



सत्यमेव जयते

MINISTRY OF
**ELECTRONICS AND
INFORMATION TECHNOLOGY**
GOVERNMENT OF INDIA

INDIA DATA **ACCESSIBILITY** AND USE POLICY

BACKGROUND NOTE

FEBRUARY 2022

Introduction

India is accelerating towards an online population with the number of active Internet users in India expected to increase by 45%¹ in the next few years and touch 900 million by 2025 from around 622 million in 2020. India is likely to have 330 million 5G subscribers by 2026 and the monthly data consumption per smartphone is expected to grow over 3-fold to 40 gigabyte per smartphone.² Its IT and telecom infrastructure provides a robust backbone wherein citizens are increasingly relying on digital services for citizen-state services. As of 2020, India's IT workforce accounts for 4.47 million in FY 2020-21³ with its relative skill penetration in technology areas like artificial intelligence (AI) being among the highest globally.⁴ Further, emerging technologies are expected to generate \$1 trillion in economic value for India -- spearheading the \$5 trillion economy ambition by contributing 20 percent to the country's nominal GDP and sustaining 60-65 million jobs by 2025.⁵ The potential five-fold increase in economic value from India's digital transformation by 2025 would create a rapidly growing market for a host of digital services, platforms, applications, content, and solutions.⁶

Against the backdrop of an opportune digital ecosystem, data occupies the position of a fundamental building block. In light of data leaks, biased algorithms and cyber-attacks, the imperative for the responsible production, access and governance of data is central to the discourse on 'India's data and digital opportunity'.

With increasing digitization and engagement, the volume of data is also increasing exponentially, providing opportunities for better governance, service delivery and innovation in sectors critical for societal transformation.

India's ambitions of becoming a \$5 trillion-dollar digital economy depends on its ability to harness the value of data. Considering this, the India Data Accessibility and Use Policy aims to enhance access, quality, and use of data, in line with the current and emerging technology needs of the decade.

¹ Internet Adoption in India; ICUBE report June 2021

² Ericsson Mobility Report

³ NASSCOM Strategic Review 2021: New World 'The Future is Virtual'

⁴ Stanford AI Index Report 2021

⁵ <https://government.economictimes.indiatimes.com/news/technology/emerging-technologies-to-create-1-trillion-economic-value-by-2025-ey/71589838#:~:text=Emerging%20technologies%20are%20expected%20to,report%20from%20consulting%20firm%20EY.>

⁶ <https://community.nasscom.in/communities/digital-transformation/smartcities/role-of-technology-in-the-government-public-sector-undertakings.html>

