Accelerating Data Sharing and Access

Implementation and Monitoring Recommendations on National Data Sharing and Accessibility Policy (NDSAP)

Wadhwani Institute of Technology and Policy (WITP)

Authors:

Shyam Krishnakumar, Technology Policy Consultant – WITP, Wadhwani Foundation

Alok Gupta, Director – Technology & Policy, WITP, Wadhwani Foundation

Executive Summary

The National Data Sharing and Accessibility Policy (NDSAP) was designed to promote open access to government data for national planning and development. The NDSAP has enabled significant sharing of open data resources. However, it is far from its desired goal because of a number of implementation, monitoring, and policy level challenges.

Implementation has lagged behind stated objectives due to multiple reasons including ambiguity around dataset classification, low data quality, lack of capacity and skills to create, manage and use high-quality datasets and no effective government-to-government data sharing mechanism. There are also monitoring challenges like lack of an effective oversight mechanism to monitor NDSAP implementation, no coordinated data sharing efforts across Ministries, and lack of mechanism to appeal in case of no response to dataset requests. This policy brief examines global best practices with a specific focus on the USA, Australia, UK and the European Union to address these key challenges and propose solutions tailored to the Indian context.

Key implementation recommendations to release high value datasets include all datasets to be considered open by default, clear definition of high-value datasets based on socio-economic value and importance to India's Artificial Intelligence (AI) strategy, publishing a priority list of high-value datasets within 6 months, and a clear decision-making framework to guide government agencies on dataset classification and privacy-preserving data sharing. To enable government-to-government data sharing, the report recommends the establishment of a searchable data inventory of all datasets available across government departments with clear metadata and data dictionaries and the creation of both a technical sharing mechanism and a standard procedure to request and receive data. The measures proposed to improve data quality include the adoption of specific and uniform data standards and metadata schema and mandatory annual data audits through an empaneled list of companies to ensure data quality and integrity. It is also proposed that the data.gov.in portal may be reviewed to enhance the usability of the platform. The report proposes empanelment of training agencies for capacity building and mandatory 2 day training programs every 6 months for officials from all NDSAP Cells to build both data management skills and data analysis skills to power data-driven policy. A cadre of data fellows must be deployed across Ministries to support data processing and release of standards-compliant data.

Key monitoring recommendations include creating Chief Data Officers Council under Chairmanship of Secretary, Ministry of Electronics and Information Technology (MeitY) and a National Data Sharing Mission headed by an Additional Secretary level officer under MeitY to provide an oversight mechanism which shall ensure implementation and monitoring of NDSAP mandate and a Project Management Unit under the supervision of the nominated officer to coordinate data sharing practices. A National Open Data Mission on the lines of the National

Digital Payments Mission headed by Secretary/Additional Secretary of MeitY can be constituted to coordinate data sharing measures and facilitate implementation and monitoring. The actual implementation will continue to rest with the respective ministry. Further, it is essential to upgrade data.gov.in portal and a dedicated Technical Support Unit is proposed under the proposed Mission which shall upgrade and maintain the data.gov.in portal. There is a need for a unified tracking dashboard for data sharing activities across Ministries to track the progress of different Ministries including the number of datasets shared, high-value data sets, frequency of updates etc. Finally, a time-bound appeals mechanism should be established to resolve inter-ministerial and public data-sharing challenges.

Additional policy measures that can be considered, is to expand the scope of NDSAP to include collation and publishing of open data of centrally sponsored schemes apart from the central sector schemes; replace the Sharable and Non Shareable dataset classification with Public, Government and Non Sharable categories and; bring the NDSAP under the IT Act or any other suitable Act to provide legal framework for execution of the data sharing activities.

1. Introduction

Open Data for Growth and Development

Data is a key resource for economic growth, innovation, job creation and social progress[1]. The Open Data Charter defines open data as "digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere."[2] The open data movement is based on the premise that vast amounts of non-sensitive government data generated using public funds can be used by the public for scientific, economic and developmental purposes. Open government data is a key asset for economic development, digital innovation, and data-driven policy-making. In addition to obvious benefits of data-driven policymaking for national development, the Organisation for Economic Cooperation and Development (OECD) estimates that public-sector data availability and sharing can generate benefits ranging up to 1.5% of gross national GDP. The number rises to up to 2.5 % of GDP when private-sector data is included[3].

Open Data To Ensure Responsible AI for All

Further, the intersection of data and Artificial Intelligence provides a massive opportunity to address India's social and economic challenges in fields ranging from agriculture to financial services to health and mobility. However the potential for AI powered public technologies depends largely on the quality and availability of suitable datasets. The European Commission considers open data to be a critical asset for developing AI systems. Similarly the United States had launched a task force to make government data available for AI innovation. The National Strategy for AI of Government of India mentions the challenge of large amount of data siloed across Ministries[4]. If high-quality, machine-readable government data is made available across sectors, it can help realise the Government's vision of "Responsible AI for all"[5] to drive inclusive growth and progress. A report of National Association of Software and Service Companies (NASSCOM) states that the value of integrated data in key sectors in India is worth nearly $500 billion, one-tenth of the Prime Minister's vision of a $5 trillion economy by 2025.[6]

Open data is also essential to ensure transparency and accountability. Data sharing significantly reduces costs within and outside the government. Further, since global indices rely heavily on publicly-available data, publishing open data can improve India's ranking across development and economic indices.

