of Ecosystem evolution**
- **Principle of Consultation**
- **Principle of No-Harm**

These are detailed in figure 5.

Figure 5: SEECoN Principles

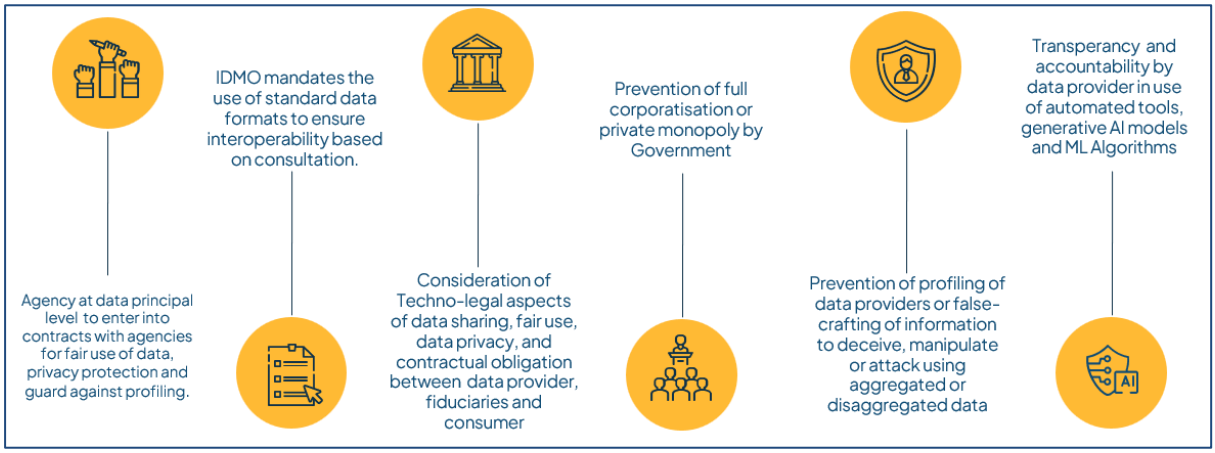


The guiding principles operate at two levels: the framework and the actual data. The framework needs to be consulted with, and for the data, consent needs to be taken. For addressing this, the following is proposed:

- **Agency at data principal level**
- **Mandate the use of standard data formats to ensure interoperability**
- **Techno-legal aspects of data sharing, data privacy, and other areas must be considered**
- **Prevention of corporatisation or private monopoly**
- **Prevention of profiling of data providers or false crafting information**
- **Transparency and accountability by data provider**

These are detailed in the figure below.

Figure 6: Proposed Actions



6 FRAMEWORK

The following sections outline the guidelines for data access, data disclosure norms, data quality and standards, user charges/pricing, usage rights, ethics and fair use of data and data monitoring and security.

6.1 DATA ACCESS AND SHARING

Access to different datasets can be based on three parameters, namely, ensure national security, galvanize innovation and prevent misuse/ abuse of data.

- *For national security protection*, the government can identify and control high-risk datasets, for example, data related to defence, nuclear installations, and critical energy infrastructure. In other words, the government therefore identifies NPD that cannot be made available for public use.
- *To galvanize innovation,* the government can define specific rules for the **flow of NPD from different agencies, namely, private, public, and government.** The government, through the IDMO, can facilitate data sharing as follows:
    - All government / public NPD can be shared with the public in the spirit of transparency.
    - All government funded projects collecting data should share the NPD collected during the course of the project in digitalized form (including biological and chemical data). The data will remain in exclusive ownership with the principal/project investigator(s) (PI) for 5 years, after which the data can be released to the public.
    - Private companies would be asked to share 'minimum data' collected by them on a regular basis. 'Minimum data' for each sector would be decided by the government and the sector representatives jointly, through a multi-stakeholder committee, under the IDMO. Thus, IDMO finalizes sector wise 'minimum data' to be shared by the private companies/agencies.

- IDMO would consult with the regulators/ nodal agencies/ leading participants in each of the sectors to provide what they consider to be the minimum data. For example, the Telecom Regulatory Authority of India (TRAI) can play a role in determining what is considered minimum data in the telecom sector. Similarly, the Ministry of Civil Aviation may be requested the minimum data that the industry can share. Furthermore, nodal agencies can determine the minimum required data in a given sector after discussing with other relevant stakeholders. For example, the National Highways Authority of India (NHAI) can collaborate with other relevant private entities to establish the minimum data requirements for roads and highways.
- In summary, a minimum amount of data shall be reported by all organizations regardless of their ownership.
    a. IDMO should come up with sector wise templates for enabling standardisation across entities. The templates should be easily accessible.
    b. IDMO should mandate all agencies to provide information about the 'minimum data' on their website.
    c. In case the agency does not have a website then the information should be passed on to IDMO to share on their website.

