Centre For Society And Policy

भारतीय विज्ञान संस्थान

IIMB

तेजस्वि नावधीतमस्तु

cdpg

Handbook on Data Management Policy Framework

Guidelines for Non Personal Data (NPD) Governance

Authors

Anjula Gurtoo

Professor

Indian Institute of Science (IISc)

Jyotirmoy Dutta

Specialist Scientist

Society for Innovation and Development (SID),

Indian Institute of Science (IISc)

Minnu Malieckal

Project Associate II

Centre for Society and Policy

Indian Institute of Science (IISc)

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

Table of Contents

About This Handbook

What is the purpose of this handbook?

The handbook 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 purpose 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.
Key objectives include:

- Elaboration of the existing data governance framework to enhance data accessibility and transparency.
- Empowering stakeholders such as data principals and processors to share and utilize data effectively.
- Establishing mechanisms for monitoring compliance and addressing potential misuse of data
- Ensuring Transparency and accountability in application of automated tools

Who is this handbook for?

This handbook 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 Collectors: Including private companies, government departments, research projects, and educational institutions that gather data.
- Data Managers: Entities like exchange platforms and data fiduciaries responsible for storing and managing data.
- Data Sharers: Organizations and individuals involved in the sharing of data, including data principals and fiduciaries.
- Data Users: Application developers, data analysts, and those involved in machine learning operations who utilize the data for various applications.

Overall, the handbook aims to empower all these parties to effectively share and manage data while adhering to established governance frameworks

How can you use this handbook?

You can effectively use this handbook, follow these steps:

1. Understand the Framework: Familiarize yourself with the governance framework outlined covering Data Access, Quality Standards, use of AI tools and Ethical Use.
2. Identify Your Role: Position yourself within the data ecosystem
3. Follow Guidelines for Data Sharing: Adhere to the established protocols for sharing Non-Personal Data (NPD).
4. Utilize Templates and Agreements: Use the provided templates for data sharing agreements and compliance
5. Engage in Continuous Learning: Stay updated on any revisions or additional guidelines by IDMO

By following these steps, stakeholders can navigate the complexities of NPD governance effectively.

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 are 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.

Chapter 1

Introduction

Key Considerations

1.1. Summary

1.2. Aim & Objectives

1.1 Summary

- NDGFP 2022 emphasizes the sharing of Non-Personal Data (NPD) to build a repository of India-specific datasets.
- Governance Framework detailed in this document outlines mechanisms for regulating NPD, focusing on:


Data Access and Disclosure


User Charges and Pricing


Data Quality and Standards


Ethical and Fair Use of Data


Usage Rights


Monitoring and Enforcement

- **Focus Areas:**
  a. This document elaborates on processes for each governance domain identified by the NDGFP 2022.
  b. It does not address data security and privacy, as these are covered under the Digital Personal Data Protection Act (DPDP) 2023.

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

1.2 Aim & Objectives

Elaborate

Elaborate on the existing data governance framework for India

Enable

Enable seamless data sharing and ethical use of data across sectors and stakeholders

Empower

Empower data principal, data fiduciary, data processors and data consumers to share and use data

Chapter 2

Applicability of the Guidelines

2.1 Applicability of the Guidelines

Data Collection

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

Data Sharing

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

Data Management

Relevant for exchange platforms, data fiduciaries, government entities consuming public data, data storage and service providers.

Data Application

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

Chapter 3

The Guiding Principles

Key Considerations

3.1. Data Empowerment Protection Architecture (DEPA)

3.2. Key Elements of DEPA Implementation

3.3. DEPA's ORGANS Principles

3.4. SEECoN Principles

3.5. Proposed Actions

3.1 Data Empowerment Protection Architecture (DEPA)

- DEPA is a techno-legal framework that empowers individuals to control their personal data transfers, enhancing agency over privacy rights through a technology interface.
- Open Standards, Revocable Consent, Granular Consent, Auditable, Notice Requirement, and Secure Design ensure user control and transparency.
- India Stack integrates Aadhaar, UPI, and Account Aggregator under the Data Empowerment Protection Architecture (DEPA).