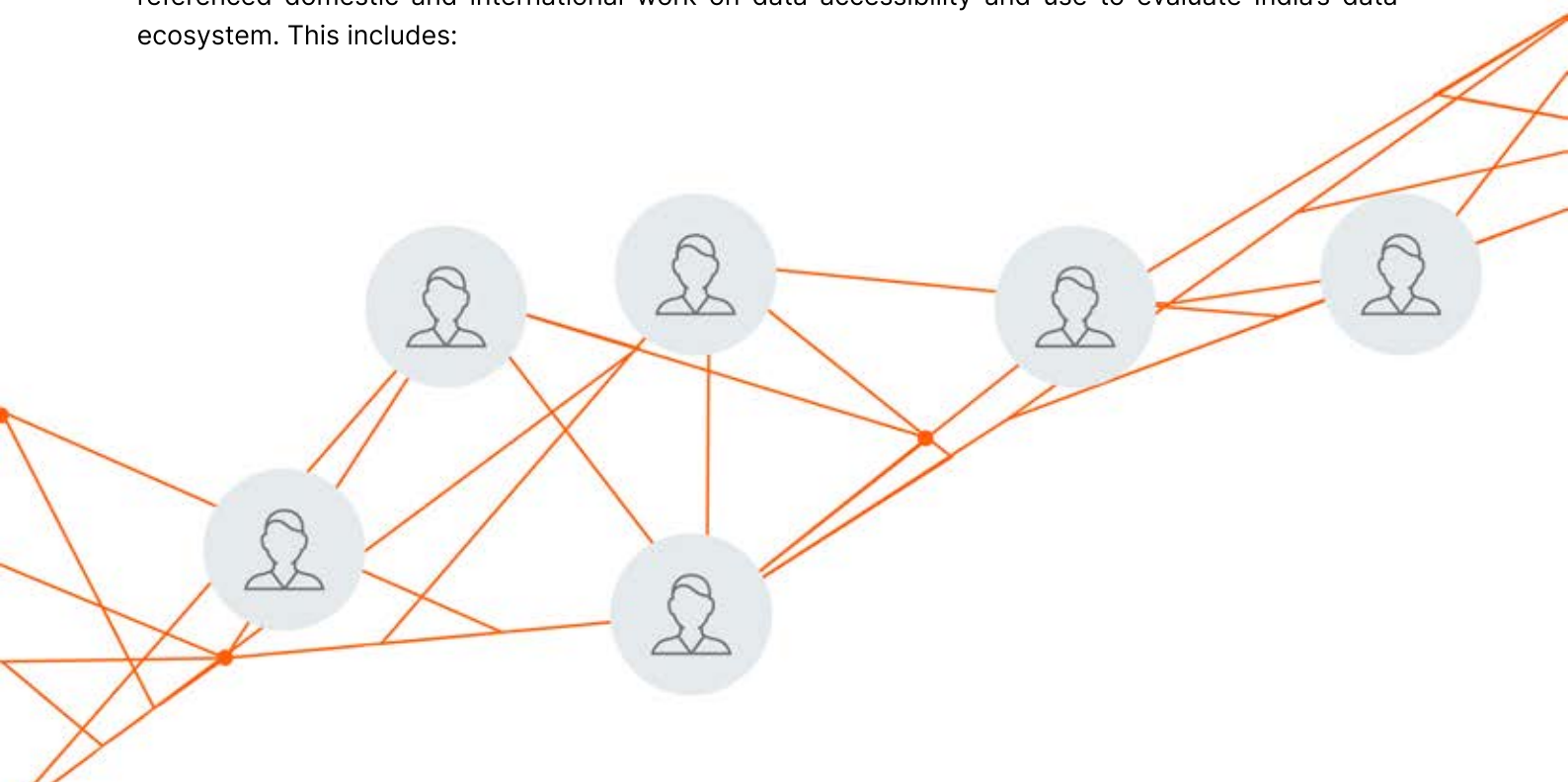
Importance of Data in the Digital Economy

Data is a valuable economic and social resource offering enormous opportunities for citizens, businesses, and governments to make better-informed decisions and develop innovative products and services

- **Huge Volume of Data:** With the rise of citizen interaction with public digital platforms and India's growing digital population, the volume of data is set to grow exponentially
- **Fostering the Digital Economy:** India's ambitions of Trillion Dollar Digital economy need the ability to harness the value of data
- **Improving Public Service Delivery:** Increased sharing of data across the public and private sectors optimizes digital welfare delivery and increases administrative efficiency, enabling citizen-centric services
- **Enhancing Efficiency & Innovation:** Improved availability of reliable, high-quality data catalyzes digital innovation for the public good, creating new economic opportunities
- **Improving Policy Making & Programs Evaluation:** Unlocking data strengthens the evidence base for policy development and public administration, bringing targeted social benefits to citizens.