---

[1] https://ec.europa.eu/digital-single-market/en/european-strategy-data
[2] https://opendatacharter.net/principles/
[3] The OECD, Enhancing Access to and Sharing of Data, page 11.
[4] http://niti.gov.in/national-strategy-artificial-intelligence
[5] https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf
[6] https://nasscom.in/knowledge-center/publications/unlocking-value-data-and-ai-india-opportunity

To empower India's data ecosystem, the Ministry of Electronics and Information Technology (MeitY) has launched flagship initiatives including the National Open Digital Ecosystems (NODEs) and India Enterprise Architecture (IndEA). However, underlying all this is a compelling need to revamp India's open data sharing policy in-order to accelerate the release of high-quality data for governmental and public purposes.

1. NDSAP Overview and Current Status

NDSAP Overview

India's Union Cabinet approved the National Data Sharing and Accessibility Policy (NDSAP)[7] in February 2012, to promote data sharing and access to government data for national planning and development. The NDSAP applies to all shareable non-sensitive data generated using public funds by central government agencies and departments. The Department of Science and Technology is the Nodal Department for the NDSAP policy and the Ministry of Electronics and Information Technology is the Ministry responsible for its implementation. The NDSAP's principles for data sharing and accessibility include: Openness, Flexibility, Transparency, Legal Conformity, Protection of Intellectual Property, Formal Responsibility, Professionalism, Open Standards, Interoperability, Quality, Security, Efficiency, Accountability, Sustainability and Privacy.

Data Classification

The NDSAP mandates all Departments to prepare a Negative List of datasets which cannot be publicly shareable. All datasets which do not fall under the Negative List would be automatically in the publicly shareable Open List.

The NDSAP has provisions for three types of data access:

1. **Open Access:** Datasets available without any process of registration or authorization.

2. **Registered Access:** Datasets that are accessible to institutions and public users through a prescribed process of authorization by respective departments.

3. **Restricted Access:** These datasets will be accessible only under government authorization.

---

[7] https://dst.gov.in/national-data-sharing-and-accessibility-policy-0
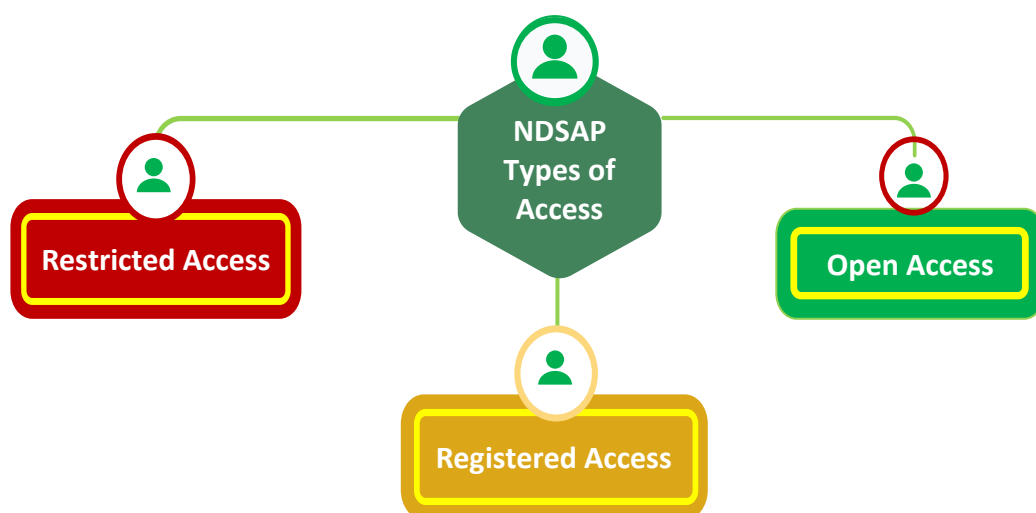
Open Government Data Platform

The Government of India (GoI) has launched data.gov.in to enable its Open Data initiative. Data.gov.in provides a unified platform for all government Ministries, Departments and organizations to publish their datasets, documents and services for public use in machine readable formats[8]. Over the last nine years, the number of resources has consistently grown to 512,588 resources in 11,096 catalogs from 195 Departments. There are 122,228 open web services from various Ministries. The resources have been downloaded 8.94 million times[9].