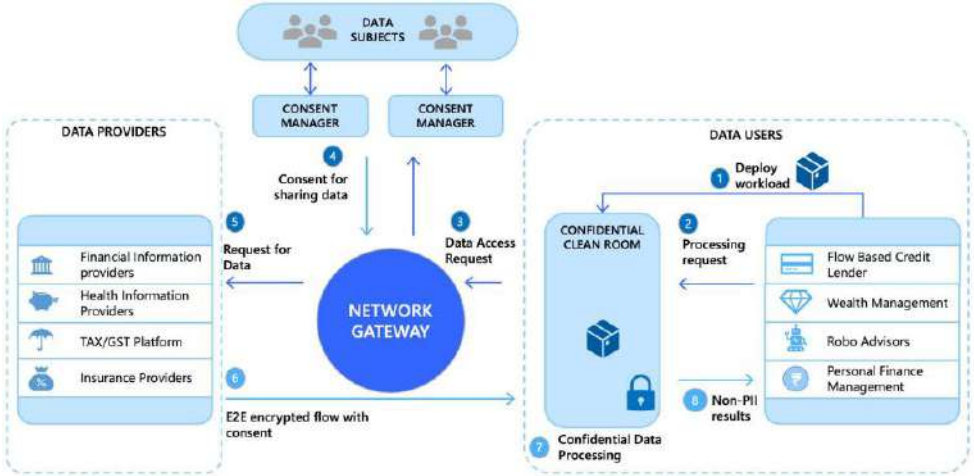


Fig.1. DEPA Workflow

Source: Aapti Institute: Building Open Networks: A Handbook

3.2 Key Elements of DEPA Implementation

Consent

Data principals must provide informed consent, should be digitally provided in a standardized manner across the ecosystem for effectiveness.

Data & Consent Flow

The data transfer workflow must prevent entities from extracting unnecessary information hence, consent flows must be separate from data flows.

Use Limitation

Data should be accessible to users in a controlled environment that limits its application to the specified purpose

Standardization

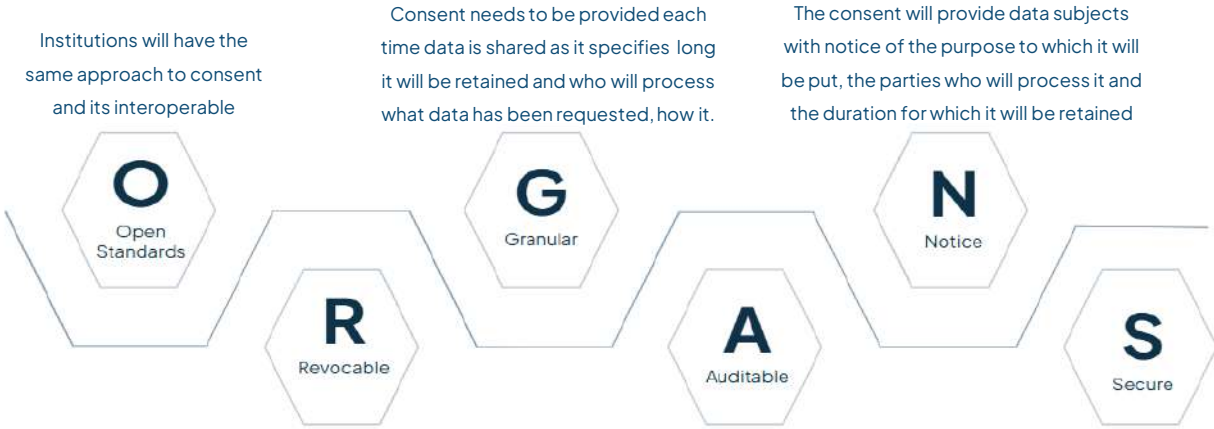
The data transfer workflow must be based on standardized protocols ensuring consistency and uniformity in information flows.

Source: https: ORGANS Principles-//depa.world/learn/about-depa/organs-principles

3.3 DEPA's ORGANS Principles

- DEPA implements a digital consent artefact through which data principals can provide their consent to individual data transfer requests. 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. These principles can be summarised as under:



Source: https: ORGANS Principles-//depa.world/learn/about-depa/organs-principles

3.4 SEECoN Principles

1

Principle of Sharing

Encouraging data sharing becomes a significant governance tenet as data creates higher value when shared

2

Principle of Ecosystem Evolution

The Indian data ecosystem is nascent and needs government support to evolve. Flexible governance policies are essential for seamless data flow across sectors and stakeholder engagement in data-driven economy.