- Private companies/agencies should be allowed/encouraged to share data beyond the 'minimum data', in a partnership approach. The additional data (can be referred as 'extended data') that the private organizations want to allow access to or choose to report can be done under specific partnership agreements with the consumer.
- Additionally, data consumers should be allowed to seek data from any data fiduciary, whether public, government or private.
- *To prevent misuse of data for public harm* and ensure compliance, India Data Management Office (IDMO), under the Ministry of Electronics and Information Technology (Meity), Government of India can be the nodal agency to ensure compliance. In other words, IDMO can ensure privacy laws and other legal requirements are adhered to while data sharing. In such case IDMO should create a compliance template for parties involved the form of a data sharing agreement (buyer agreement, seller agreement, fair use agreement and similar).
- The relevant agreements for data sharing, for example, buyer agreement, seller agreement, the authenticity of data agreement, continuity agreement for a guaranteed flow of data, fair use agreement and similar (Annexure 1), should be provided by the IDMO.
- Services should be classified into essential (Directive (EU) 2022/2556 of the European Parliament and of the Council of 14 December 2022 on the Resilience of Critical Entities and Repealing Council Directive 2008/114/EC.) and non-essential activities. Contracts should be devised in such a manner that the state has the first right to data on activities that fall under essential services. Even in scenarios where (private) third parties are collecting data, the data ownership and first right are with the government. The third party can have access to data and can be asked to provide additional services other than collecting data if the contract is an outcome-based one. In scenarios where the state does not require third parties to provide

additional services, the third party need not have access to the data collected. The contracts or agreements should be based on the principles of fair use, privacy, and protection against profiling.

6.1.1. Data Exchange Process –Support from Contract Laws, MOSIP, and DEPA

Some of the existing laws and structures, like law for personal digital data protection, contract law for seller-buyer agreements, and Modular Open-Source Identity Platform (MOSIP) can provide a broad framework for managing the exchange of data. MOSIP is an open-source platform meant for governments or international organisations to build a foundational identification system in a cost-effective way. A functional identity system enables individuals to get a unique identity from the government to avail various services such as financial, social security, etc. Nations can use the platform when they want to build their own identification system. It provides a vendor-neutral and interoperable approach allowing governments to configure their systems with high accuracy. Apart from that, the platform gives ways to address various challenges when building a national functional system that helps meet the essential needs. The architectural principles of MOSIP can help frame the architecture for data exchange.

For example, the MOSIP architecture and Data Empowerment and Protection Architecture (DEPA) can guide the data exchange architecture to ensure privacy and data security. MOSIP platform is available with security and privacy features that will help protect the data from potential threats. The consent framework in the platform takes care of user privacy that lets users choose what they want to share with who and when. Apart from that, it enables users to lock authentication features that pave the ways to reach the next levels. The platform makes feasible methods to encrypt all the information that is inaccessible by both external and internal parties without user consent. Some other security features include license keys, policies, and infrastructure security that will help minimise potential risks. Thus, the MOSIP architecture can ensure privacy, data security and prevent occurrence of hacking and fraud.

The contract law can form the basis for data sharing agreements. For example, the basic agreement for transaction (buying agreement and selling agreement) can be guided by the contract law. Personal data protection can be ensured through the DPDP Act 2023.

6.1.2. Data Breach

IDMO will set the rules and regulations for handling data breach. The following two rules can be established:

- Rules and procedures to notify concerned data principals and data fiduciaries in case of a digital data breach (to ensure security).
- Penalties and Charges to be applied to the agency which that has caused and/ or been subject of a data breach (to ensure accountability and privacy protection).

6.2 DATA DISCLOSURE NORMS

Data disclosure norms refer to rules formulated around notifying the IDMO on the data collected and data stored.

- All companies /agencies (private, public and government) should disclose information on **metadata** collected by them to the IDMO at regular intervals[3].
- The metadata should be published on the company websites to make it easily available to the public.
- Furthermore, all companies/agencies to put the 'minimum data' (Section 5.1) on their website separated out as follows:
    - free data and priced data.
    - sample data, uploaded in the IDMO provided template.