Assessment of India's Current Data Accessibility & Use Ecosystem

The Ministry of Electronics and Information Technology has instituted multiple commissions and referenced domestic and international work on data accessibility and use to evaluate India's data ecosystem. This includes:



1. Outputs of the **OGD Data Audit** by **NASSCOM Task Force**
2. Review of the **Data Sharing Policies of different Ministries / Departments / States / UTs**
3. Assessment of other domestic data initiatives (such as the **NDAP portal** by NITI Aayog, State data portals, smart cities data portals and India Urban Data Exchange (IUDX) among others)
4. **Non-Personal Data Protection Framework** - 2020 Draft
5. PDP Bill 2019; **JPC Report on PDP Bill**
6. Body of Work Published on the **e-gov standards portal** by STQC, including **sectoral metadata & data standards, Interoperability framework, policy on open standards, etc.**
7. Outputs of the **Open Data Working Group** - IIC and IDFC Institute
8. Outputs of the **CDO Surveys** - IIC and CPR's Accountability Initiative
9. Study of **International Public Sector Data Policies & Papers** including but not limited to:
 - a. Data Sharing & Release Legislation Consultation Paper (Australia)
 - b. Data Availability and Transparency Bill 2020 (Australia)
 - c. European Data Strategy 2020
 - d. EU Open Data and Re-Use of Public Sector Information Regulations 2021
 - e. EU Data Governance Act 2021
 - f. Digital Economy Act (DEA) (UK)
 - g. Public Sector (Governance) Bill (Singapore)
 - h. Federal Data Strategy (USA)
 - i. Estonia's Public Information Act

Major Past and Ongoing Data Initiatives

The government of India has taken several measures on data as a part of the Digital India programme that inter-alia includes e-Governance initiatives. Some of the major initiatives are as follows:

1. **Open Government Data (OGD) Platform** - In pursuance of the NDSAP- Policy notified by the Government of India in March 2012, MeitY through NIC has set up the Open Government Data (OGD) Platform India - <https://data.gov.in/> to provide open access through the proactive release of data available with various ministries/ departments/ organizations of Government of India. The portal has over 5 lakh resources, 12 thousand catalogues and over 1 lakh APIs.

2. **Open Government Data License** - Open Government Data License of India has been released in 2017 to ensure that the data sets released are not misused or misinterpreted (for example, by insisting on proper attribution), and that all users have the same and permanent right to use the data.
3. **Ministry/State/Department Data Sharing Policies** - Since notification of the NDSAP policy, various state governments, ministries and departments have published their own detailed data sharing policies and guidelines that are based on the NDSAP policy. This includes the data sharing guidelines issued by Telangana, Punjab, Geological survey of India among others.
4. **Policy on Open Application Programming Interfaces** - The policy sets out the Government's approach on the use of "Open APIs" to promote software interoperability for all e-Governance applications & systems and provide access to data & services for promoting the participation of all stakeholders including citizens.
(https://www.meity.gov.in/writereaddata/files/Open_APIS_19May2015.pdf)
5. **API Setu** - API Setu was started in March 2020 and aims to bring all the APIs from Government to a single place and make them available for consumption by the Government departments and the industry. (<https://apisetu.gov.in/>)
6. **Meta Data & Data Standards** - The standards for metadata and data are significant for data sharing and interoperability among the systems. The domain-wise metadata and data standards are worked out by some Ministries and the same are notified at <http://egovstandards.gov.in/>. It is important to cover MDDS for the remaining domains.
7. **State & Sectoral Instances on OGD Portal** - Few States are contributing data directly on data.gov.in. Sikkim is the first State with its own Data Portal (<https://sikkim.data.gov.in/>), which was launched on 10th June 2016. The Open Government Data Portal of the Surat Municipal Corporation (<https://surat.data.gov.in/>) was launched on 23rd October 2016.
8. **IUDX** - IUDX is an open-source data exchange that facilitates secure and authenticated exchange of data amongst various data platforms, 3rd party applications, data producers and consumers, both within a city to begin with, and scaled up across cities eventually at a national level, in a uniform & seamless way.
9. **Cyber Security** - The safety and security of cyberspace is important because of data security, data privacy and data protection concerns. Incident based response and proactive guidelines on cyber security are issued on a regular basis by CERT-In. The National Cyber Security Policy has been published and the same is being revised considering the latest advancements in technologies. National Cyber Security Policy-2013 | Ministry of Electronics and Information Technology, Government of India

Current Challenges with Data Sharing & Use

Recognizing open data as a valuable public resource, the Ministry of Electronics and Information Technology has instituted multiple initiatives to assess the current challenges and improve opportunities in data accessibility and use.