3

Principle of Consultation

Public participation should guide data governance by ensuring consultation with the public, incorporating feedback on decisions, and reflecting public interest and welfare.

4

Principle of No-Harm

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.

3.5 Proposed Actions

Agency at data principal level to enter into contracts with agencies for fair use of data, privacy protection and guard against profiling.

IDMO mandates the use of standard data formats to ensure interoperability based on consultation. Refer Annexure 2

Consideration of Techno-legal aspects of data sharing, fair use, data privacy, and contractual obligation between data provider, fiduciaries and consumer

Prevention of full corporatisation or private monopoly by Government

Prevention of profiling of data providers or false-crafting of information to deceive, manipulate or attack using aggregated or disaggregated data

Transperancy and accountability by data provider in use of automated tools, generative AI models and ML Algorithms

Chapter 4

Understanding The Framework

Key Considerations

4.1 Data Access & Sharing

4.2 Data Disclosure Norms

4.3 Usage Rights

4.4 User Charges

4.5 Ethical & Fair Use of data

4.6 Monitoring & Enforcement

4.7 Data Quality

4.1 Data Access & Sharing (1/2)

Access to datasets is governed by following three key parameters: ensuring national security, preventing data misuse, and Galvanizing innovation. Refer Annexure 2.

- **National Security:**
  - The government will identify and control high-risk datasets, such as those related to defense and critical infrastructure, which will not be publicly accessible.

- **Preventing Misuse:**
  - The IDMO will serve as the nodal agency to ensure compliance with privacy laws during data sharing, creating templates for various data-sharing agreements (e.g., buyer and seller agreements).
  - Services will be classified into essential and non-essential categories, with the government retaining the first right to data related to essential services.
  - Contracts must adhere to principles of fair use, privacy, and protection against profiling.

- **Galvanizing Innovation:**
  - The government will establish rules for the flow of non-public data from various sectors.
  - Transparency: Government/public NPD can be shared with the public.
  - Data from Funded Projects: Data collected from government-funded projects must be shared digitally after five years of exclusive ownership by the principal investigator.
  - Minimum Data Sharing: Private companies will be required to share a defined "minimum data" set regularly, determined collaboratively by the government and sector representatives through the India Data Management Office (IDMO).
  - Consultation with Regulators: IDMO will work with relevant agencies to define minimum data requirements for different sectors.

4.1 Data Access & Sharing (2/2)

- **Data Exchange Process Supported by Contract Laws, MOSIP, and DEPA**

  The data exchange process can be effectively managed using existing frameworks such as personal digital data protection laws, contract law, and the Modular Open-Source Identity Platform (MOSIP) Architecture. Key points include:

  - **MOSIP Architecture:**
    - An open-source platform designed for governments to create foundational identification systems cost-effectively.
    - Provides a unique identity to individuals for accessing various services, ensuring vendor neutrality and interoperability.
  - **Data Exchange Architecture:**
    - MOSIP architecture, along with the Data Empowerment and Protection Architecture (DEPA), guides secure data exchange while ensuring privacy.
    - Features include user-controlled consent for data sharing, encryption of sensitive information, and robust security measures to prevent data breaches.
  - **Contract Law:**
    - Forms the basis for data sharing agreements, ensuring personal data protection through the Digital Personal Data Protection (DPDP) Act 2023.
- **Data Breach Management:**
  - The India Data Management Office (IDMO) will establish rules for notifying affected parties in the event of a data breach and impose penalties on responsible agencies to ensure accountability.

Source: https://docs.mosip.io/1.1.5/architecture/mosip-architecture; https://depa.world/learn/about-depa/an-introduction-to-depa

4.2 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.
- 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 to be uploaded using the IDMO provided template.