---
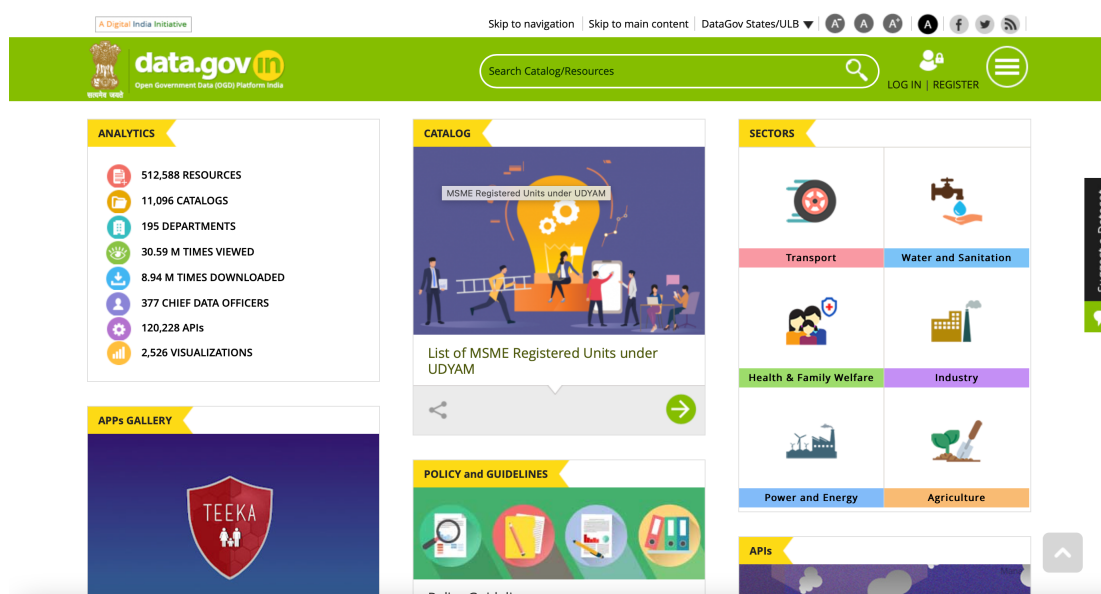
[8] https://www.meity.gov.in/writereaddata/files/OGD_Overview%20v_2.pdf

[9] https://data.gov.in/ as accessed on 25/10/2021

Government Open Data License

The Government Open Data License grants a royalty-free, non-exclusive license to use, adapt, publish and create derivative works for commercial and non-commercial purposes. However, legal ownership of the data will remain with the Ministry or Department which collected them[10].

---

[10] https://data.gov.in/government-open-data-license-india

2. Key Challenges

NDSAP has enabled data sharing and significantly lowered barriers to public access of non-sensitive government data. However, there are still significant implementation and monitoring challenges to the effective sharing and use of high-quality government data. India's position in the State of Open Government Data Index by Open Knowledge Foundation has slid from 10th place out of 97 countries in 2014 to 32nd place out of 94 countries in 2017[11].

1. Implementation Challenges

a. Implementation varies widely across Ministries

The number of datasets published and frequency of updates varies widely across Ministries. While some Ministries have been proactive in data sharing, others have lagged behind. The Chief Data Officer nominations on data.gov.in have not been updated since 2014 for some of the Ministries which provides an impression that the NDSAP cells have become inactive in those Ministries.

b. Ambiguity around data classification

The current guidelines for classifying data are limited to basic principles. Lack of specific and effective guidelines for data classification creates ambiguity within Ministries on whether certain datasets should be shared. The absence of clarity leads to a risk-averse approach and a focus on downside risks, possibly resulting in non-sharing. Specifically:

   i.   There is no standard mechanism to define a negative list.
   ii. Lack of clarity on how to classify datasets as Open Access, Registered Access and Restricted Access and the mechanisms for Registered and Restricted Access to data.

c. Poor data quality

Despite hosting hundreds of thousands of datasets, the use of India's open data remains limited due to issues of quality, diverse standards and gaps in high-value data. Quality, consistency and completeness are essential for data analysis to unlock insights. The data quality challenges in India's open data include:

   i.   Inconsistent or incomplete Datasets.
   ii.   Non-updation of data at regular intervals leads to outdated data in substantial cases.
   iii. There is a lack of common data and metadata standards.
   iv.   In case of variations in data-attributes, there is absence of concordance in data over a period of time.

---

[11] https://index.okfn.org/place/

v. Unfriendly presentation of user data leads to usability challenges for example, in certain datasets each time period is stored as a different file, sometimes with even different labels, which makes analysis challenging.

d. **No Effective Government-to-Government Mechanism**
While data.gov.in enables sharing of open access data with the public, there is no effective government-to-government sharing process or data-architecture. This results in siloed data within Ministries and or Departments.

i. Government Departments have no visibility of data held in other Departments or even within the same Department.
ii. The lack of effective intergovernmental data sharing mechanisms results in lack of high-quality data for decision-making and national development.
iii. There is significant duplication of effort and cost in collecting similar datasets across Ministries.

A prime example of the challenges caused due to lack of data sharing can be seen with the multiple schemes that target pregnant or lactating mothers. The Ministry of Women and Child Development runs Anganwadi services, Poshan Abhiyaan and Pradhan Mantri Matru Vandana Yojana for pregnant or lactating women using the same Anganwadi platform. However, due to a lack of intra-ministerial data sharing, the beneficiary data is separately collected for each scheme, leading to unnecessary repetition of effort, increased costs, and major inconvenience for the intended beneficiaries. Similar issues arise due to the lack of data-sharing between Ministry of Women and Child Development and Ministry of Health and Family Welfare who again target the same group of children, pregnant women and lactating mothers through multiple schemes using different beneficiary IDs which makes it extremely difficult to use the data of two Ministries together for measuring impact assessment or to perform any other analysis. Many schemes of the government have same or similar eligibility conditions. The absence of data sharing leads to delayed identification of beneficiaries under newer schemes as it requires fresh round of KYC by the beneficiaries, which can be easily done away with.

e. **Lack of capacity and skills to create, manage and use high-quality datasets**
Releasing high-quality datasets is time-consuming work that requires technical expertise. But many Government Departments do not have adequate personnel specialising in data quality and data release. Further, systems are often not designed with a focus on data quality and ability to share.

i. There is a skill gap related to processing and release of datasets in a privacy-preserving manner in NDSAP cells, especially with respect to Data Contributors who are responsible for compilation of data which is required to be published. This prevents or slows the release of datasets even when there is consensus on public sharing.
ii. Ministries have a shortage of personnel with data analysis skills within Ministries to leverage government data for data-driven decision making.