6.3 USAGE RIGHTS

The National Data Governance Framework Policy 2022[4] ensures data usage rights along with permissioned purposes to be with the **Data Principal**.

For the purpose of data sharing, the **usage rights/data sharing rights** for 'monetary compensation for exchange' to be with the fiduciaries.

IDMO should establish this distinction between rights among data principal and data fiduciary. IDMO template for data sharing (discussed in section 5.1) could also include clear mention of these rights.

The data fiduciary secures the data principal's consent for the use and processing of NPD for the defined purposes using the DEPA architecture. The data fiduciary acquires data sharing rights only after obtaining consent from the data principle in accordance with the DEPA architecture.

To prevent any discrepancies, data fiduciaries seeking monetary compensation for exchange should register with the IDMO.

- **Data Fiduciary** (individual, company, or community), therefore, to be the nodal agency for data exchange.

Data fiduciary can function as follows.

1. Sign agreements with the Data Principal to get consent and processing NPD for the specified purposes following the DEPA architecture.

[3] **Alternately,** all government agencies can share information about the nature of data collected to the IDMO. Private and public agencies can be directed to **store** the 'nature of data' (metadata) in a predetermined format including the use declared by the data principal and agreements signed, for the IDMO to ask for such information in case of national interest, dispute, breach, or public harm.

[4] https://www.meity.gov.in/writereaddata/files/National-Data-Governance-Framework-Policy.pdf

2. Interact and sign agreements with data users, including terms and conditions for processing data for specified purposes.
3. Sign agreements with data exchanges where the data fiduciary is not acting as the exchange. The agreements include conditions for providing data for exchange, as well as terms and conditions for processing for specified purposes. Data exchanges can also be data fiduciaries. However, not all data fiduciaries are data exchanges. Thus, in other words, the data fiduciary, as the custodian of the data, will sign separate agreements with the specific stakeholders directly involved with the data sharing.

- A government department or public sector organisation can be a data fiduciary. However, they should register themselves as data fiduciary with the IDMO.

6.4 USER CHARGES

The value of data and pricing is driven by several factors such as data collection cost, data quality, interoperability, value-added services, and similar factors.

- Two factors are used in this document to determine the pricing mechanism, profitability and fairness of access.

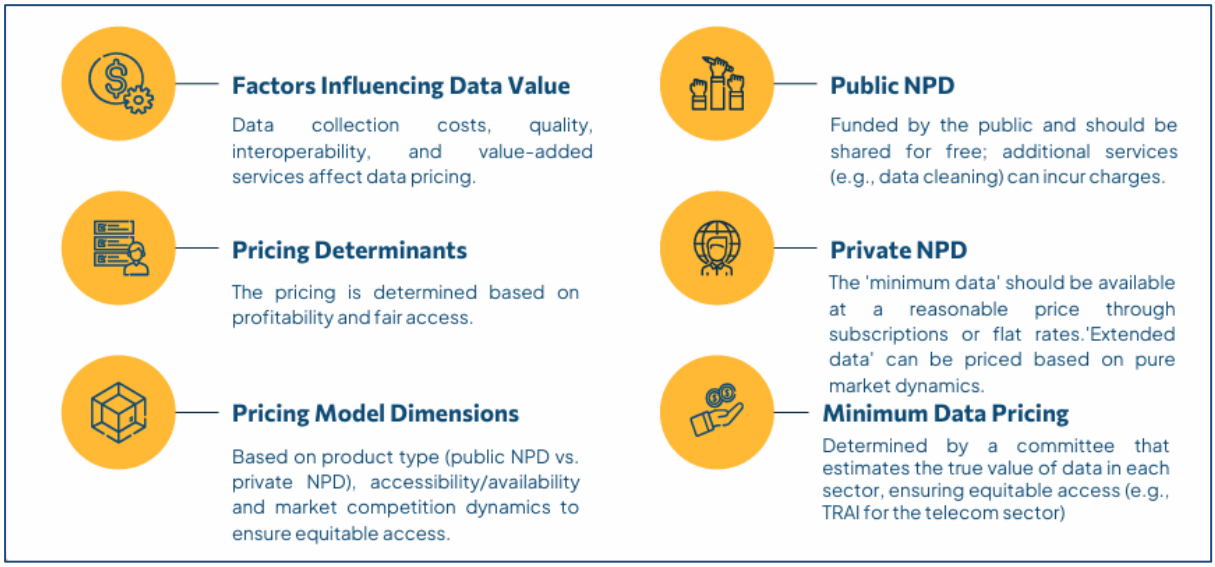
Table 1: Price models based on Type of Data and Equitable Access