4.3 Usage Rights

- **Data Usage Rights:**
  - Data Principal retains usage rights while data fiduciaries hold rights for monetary compensation in data exchanges.
  - IDMO will clarify these distinctions in data sharing templates.
- **Consent and Processing:**
  - Data fiduciaries must secure consent from data principals for processing non-public data using DEPA architecture.
  - Data fiduciaries seeking monetary compensation must register with IDMO.
- **Roles of Data Fiduciaries:**
  - Sign agreements with data principals for consent and processing NPD.
  - Interact with data users and establish agreements for specified data processing purposes.
  - Sign agreements with data exchanges, outlining conditions for data provision and processing.
- **Government Involvement:**
  - Government departments can act as data fiduciaries but must register with IDMO.
- **IDMO Responsibilities:**
  - Establish rules for handling data breaches, including notification procedures and penalties for breaches.
  - Ensure compliance with privacy protection laws and manage the overall framework for data sharing.

4.4 Usage Charges (1/2)

Factors Influencing Data Value

Data collection costs, quality, interoperability, and value-added services affect data pricing.

Pricing Determinants

The pricing is determined based on profitability and fair access.

Pricing Model Dimensions

Based on product type (public NPD vs. private NPD), accessibility/availability and market competition dynamics to ensure equitable access.

Public NPD

Funded by the public and should be shared for free; additional services (e.g., data cleaning) can incur charges.

Private NPD

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.

Minimum Data Pricing

Determined by a committee that estimates the true value of data in each sector, ensuring equitable access (e.g., TRAI for the telecom sector)

4.4 Usage Charges (2/2)

Access	Type of Data	Pricing	Examples
Free Pricing	For data generated by public authorities or government agencies.	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 encourage 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.

4.5 Ethical & Fair Use of Data

- **Principles for Ethical Data Use:**
  - IDMO should establish rules to prohibit unfair and illegal data use, supported by a committee of competent authorities.
- **Active Consent Framework:**
  - Data principles and providers must have a defined time frame (hours to days) to revoke consent before data publication; data is published only if no complaints are raised within this period.
- **Awareness Campaigns:**
  - IDMO should conduct consumer and public awareness campaigns on ethical data use and regulations against unfair practices.
- **Penalties for Misuse (Abuse):**
  - IDMO should define penalties for unethical or unfair data use, with conflicts potentially resolved through the judicial system.
- **Data Tribunal Establishment:**
  - A specialized body, either judicial or quasi-judicial, should be created to adjudicate data-related cases, allowing individuals to seek relief and compensation for damages from data breaches or leaks.

4.6 Monitoring & Enforcement

Three distinct agencies should oversee monitoring and enforcement of data usage regulations.

India Data Management Office (IDMO)

Plays a regulatory and oversight role for violations, appeals, complaints, and redress mechanisms.Monitors compliance with fair use, conducts internal investigations, and ensures legal enforcement.

Mandates organizations to share information about their data security and privacy protection systems.

Data Fiduciaries

Responsible for monitoring data security and privacy during exchanges and ensuring legal compliance in data use and storage.

Manage data flow from the cloud to consumers and interact with data principals; responsibility shifts to data exchanges when involved.

Data Exchange Agency

Responsible for internal monitoring and enforcement of data protection procedures.

Conducts regular internal audits and assessments to ensure safety, security, and compliance with legal provisions.

4.7 Data Quality (1/4)

- 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**:
  - Pre-established standards + Market-based standards.
  - Combines user and demand perspectives.
- **IDMO Parameters for Quality**:
  - Consistent across sectors, source-agnostic, quantifiable, and automatable.
- **Key Data Quality Dimensions**:
  - **Accuracy**: Error-free, precise, and unique data values.
  - **Completeness**: All required data elements are available.
  - **Timeliness**: Data is current and up-to-date.
  - **Uniqueness**: Ensures minimal duplication.

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

Table 1: List of dimensions and quantifiable indicators

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

4.7 Data Quality: Understanding Metrics (2/4)

1. **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.

2. **Timeliness Metric (TC)**:

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

3. **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.

4.7 Data Quality: Understanding Metrics (3/4)

4. 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.

5. Granularity Metric (GM):

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

4.7 Data Quality Assessment and Standards (4/4)

6. IDMO's Quality Assessment Software:

Developed by IDMO as a downloadable or web-based tool, enabling companies to assess data quality metrics (e.g., duplicates, completeness, precision).