This restricts the use of data within the government and lack of compelling use cases reduces commitment to data processing and release.

iii. Chief Data Officers who are Joint Secretary level or above officers, handle the responsibility in addition to their current responsibilities. Similarly, the NDSAP cell members are likely to have other responsibilities as well. The 'additional work' aspect with the overworked staff displaces the focus required in good quality data generation and thus impacts the data sharing substantially.

iv. The data.gov.in displays the outdated or nil information with respect to Chief Data Officer's nomination of several Ministries which suggests an ineffective NDSAP cell in that Ministry. This echoes the idea that the nominations of Data Contributors may also not be a smooth process which ultimately impacts data governance.

2. Monitoring Challenges

a. No effective monitoring mechanism to oversee NDSAP implementation
While the NDSAP mentions the constitution of an oversight committee, there currently exists a compelling need for an effective monitoring mechanism to coordinate data sharing across Ministries and periodically evaluate the performance of various Ministries in data sharing on parameters like number of datasets shared, high value datasets shared, data quality, NDSAP and CDO nomination and functioning etc.

b. Absence of central coordination to facilitate data sharing.
The NDSAP Programme Management Unit's (PMU) functions are highly technical in nature and restricted to managing the Open Government Data (OGD) platform, providing technical advice, supporting dataset contribution and capacity building. The operational efficiencies in data sharing requires coordination, guidance and handholding to prioritise high-value datasets, to ensure release of high quality, high value datasets and submit periodic reports on progress. This requires a central coordinating entity, a Program Management Unit (PMU) responsible for coordination and monitoring, which has not been institutionalized in the current setup.

c. There is no appeals mechanism in case datasets requests are rejected
While citizens can request for specific datasets, there is no mechanism for follow up in case the concerned Ministry does not respond. Similarly, there is no appeal mechanism in the case of inter-governmental data sharing requests. There is a need for a time-bound appeals mechanism to create accountability.

3. Global Best Practices

Many governments across the world have passed legislation, published data strategies and built tools for open and inter-governmental data sharing. While many elements of each country's approach exist within a specific institutional structure and legal framework, nevertheless, there are certain key elements that can be contextualised to India's requirements. This section outlines some of the best practices from different Governments worldwide on five key issues:

a. Identifying High Quality Shareable Datasets
b. Enhancing Data Sharing with Privacy Safeguards
c. Monitoring Mechanisms
d. Government-to-government Data Sharing
e. Building Data Analysis and Data Management Skills

1. Identifying High Quality Shareable Datasets

a. Open by Default Approach

Many countries worldwide have an Open by Default requirement that mandates that non-sensitive Government data must be shared publicly unless there is a greater public/national interest for withholding it. 30 out of 32 OECD countries have an Open by Default requirement[12]. While there is overall consensus on the Open by Default requirement globally, there are concerns that it is a vague idea that lacks a clear definition needed to help Governments put it into practice[13]. In the United States, the Foundations for Evidence Based Policymaking Act 2018 requires agencies to make federal data publicly available by default and requires a comprehensive searchable data inventory of all datasets collected by each agency. Australia and many EU countries like France have an open by default requirement as well.

b. Prioritisation Mechanisms

Government Departments may not have the bandwidth to release all non-sensitive data and the public value of datasets varies widely. Therefore, prioritisation mechanisms for release of data are necessary. Governments worldwide are adopting two broad approaches to address the challenge of prioritisation:

i. Identifying High-Value Datasets:

The European Union's Open Data Directive, which aims to facilitate the creation of EU-wide cross-border use of public sector data, provides a "thematic list" of high-value categories of data that have important socio-economic benefits. The indicative list includes postcodes, geospatial data including national and local maps, demographic and economic indicators

---

[12] https://www.oecd.org/gov/digital-government/ourdata-index-policy-paper-2020.pdf
[13] https://medium.com/opendatacharter/a-map-for-open-by-default-53fe045951be

(statistics), business registers and registration identifiers (companies and company ownership) and mobility data road signs and inland waterways (mobility). Specific datasets under each of these categories will be notified. There is also provision to add to the thematic list of high-value categories[14]. Australia's Productivity Commission Report recommended that datasets which have significant community-wide benefits should be designated National Interest Datasets. National Interest Datasets would be aggregated and linked across Departments and prioritised for wider sharing and release.

ii. Demand-driven Approach

It is challenging for Government Departments to foresee the creative ways in which industry, researchers and civil society can interlink and use their datasets. Some Government agencies like the US Department of Health and Human Services (HHS) have adopted a demand-driven approach for prioritising datasets where stakeholders from civil society, academia and other government organizations put forth what data they require. This approach helped the HHS prioritize the release of its most-relevant data and enabled better allocation of limited manpower and resources.