The list of bottlenecks identified is outlined below:

- 1. Policy Monitoring & Enforcement** - Lack of a well-structured mechanism for monitoring the data sharing efforts of various line ministries/departments/agencies and ensuring the regular release of high-quality data sets.
- 2. Data Discoverability & Inter-Government Data Sharing Challenges**
Absence of comprehensive and updated data inventories leading to inadequate data discoverability, sub-optimal inter-government data sharing, duplication of data assets, and poor planning.
- 3. Siloed Data Platforms/Portals/Initiatives** - Multiple line ministries, departments, agencies, and state governments maintain their own data portals. There are several instances where these data portals/dashboards are not integrated with the Open Government Data platform or there is manual, inconsistent, and delayed integration with the OGD platform.
- 4. Absence of Data Sharing Toolkit/Playbook** - Presently there is no model data sharing toolkit to help chief data officers in assessing and managing risk associated with the sharing and release of data sets. As a result, many data cells follow a risk-avoidance culture and are reluctant to share detailed data sets.
- 5. Lack of Framework for Identification of HVDs** - There are no standard classification criteria to assist chief data officers in identifying High-Value Data Sets (HVDs). As a result, many line ministries/departments have not been able to consistently identify & maintain high-value data sets. Further, there is absence of a mechanism for identifying datasets that may benefit from linking.
- 6. Absence of Data Licensing & Valuation Frameworks** - Absence of innovative licensing frameworks, guidance on licensing approach, pricing data sets, criteria for valuation, and reference valuation models. Further, policy constraints towards the release of priced datasets through the OGD platform also inhibit sharing of detailed data sets.
- 7. Absence of Data Anonymization & Privacy Preservation Tools** - While the Data Protection Authority (DPA) gets established and defines data anonymization standards in line with the PDP bill, there is a need for reference anonymization and de-identification decision making frameworks & tools for assisting data cells & CDOs.

8. **Limited Focus on Restricted Data Sharing** - The current policy provisions only apply to non-sensitive data with a focus on highly aggregated & anonymized data. This prevents restricted sharing of detailed high-value data sets with trusted users for research & developments purposes under this policy.
9. **Lack of Meta-Data & Data Quality Standards** - While various sectors have published their own meta-data & data quality standards, this is still not done across sectors. Further, there is a need for ensuring that these standards are met while uploading data on the OGD portal.
10. **Inadequate Government Data Capacity** - There is a lack of comprehensive, continuous, capacity building strategy that focuses on building data literacy, competency, and knowledge of officers. There is a need for injecting external capacity into the data cells of respective line ministries/departments/agencies.
11. **Limited Citizen Engagement & Data Stewardship** - Current OGD platform lacks adequate mechanisms for community engagement and data stewardship. There is no mechanism for ensuring community requested data sets are added or for citizens to flag data sets due to privacy/security concerns or potential misuse of data

Way Forward - India Data Accessibility & Use Policy 2022

In order to transform the data accessibility & use ecosystem in India, the government of India is in the process of drafting the India Data Accessibility & Use Policy. This policy aims to enhance access, quality, and use of data, in line with the current and emerging technology needs of the decade.