Access	Type of Data	Pricing	Examples
Free Pricing	For data generated by public authorities or government agencies.	The public has indirectly paid for generation of such data and hence should be free.	Data on public utility like sanitation, water and electricity at aggregate level
On Demand Pricing	For data generated on a continuous/ periodic frequency.	Given the variable cost of continuously collecting and updating data is non-trivial and positive It has to be charged depending on quantum	Data on metrological conditions / alerts
Value Added Pricing	For data generated through public funded projects	Composite of free and premium services can be charged.	Data on Fin services, aggregate data to be made free and specific segment level data, priced appropriately.
Subscription/ period pricing	For 'minimum data' shared by private agencies	Price is determined by fixed price options.	One-time payment irrespective of scope of services, flat rate periodic subscription fees, and recurring prices for real-time datasets.
Dynamic Pricing	For 'extended data' shared by private agencies	Price is continuously determined by market dynamics like demand, data quality and other market parameters.	The cost of electricity varies based on real-time supply and demand conditions. This encourages consumers to adjust energy usage according to price fluctuations throughout the day, leading to more efficient energy consumption and reduced peak demand. It contributes to grid stability and minimizing the risk of blackouts during high-demand periods.

- The pricing model is based on 1) product type (public NPD/private NPD), 2) accessibility/availability, and 3) market dynamics of competition. The three dimensions will ensure marketplace galvanization without discrimination in access (unfair access).
- Public NPD is indirectly paid for by the public and should be shared with the public for free. Any additional services including data cleaning and customization on public NPD can be charged by the agency that undertakes the activity.
- In case of NPD from private agencies the 'minimum data' should be available at a reasonable price through subscriptions or flat rates. 'Extended data' can be priced based on pure market dynamics
- The pricing of 'minimum data' from private agencies can be determined by the committee that estimates the true value of data in each sector, ensuring equitable access (e.g., TRAI for the telecom sector) the 'minimum data' in each sector

The logic for the pricing outlined in is summarised in the figure below.

Figure 7: Usage charges



6.5 ETHICAL AND FAIR USE OF DATA

- The IDMO should define the principles for ethical and fair use of data (in other words, rules prohibiting unfair and illegal use). A committee with competent authorities can be set up for the same purpose.
- In the spirit of active consent, the data principal and the provider must be given a time frame by which they can revoke consent before publishing the data. The time frame could range from a few hours to a few days. The data gets published only after this time period if the data provider has not raised any complaints or requests for withdrawing their data.

- Consumer and public awareness campaigns should be run by the IDMO on ethical use of data and on regulations prohibiting unfair and illegal data use.
- IDMO should define penalties for unethical or unfair data use, with conflicts potentially resolved through the judicial system.
- Data Tribunal- A specialized body equipped with expertise solely for the purpose of adjudicating data related cases can be set up. It can either be a judicial body akin to the Environmental Tribunal or a quasi-judicial system like that of Central Electricity Commission. Any person seeking relief and compensation for damages due to data breaches/leak may approach the data tribunal.

6.6 MONITORING AND ENFORCEMENT

Three agencies should take responsibility for monitoring and enforcement.

- IDMO should play a regulatory/ oversight role of dealing with violations, appeals, complaints, and redressal mechanisms.
    - Should define the principles for the ethical and fair use of data.
    - Should monitor fair use, takes up internal investigations, and ensures enforcement of legalities.
    - Should mandate information sharing by organizations on the systems established for data security and data privacy protection.
- Data fiduciary to monitor data security and privacy procedures during data exchange and legal compliance in use of data and storage.
    - Data fiduciaries are responsible for managing data flow from the cloud to consumers as they are the agency responsible for interacting and signing agreements with the data principal. In scenarios where data exchanges are involved, this responsibility will fall on data exchanges.
    - Data security is well understood and there are technologies for the same. Data travels over the internet through a combination of wired and wireless connections. HTTP (hypertext protocol) protocol used for transferring data over the internet ensures data security, through data encryption.
    - However, data privacy is a complex issue, and it is close to impossible to keep data private while it is shared and used. However, there are methods that allow for sharing of data without loss of privacy. Differential privacy is one such practice that adjusts the contents of the dataset to conform to a certain acceptable threshold, ensuring that the dataset's usefulness and precision remain intact.
- Data exchange agency to take responsibility for internal monitoring and enforcement. The agency to follow appropriate internal data protection procedures, conduct regular internal audits

and undertake assessments of internal controls for safety and security, and adhering to legal provisions.