Provides a report, the "IDMO quality score," which can be displayed by companies to ensure data quality and consistency across sectors.

7. Quality Standards:

IDMO's standards promote data quality control and standardization across industries. Agencies may further customize quality metrics based on needs like customer feedback and relevance

Chapter 5

Conclusion

5.1 Conclusion

Growing Importance of Data

The significance of data in the economy has become increasingly clear, necessitating a robust framework to enhance the data ecosystem

SEECON Principles

The principles of Sharing, Ecosystem Evolution, Consultation, and No-Harm guide the framework, encouraging public engagement and safeguarding against misuse.

Governance Framework

This report outlines governance processes for data access, disclosure norms, quality standards, pricing, usage rights, and ethical considerations for Non-Personal Data (NPD).

Facilitating Innovation

Improved data availability is crucial for stimulating research, innovation, and development in India's data-driven economy.

Balancing Data Rights

Recommendations focus on ensuring that data rights do not restrict NPD use, promoting efficiency and preventing monopolistic practices while maintaining privacy and public welfare

Call to Action

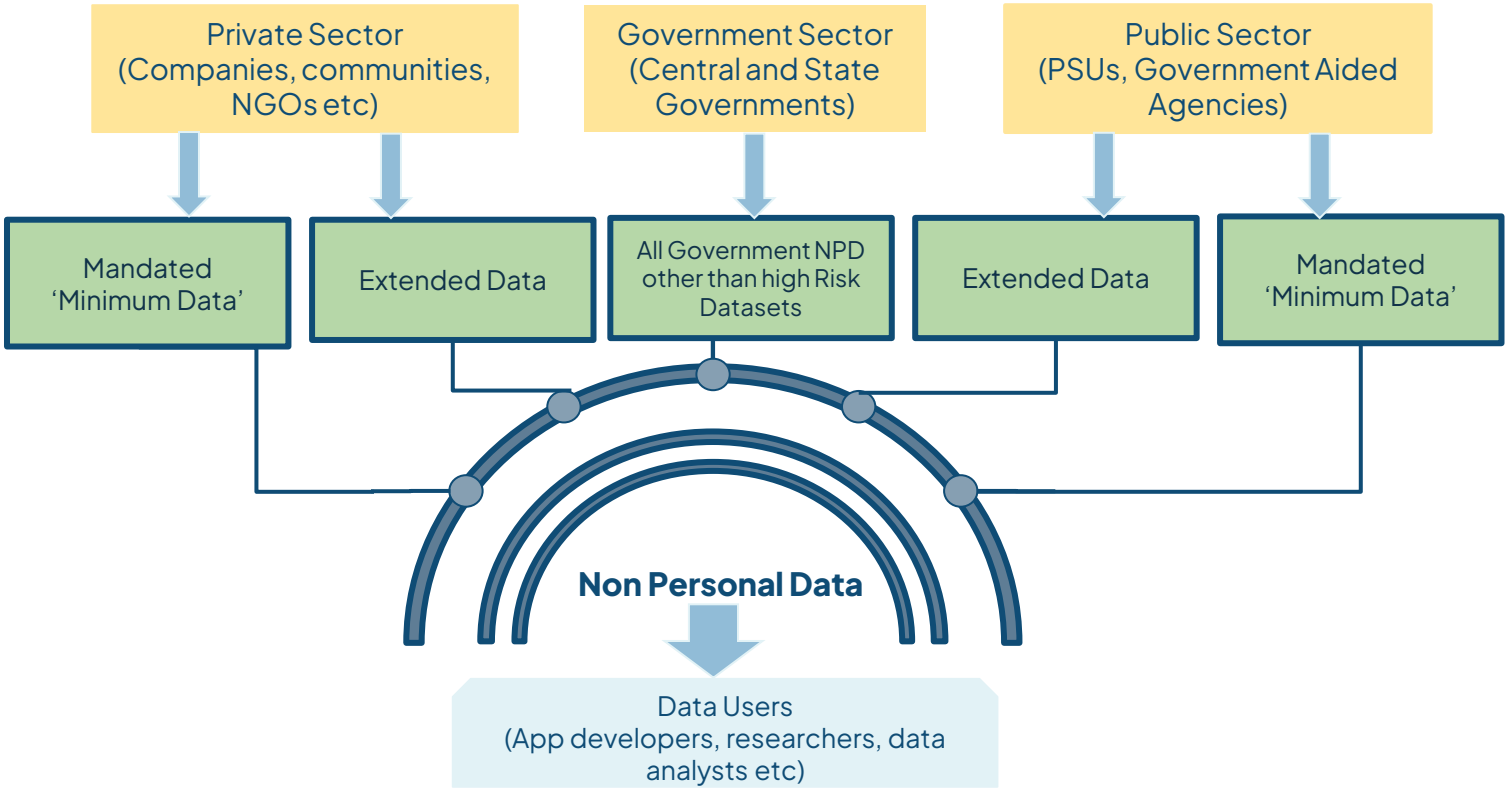
The report advocates for a cultural shift in data management through proactive regulation, fostering an environment conducive to informed decision-making and enhanced service delivery.