Key Issue	Global Best Practices			
	USA	Australia	EU	OECD
To increase sharing of datasets	Open by Default mandated by the Foundation for Evidence-Based Policymaking Act of 2018	Australia's Public Data Policy Statement commits to release non-sensitive data as open by default	Many EU countries including France have Open by Default requirements.	30 out of 32 OECD countries have an Open by Default Requirements.
Identifying High Value Datasets	Demand-driven approach by the Department of Health and Human Services which involved the public in prioritization.	Australia's Productivity Commission recommended identification of National Interest Datasets, which have substantial benefits for the Australian population.	The EU Open data directive provides a "thematic list" of high-value datasets with provisions for further addition.	-

---

[14]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG

2. Enhancing Data Sharing with privacy safeguards

a. A searchable inventory of all Government data

A key challenge to data sharing is the lack of knowledge of what data is held by different Departments. To address this, the United States Evidence Based Policymaking Act mandates each agency to maintain a comprehensive data inventory of all data assets created, controlled and maintained by the agency including those which are not yet shared. The data inventory must be updated within 90 days of creation of a new dataset. The act also mandates a Federal Data Catalogue to allow government agencies and the public to understand which data is collected by which agency and have access to inventories that summarise the key characteristics of the data. By 2016, all EU member countries had established searchable national catalogues indexed by metadata. The European Data Portal aggregates data from all national portals.[15]

b. Privacy-Preserving Data Sharing

The more detailed and granular the data, the more valuable it is. However, there is a risk of divulging data that could directly or indirectly lead to re-identification of individuals or other privacy violations. A study showed that two-thirds of the US population can be identified with basic census data like gender, date of birth and zip code[16]. Concerns over privacy lead to the government department's hesitancy in sharing data. Overcoming this requires both a clear decision making framework on what data can be shared and tools that allow sharing while preserving privacy.

i. Privacy Preserving Tools

Australia's data innovation centre Data61 has done pioneering work on developing privacy preserving tools that enable government departments to securely share data. Their Personal Information Factor (PIF) tool uses data analytics to identify the security and privacy risks of re-identification of individuals from datasets. The tool is being used by the New South Wales Government to apply protection measures to granular Covid-19 datasets before release as open data[17]. The Re-identification Risk Ready Reckoner (R4) is a tool to understand the risk of re-identification in datasets. R4 simplifies the process of releasing a dataset by highlighting problematic records and suggesting corrective measures to be applied.[18] Finally Data61's Confidential Computing platform enables analysis of sensitive data across government departments without compromising privacy. The tool uses homomorphic encryption and secure multiparty

---

[15] https://www.stateofopendata.od4d.net/chapters/regions/european-union.html

[16] https://dataprivacylab.org/projects/identifiability/paper1.pdf

[17] https://www.csiro.au/en/news/news-releases/2021/new-data-privacy-tool-ensures-anonymous-covid-19-data-remains-secure-and-private

[18] https://data61.csiro.au/en/Our-Research/Our-Work/R4

computing to train machine learning models across multiple government datasets without any data leaving their respective secure databases. Homomorphic encryption ensures that data processing is done on encrypted data. There is no loss of accuracy due to encryption and distributed processing[19].

ii. Decision Making Frameworks

The United Kingdom Anonymisation Network has developed 'The Anonymisation Decision-Making Framework' to help Government officials make decisions on whether they can share a dataset and to what degree of detail[20]. In the US, the Federal Committee on Statistical Methodology is developing a data protection toolkit of best practices to maintain privacy and confidentiality[21]. Australia's Information Commissioner's Office has released a De-identification Decision Making Framework to enable government departments to help departments make sound data release decisions based on best practices.

Key Issue	Global Best Practices			
	USA	Australia	EU	UK
Searchability of data held across all government agencies.	A data inventory and federal data catalogue are mandated by the Evidence Based Policymaking Act of 2018.	-	All EU member countries have searchable national data catalogues. The European Data Portal aggregates all national data.	-
Facilitate data sharing while preserving privacy	Data protection toolkit of best practices.	Australia's data innovation centre Data61 has developed a suite of new tools and technologies to enable data access while managing privacy risks with sensitive data through homomorphic encryption and secure multi-party computing. Deidentification Decision Making Framework	The EU Open data directive provides a "thematic list" of high-value datasets with provisions for further addition.	The UK has developed 'The Anonymisation Decision-Making Framework' to help decision makers understand if they can share a dataset and in what degree of detail.

---

[19]https://data61.csiro.au/en/Our-Research/Our-Work/Safety-and-Security/Privacy-Preservation/Confidential-computing

[20] https://ukanon.net/framework/

[21] https://nces.ed.gov/fcsm/dpt

3. Monitoring Mechanisms

An effective whole-of-government oversight mechanism is essential to monitor progress and facilitate inter-ministerial coordination in data sharing. Countries like Canada and France have adopted a Chief Data Officer approach. Canada has a Government Chief Data Officer[22] while France has Administrateur Général des Données to oversee its data strategy[23]. The Administrateur Général des Données reports to the Prime Minister and coordinates the government's actions with regard to the inventory, governance, production, release and use of data[24]. Canada's Chief Data Officer is responsible for implementing Canada's data management strategy, ensuring consistency in data quality and overall data governance. The Chief Data Officer reports to the Prime Minister[25].

The United States has taken a more federated approach with a Chief Data Officer Council headed by a Council Chair. Chief Data Officers from all Government Departments serve as council members. The council is responsible for establishing best practices in data generation, management, use and release, facilitating inter-departmental sharing of data and improving public access to government data[26].