6.7 DATA QUALITY

Data Governance Quality Index (DGQI) Toolkit 2021, from the Government of India has provided a base for data quality and data standards.

Measures of quality vary based on nature of data sets, and purpose/use of the data. The most practical approach will be an amalgamation of data user perspective and data demand perspective. In other words, pre-established data standards (Step 1) and market dynamics-based standards (Step 2) can together form the measure of data quality, as follows:

- Basic IDMO defined: For consistency across sectors IDMO to mark the following data quality parameters to be calculated for all data. **These data quality parameters are source agnostic, sector agnostic, quantifiable and automatable.**
- Private companies sharing minimum and extended data should ensure relevance of the data published.
- Table 1 provides the quality dimensions with quantifiable indicators to determine the data quality from an exchange perspective.

Completeness and timeliness dimensions, and 3 quality indicators of accuracy, are quantifiable, sector agnostic, domain agnostic and automatable.

Table 2: List of dimensions and quantifiable indicators

Dimension	Indicator	Definition
Accuracy	Correctness	Error-free representation of data
	Precision	It can be represented by small quality
	No-duplication	Contains distinct data values
Completeness	Comprehensive	Availability of data satisfies the user's needs.
Timeliness	Currency	Sufficiently update for new or existing task

Source: Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for Data Quality Assessment and Improvement. ACM Computing Surveys, 41(3)

Figure 8: Data quality dimensions.

- IDMO will define the template through a downloadable quality assessment software. IDMO will develop a downloadable/ web-based Data Quality assessment software that companies and other data holders can use. The template acts like a certificate. In other words, companies upload files in the IDMO website to check data quality and receives results with scores indicating information on duplicates, completeness, format, precision etc in the form of a data quality report. This can be downloaded by companies and uploaded in the company website for users.
- The quality assessment software will be situated on the IDMO website for fiduciaries to use and get 'IDMO quality score' for their data. The 'IDMO quality score' is must for all data before it is shared. Thus, NPD will have consistency across sectors and agencies.
- The parameters defined above will ensure basic quality control, quality standardization and quality management across all sectors and data types.
- Market defined: Data agencies can additionally define other quality standards for their NPD datasets as per their requirement, for example, based on customer feedback, customer rating, relevance /usefulness and value addition.

Data Quality Dimensions

Accuracy Metric (AC):

- **Purpose:** Measures the degree to which data values accurately represent real-world situations.
- **Two Sub-Metrics:**
  - **Correctness Metric (A-C1):** Indicates the proportion of data points within a defined range interval, based on a reference dataset.
  - **Precision Metric (A-P2):** Measures data centering within a 95% confidence interval, enhancing prediction accuracy.

Timeliness Metric (TC):

- **Purpose:** Measures the uniformity of time intervals in a time series dataset, ensuring consistent data flow for smooth processing.

Completeness Metric (CM):

- **Purpose:** Measures the completeness of a dataset by assessing the proportion of missing elements, impacting research accuracy.
- **Types of Completeness:**
  - **Schema Completeness:** Presence of entities and attributes in a schema.
  - **Column Completeness:** Missing values in table columns.
  - **Population Completeness:** Missing data in relation to a reference population.
- **Global Completeness Metric (CM1):** Calculates the overall completeness, with a value between zero (all values missing) and one (no missing values).
- **Variable-Specific Completeness (CM2):** Assesses completeness of individual variables over time, also ranging from zero to one.

Uniqueness Metric (UM):

- **Purpose:** Evaluates the percentage of duplicate data in a dataset; a low duplicate rate indicates unique, efficient, and representative data, while a high rate may cause redundancy and bias in analysis.

Granularity Metric (GM):

- **Purpose:** Evaluates extent of granularity as opposed to aggregate nature of variable.