Annexure 1

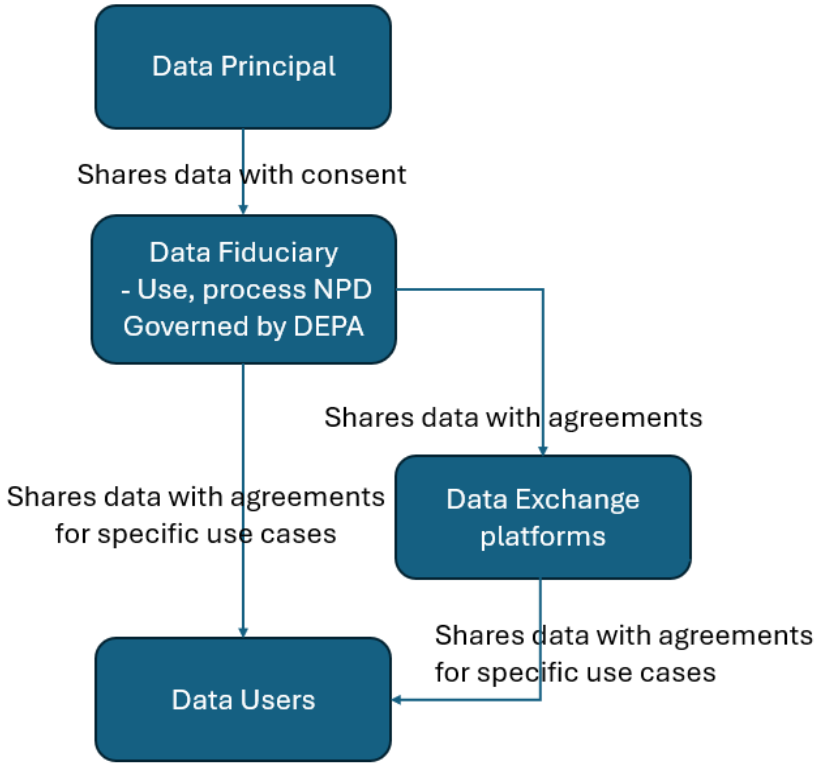
Agreements for Data Sharing/Exchange:

- 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

Annexure 2 : Data Flow Diagram



Annexure 2: Process Flow Diagram



Credits

- **Lead Organizations:**
  - Centre for Society and Policy, Indian Institute of Science, Bangalore
  - Center for Digital Public Goods, Indian Institute of Management Bangalore
- **Collaborative Effort:** This report is a result of a partnership between leading academic institutions focused on enhancing data governance in India.

Acknowledgements (1/2)

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

- *Names Included with consent of the stakeholders*

Acknowledgements (2/2)

Sr. No.	Name
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*

Contacts

- **Prof. Anjula Gurtoo**

  Indian Institute of Science, Bangalore

  Email: [office.csp@iisc.ac.in]

- **Prof. Srinivasan R**

  Center for Digital Public Goods

  Indian Institute of Management Bangalore

  Email: [cdpg@iimb.ac.in]

Notes

Notes

Notes