---

[22]https://www.canada.ca/en/public-service-commission/services/publications/roles-responsibilities-with-respect-data-at-the-psc.html#2_1

[23] https://www.etalab.gouv.fr/administrateur-general-des-donnees

[24] https://www.etalab.gouv.fr/administrateur-general-des-donnees

[25] https://www.canada.ca/en/public-service-commission/services/publications/roles-responsibilities-with-respect-data-at-the-psc.html#2_1

[26] https://www.cdo.gov/about-us/

Key Issue	Key Oversight Agency for Data Sharing
Global Best Practices	
USA	US Federal Chief Data Officer Council with a Council Chair
EU	All EU member countries have searchable national data catalogues. The European Data Portal aggregates all national data
AUS	Office of the National Data Commissioner The National Data Advisory Council with members from industry, academia and government advises the National Data Commissioner on community expectations, technical best practices, and industry developments and ethical use of data
FRA	Administrateur Général des Données
CAN	Government Chief Data Officer.

4. Government-to-Government Data Sharing

Government data present in ministerial silos hinders effective data-driven decision-making and leads to duplication of data collection and management efforts. In this regard, Estonia's Information Sharing Data Sheet initiative(X-Road) facilitates intergovernmental data exchange by linking key national databases. It is motivated by the "once-only" principle where public agencies should only collect data that is not previously maintained in any other government database. Australia has replaced the cumbersome Memorandum of Understanding process for data sharing between government agencies with a simple letter of exchange. Further, government entities must share data by default with other government entities unless there are legislative barriers or risks to privacy, security or confidentiality.[27]

a. SenDa: Workflow Tool for Data Sharing
Australia' Data61b has developed SenDa, a workflow tool to manage access to sensitive datasets. It is designed to replace the offline or email-based workflow for

---

[27] https://www.pmc.gov.au/resource-centre/public-data/guidance-data-sharing-australian-government-entities

obtaining data between government agencies. SenDa allows users to explore a catalogue of available data, select a desired subset, make an access request and receive data through an online workflow. Each request has its own audit trail. The tool is designed to improve efficiency in access to sensitive datasets while ensuring confidentiality and data governance.

Key Issue	Government to government data sharing
Global Best Practices	
ESTONIA	Information Sharing Data Sheet initiative(X-Road)
AUS	Government agencies must share data by default with other agencies. Data61's SenDa data sharing tool to enable data-sharing between government departments.

5. Building Data Analysis and Data Management Skills

Maintenance, processing and public release of government data requires considerable upskilling of government officials who may not necessarily be trained for this particular task. The Australian government has created Data Training Partnerships with a list of empanelled partner organizations. Government departments can approach them for customised training and capacity building programs[28]. The United States Federal Data Strategy Action Plan mandates the creation of a curated skills catalogue to build data management and data driven decision-making skills in the government[29]. Practice 27 of the Federal Data Strategy Framework 2020 mandates training for all agencies to improve data analysis, evaluation, data management and privacy protection competencies in the government workforce.[30]

---

[28] https://www.pmc.gov.au/sites/default/files/publications/data-skills-capability.pdf
[29] https://strategy.data.gov/action-plan/#action-13-develop-a-curated-data-skills-catalog
[30]https://strategy.data.gov/practices/

Key Issue	Capacity building within the government
Global Best Practices	
USA	Curated data skills catalogue
AUS	Data Training Partnerships with partner organizations.

4. Recommendations

The focus of this paper is to suggest recommendations that are limited to implementation and monitoring challenges. Based on the current challenges and a study of relevant international best practices, we recommend the following measures:

A. Implementation Level

1. **Datasets Must be Open By Default**

   The NDSAP should adopt an explicit Open by Default requirement to enable Ministries and Agencies to shift from a *need-to-know* to a *share-by-default* approach.

   a. **Datasets must be open by default unless the Ministry/Department provides a specific reason to include it in the negative list.**

   b. Negative lists of all Ministries/Departments must be periodically, preferably every 6 months, reviewed by the oversight committee.

2. **Identifying High Quality Shareable Datasets**

   a. **Publish a specific and actionable mechanism to identify high-value datasets.** The current NDSAP definition of high-value datasets focuses on data quality principles like completeness, timelines, ease of access, machine readability and "performance". It also permits each Department to have its own criteria of high value datasets, creating ambiguity and lack of accountability.

      i. **High value datasets** must be defined based not on data quality but according to **value of information and the degree of socio-economic benefits.** The EU Open Data Directive provides a possible model.

      ii. Data that is necessary to power India's AI strategy in key sectors can be considered high-value data. This can be identified through consultation with government and public stakeholders.

      iii. Finally, datasets which determine performance of the country in key global indices should be considered as high value datasets.

   b. **To start with, MeitY may identify and publish a list of important high-value datasets of various ministries to be released within 6 months.** High-value datasets may be distributed across Ministries. Please refer to **Appendix 1** for an indicative list of high-quality datasets.

   c. High-value datasets must have **good data quality standards** and should be **updated minimum every quarter.**

   d. Individual Ministries and Departments can **adopt a demand-driven approach** to prioritize and release pipeline of datasets.

---

e. Government officials should be provided with **simple questionnaire-based tools to help them classify** if data is publicly shareable, shareable within the government or not shareable.