7

CONCLUSION

In recent years, the importance of data in the economy has become increasingly apparent. This document presents a framework to develop/ galvanize the data ecosystem. This report discusses possible governance processes and structures for data access, data disclosure norms, data quality and standards, user charges/pricing, usage rights, ethics and fair use of data, and data monitoring and security for NPD.

We are particularly concerned and interested in ways different data rights determine data use in the economy, and thus affect output, privacy, and public welfare. Thus, our recommendations take care that any policy allocation one considers may not limit the use of NPD by one user or another. For example, a policy that succeeds in generating market-controlled prices, may potentially limit the use of the data and also create inefficiency that arises from a nonrival input not being used at the appropriate scale. Furthermore, we consider policy structures where data owners and providers have the control to balance concerns about privacy against the economic gains that come from selling data to all interested parties. This equilibrium is essential. The equilibrium balances consumption and welfare. The SEECoN principles ensure the same. The principles, for example, talk of flexibility to allow for data to flow seamlessly while emphasizing the principle of consultation where consulting with the public and getting feedback on potential decisions forms a significant tenet. Furthermore, SEECoN applies the harm principle to data use, in other words, data should be provided to be used freely for all purposes unless the collection, management, sharing and application of data can cause harm to individuals, groups, or national interest, directly or indirectly. In such case, encouraging open access to certain datasets facilitates better public and private reach and can allow for better intervention by non-governmental organizations to support local economic and development projects.

In summary, the suggestions in the report can expand the possibilities for better, more informed decision making, and more efficient and data led service delivery. Given the urban ecosystem the report can stimulate research, innovation, and growth in the Indian Data and AI based

research. Improving data availability is absolutely critical as scarcity of data has been the primary hindrance in data led research, policy and governance. Making data available to third parties can create a virtuous cycle of development and data led research. The report aims to induce this change through regulation and a shift in data culture.

8

GLOSSARY & ABBREVIATIONS

- **AI Model:** Artificial intelligence (AI) model is a program that has been trained on a set of data to identify specific patterns or make decisions without the need for human intervention.
- **Consent Artefact:** Machine-readable electronic document that specifies the parameters and scope of data share that a user consents to in any data sharing transaction.
- **Consent Managers:** Third-party entities that facilitate user consent for data sharing, ensuring users can control how their data is used.
- **Data Breach:** An incident where unauthorized access to or disclosure of data occurs, requiring notification and accountability measures.
- **Data Exchange:** The process of securely passing encrypted information back and forth between trusted parties to ensure message integrity and prevent alterations during transit.
- **Data Fiduciary:** An organization or individual that manages data on behalf of a data principal, ensuring ethical handling and compliance with regulations.
- **Data Governance Framework:** A structured approach that outlines the processes, policies, and standards for managing data effectively and ethically.
- **Data Principal:** An individual or entity that owns the data and has the right to control its use which is also Data Subjects.
- **Data Processor:** An entity that processes data on behalf of a data fiduciary, typically involving storage, analysis, or transformation of data.
- **Data Set:** A data set is a collection of data that can be organized in a table, spreadsheet, database, or other format.
- **Data Subject:** An individual or entity that owns the data and has the right to control its use, which is also a Data Principal.
- **Digital Personal Data:** Personal data is any information which are related to an identified or identifiable natural person.
- **Digital Public Infrastructure (DPI):** A framework that enables digital transformation and public service delivery through technology, including systems like digital identification and payment infrastructure.
- **Extended Data:** Additional data that organizations may choose to share beyond the minimum requirements under specific partnership agreements.
- **Metadata:** Data that provides information about other data, including its characteristics and management details.

- **Minimum Data:** The essential amount of data that must be shared by organizations, as determined collaboratively by the government and sector representatives.
- **Non-Personal Data (NPD):** Data that does not identify an individual and can be shared without compromising personal privacy.

9

ANNEXURE

ANNEXURE 1: AGREEMENTS FOR DATA SHARING/ EXCHANGE

- Buyer Agreement: An agreement between the buyer and seller outlining the terms of data purchase.
- Seller Agreement: An agreement detailing the terms under which data is sold to a buyer.
- Authenticity Agreement: Certifies that a dataset is genuine, credible, and reliable.
- Fair Usage Agreement: Ensures that the data will be used fairly, preventing fraud and abuse; violations may incur charges.
- Additional Service Agreement: Specifies terms for collaborative data schemes or other services provided.
- Continuity Agreement: Confirms the time duration for which data sharing is guaranteed to the consumer.