The envisioned policy outcomes include:

1. **Unlock high-value data across the economy:** While India's private sector remain early adopters of technology and digital products, the thrust towards e-governance has meant that data is increasingly becoming central to fundamental citizen services. Likewise, the private sector relies on high-value public data repositories in sectors such as health, geo-spatial mapping, and energy to successfully inject innovation and propel a data-driven economy. A future-forward data policy framework should accordingly engage with data from multiple sources (public and private) to make accessible through G2G, G2B, B2G and B2B channels. Such a framework will not only enable more informed policymaking and efficient public services but will also allow a new generation of start-ups to bolster digital innovation for high-priority use cases and enter new markets, driving growth in the Indian economy.
2. **Facilitate a congruent and robust governance strategy:** Data governance policies are increasingly viewed as a tradeoff between privacy and innovation. Open data accordingly, is perceived with a degree of suspicion, not only because of the propensity it creates for accountability but also for the lack of understanding around misuse of data and safeguards against it. Principles of data privacy, security, and its implications on the safety of the data principal are tantamount to responsible technology and innovation, instilling trust in governmental data fiduciaries to adopt a rights-based approach to data sharing.

A responsive data governance policy framework focuses on devising safeguards to optimize principles of value to governance (like privacy and safety of the data principal) and one that adopts a risk management approach over risk avoidance.

3. **Realize an interoperable digital infrastructure:** Federated and interoperable architectures allow data sets and platforms to engage with each other's insights and functions such that the value of the architecture expands with the input of more data, without compromising the ability to update and improve its layers of functionality. A data architecture that is interoperable not only mandates adherence to strict data and metadata standards but also creates robust enforcement structures for standards across the data ecosystem. The flow of data within this architecture is further governed by comprehensive consent frameworks, technical safeguards for anonymization and streamlined sharing. This will be a crucial step in breaking data silos that exist today.
4. **Data skills and data-driven culture:** Skill investments within the government for building capacity in data science, analytics, emerging technologies, and ethics is crucial for ensuring the dissemination of high-quality data and its use.

The government aims to release the policy draft for stakeholder consultation, inviting views from all sectors including research, academia, startups, and industry on how the policy can institutionalize a data-sharing framework for the next decade. Along with the policy, the government also aims to publish detailed implementation guidelines and a comprehensive data sharing toolkit in a similar multi-stakeholder effort to ensure effective implementation in line with the policy objectives.

Annexure - I

1. **Open Data** - A dataset is said to be open if anyone is free to use, reuse, and redistribute it. It should be machine-readable and easily accessible.
2. **Open Government Data Portal** - The Open Government Data Platform developed by MeitY through NIC (Data.Gov.In). The OGD platform is a single point of access to datasets in open machine-readable format published by Ministries/Departments under the NDSAP 2012 policy.
3. **Negative List** – The list of data assets deemed non-shareable by ministries/departments; Datasets that are confidential in nature and/or are in the interest of the country's security in not opening to the public.
4. **Restricted Access Data Sharing** - Datasets that are accessible only through a prescribed process of registrations and authorization by respective departments/organizations since it could lead to a threat to life or loss of public assets or critical infrastructure.
5. **Data Inventories** - An exhaustive catalogue of existing data assets maintained by an entity.
6. **Licensing Framework** - It refers to the agreed legal framework for the exchange of data between two or more entities, the permitted use of datasets and the access term for those datasets.
7. **Open Government Data License**- Open Government Data License of India has been released in 2017 to ensure that the data sets released are not misused or misinterpreted (for example, by insisting on proper attribution), and that all users have the same and permanent right to use the data.
8. **Meta Data & Data Standards** - Metadata standards are intended to establish a common understanding of the meaning or semantics of the data, to ensure correct and proper use and interpretation of the data by its owners and users.
9. **E-Gov Standards** - E-Gov Standards ensure sharing of information and seamless interoperability of data across e-Governance applications for better decision-making, cooperation and coordination among the stakeholders.
10. **e-Govstandards Portal** - It is the official portal for e-Governance Standards to provide a platform for sharing ideas, knowledge, and draft documents among the members of various committees involved in the standards formulation process. The duly approved standards by Government's Apex body would be released on this portal for free download and usage.
11. **Data Anonymization** - The irreversible process of transforming or converting personal data to a form in which a data principal cannot be identified, which meets the standards of irreversibility specified by the competent authority.



सत्यमेव जयते

MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY
GOVERNMENT OF INDIA