3. **Data Quality**
   a. **Enforcement of specific and uniform data and metadata standards to ensure data quality.**
      The current data and metadata standards are quite basic and require a thorough review and revision to keep pace with the growing data requirements.
      i. **Revise and publish updated data standards.** Data standards enable consistency in data and provide rules to reliably structure datasets so that they can be read, indexed and shared. NITI Aayog's Government Data Quality Index can serve as a possible reference[31].
      ii. **Review and publish updated standard metadata schema.** Metadata Schema is critical for discoverability of datasets, for governmental and public data users to find the dataset through the Inventory search engine. If metadata is machine readable, it greatly increases its utility. The US DCAT Metadata Schema provides a good model for metadata standards for government data[32]. Interestingly, the schema has an access Level field which specifies if a dataset is for public, intergovernmental or restricted use and a Rights field for departments to record the reason for not releasing a dataset.
      iii. **Timeline based approach to ensure that all Ministries adopt standards.** Within one year of release of updated data and metadata standards, the head of the IT division or concerned authority in every Ministry should give a declaration that all existing systems conform to data and metadata standards prescribed by the Government.
      iv. **All new systems must be standards compatible.** Before making any Government system 'live', the concerned authority should give a declaration to MeitY that the system complies with all the latest prescribed data and metadata standards. A mandatory audit to ensure compliance of metadata standards may be implemented just like currently the application security testing is mandatory before it gets hosted.

   b. **Mandatory Annual Data Audits to ensure data quality and integrity**

---

[31] https://dmeo.gov.in/sites/default/files/2021-03/DGQI_Toolkit.pdf
[32] https://resources.data.gov/resources/dcat-us/

i. All Ministries and agencies must conduct mandatory annual data maturity assessments of all systems to ensure the integrity of data generation, processing and release.

ii. Data Audits should evaluate the adoption of data standards and metadata standards.

iii. The Standard Operating Procedure for data audits may be defined by MeitY.

iv. The data audit can be conducted by the companies that have been empaneled by MeitY for IT audits. This mechanism is similar to the IT audits mechanism already in place for all Departments.

c. Enhance the ease-of-use of data.gov.in

The uptake of data consumption can hugely drive the data generation requirements. As the data.gov.in is the sole platform to enable data sharing, it is important to review and lay down the roadmap for making it more popular and user friendly for data publishing and consumption. This may include improving the user-friendliness, better search functionality to discover relevant datasets, faster load times, etc.

4. Government-to-Government Data Sharing

a. Establish and frequently-update searchable data inventories

There is a compelling need for a searchable index of all datasets available across Ministries and agencies within the government, so that all Departments have a whole-of-government view of what data is already available for usage. This will improve data use with the government, power data-driven decision making on key policy issues and reduce duplication of effort. The data sharing between Ministries can lead to holistic governance and economic growth which is very well highlighted in the recently launched GatiShakti Programme where around 16 Ministries will share the data for common oversight on the infrastructure development projects across the country. It is therefore suggested that -

i. **Each Ministry may be mandated to maintain a searchable inventory of all datasets created, controlled and maintained by the Ministry and all Departments/agencies** under it including those which are not yet shared. The data inventory will include datasets that are publicly shareable, shareable within government and restricted.

ii. The data inventory must have **adequate metadata and a data-dictionary** of all fields present in the datasets. It should also make it easy to understand what data is collected by the Ministry and bodies under it and how the inventory can be accessed.

iii. These inventories can be federated into a **government-wide searchable inventory of datasets accessible only to the government. Upon**

sufficient maturity, the MeitY could consider **opening the searchable inventory to the general public** as well by aggregating these datasets at the data.gov.in portal.

iv. **Agencies and departments must update dataset information every 90 days.** Every 90 days all agencies should identify missing or incomplete dataset listings in their data inventories and ensure metadata is comprehensive for high value datasets.

v. **Eventually the Data Inventories can help power multiple sectoral NODEs** "open and secure delivery platforms, anchored by transparent governance mechanisms, which enable a community of partners to unlock innovative solutions, to transform societal outcomes".

b. **Create standard intergovernmental data-request and sharing mechanisms** Enabling data flow across Ministries and agencies requires both a technical sharing mechanism and a standard procedure to request and receive data.

i. **Pull-based mechanisms must be established to allow Ministries to request data from other Ministries**. A workflow-management solution like Data61's SenDa can enable the secure and efficient inter-governmental sharing of data through a clear approval mechanism.

ii. **Data exchanges can be used to facilitate the secure and authenticated exchange of data across various government platforms.** A key precedent is the Indian Urban Data Exchange under the Smart Cities Mission, which facilitates the sharing of urban data between government repositories[33].

iii. **A suitable appellate authority is needed to adjudicate government-to-government data requests** to handle cases of lack of consensus. This could be handled by the oversight committee.

5. Enhancing Data Sharing with privacy safeguards

a. **Develop and distribute a clear Anonymisation and De-identification Decision Making Framework** modelled on the UK or Australian approach along with a set of best practices to guide government agencies on privacy-preserving data sharing.

---

[33] https://smartcities.gov.in/India_Urban_Data_Exchange

b. Provide simple privacy-preserving tools for Ministries to share data with confidence. The government can initiate time-bound projects with academic and research organizations such as the IITs, IIIT's and IISC to develop privacy-preserving and anonymisation tools like those developed by Data61 to highlight problems in datasets and suggest corrective measures. These institutes may also be considered to help government develop **simple questionnaire-based tools to assist the government officials classify the data** into publicly shareable, shareable within the government or not shareable.

6. Capacity and Skill Building Measures

a. Mandatory refresher training sessions on data management and data skills

There must be mandatory training and capacity building programs for 2 days every 6 months for officials from all NDSAP Cells to build capacity in all government agencies to both manage and make use of data. The training programs will serve a two-fold objective:

    i. To build understanding of open data and skills to create, manage and publish high-quality open datasets

    ii. To build awareness of how data can be used within the public sector and data analysis skills to enable data-driven decision making to get better insights from available data. Building data usage skills across Ministries will naturally create the demand for quality datasets across Ministries. When the US Postal service standardised its datasets, its data modelling team could use the data to identify nearly $100 million in savings and recover over $30 million lost to possible fraud[34].

    iii. MeitY may empanel training providers and make them available to Ministries so that they select an agency to conduct the required training.

b. Create a platform for sharing best practices related to data governance between Ministries.

c. A cadre of Data Fellows should be deployed in Ministries to create dedicated capacity for data management and release. Dedicated data fellows will expand the capacity to process data internally and release datasets. Processing and release of quality-data conforming to data

---

[34] https://www.datacoalition.org/what-the-data-act-means-for-anti-fraud-analytics/

standards requires expertise and can be time-consuming. Data Fellows can address the capacity and skill challenges faced by multiple Ministries in implementing the open data sharing mandate. Data Fellows within a Ministry can assimilate data coming from different divisions and facilitate standards-compliant data release. They will function as a middle-layer between Data Contributors and the Chief Data Officer of the Ministry.

B. Monitoring Level

1. Chief Data Officers Council

In absence of a law which mandates Ministries to follow the guidelines, it is suggested to constitute a Chief Data Officers Council under chairmanship of Secretary MeitY with CDOs of Ministries as members. Subsequently, State IT Secretaries may be adopted as members of this council. The Council may restrict itself to policy and implementation issues.

2. National Open Data Mission

The data sharing activities require lot activities to be carried out which necessitate a requirement of an institutionalized mechanism which can manage the Open Data Sharing process in a Mission Mode. It is therefore proposed to have a dedicated structure which can be designated as National Open Data Mission just like National Health Mission or National Digital Payment Mission. This Mission can be headed by Secretary/Additional Secretary MeitY and have members from Ministries, Industry and Academia. The role of Mission shall be promote data sharing and facilitate the implementation and monitoring activities while the actual implementation shall be carried out the respective Ministries Data Cells under the leadership of respective CDOs.

3. Dedicated mechanism housed in MeitY to ensure implementation and monitoring of open data mandates.

There is a need for a dedicated mechanism headed by an officer of the Additional Secretary level under MeitY. This mechanism must have a separate Project Management Unit (PMU) and a dedicated Technical Support Unit (TSU). The PMU and TSU can be housed under this Mission who will carry out the necessary role of coordination, handholding, formulation of guidelines, monitoring and technical support. The PMU will also be responsible for facilitating periodic review of data sharing practices, review results of data audits and monitor the enforcement of data and metadata standards. The TSU may function under the supervision of NIC and the current data.gov.in team may become the part of the proposed TSU.

4. **Build a unified tracking dashboard of data sharing activities across all Ministries to enable oversight and monitoring.**
While data.gov.in provides an analytics dashboard for Chief Data Officers of various Ministries, there is a need for a whole-of-government dashboard that allows the oversight mechanism to effectively track and monitor the progress of different Ministries and agencies in data sharing including number of datasets shared, high-value data sets, frequency of updates etc.

5. **Establish an appeals mechanism to enable public appeals and resolve challenges in data sharing.**
It is important to provide a mechanism to address the issues which may emanate out of the data requests placed, sharing of data process, and the data shared. Therefore, it may be appropriate to establish an appellate which may review such appeals and provide appropriate resolutions.

C. Further Policy Measures

Apart from the above implementation and monitoring measures, the below policy measures could be considered in the medium term:

1. **The scope of the NDSAP can be expanded to include collation and publishing of all open data of centrally sponsored schemes also apart from the central sector schemes.** Mechanisms can be created to incentivise States to share their data on data.gov.in or equivalent platforms.

2. **The data classification into Shareable and Non-Shareable categories should be replaced with Public, Government and Non-Shareable Categories.**
This will facilitate greater clarity in implementation and to specifically identify data that can be shared across different government agencies. A sample classification of data for few Government Schemes is provided at **Appendix 1**.

3. **The NDSAP can be brought under the IT Act or any other relevant Act** in order to provide a legal framework for execution of the data sharing activities.

4. Government may consider enacting an evidence based policy making law like the Evidence Based Policymaking Act 2018 of USA which will bring in better enforceability to the measures contained in NDSAP 2012.

5. Conclusion

The Open Data Sharing requires progressive steps and it is important that MeitY lays down a clear roadmap on the steps that are required to be taken. However, to ensure the execution of all the required activities, it is important to establish a National Open Data Mission and putting in place a monitoring structure with a dedicated PMU and TSU on priority. Further, the MeitY needs to carry out one-on-one meetings with Ministries which have potentially high-value data. It also needs to issue guidelines for appointment of Data Fellows in all the Ministries apart from the Data Contributors that are in place. Data and Metadata standards should be reviewed and updated. Other measures that could be taken up in the short run include:

a. Adopting an open by default approach
b. Publishing a list of priority high-value datasets
c. Beginning capacity building measures to manage and analyse data across Ministries.

The interlinkages between the NDSAP and other data related measures including NODE, IndEA, National Data and Analytics Platform and sectoral strategies to build data exchanges must be systematically studied to ensure a coherent national data strategy that enables greater socio-economic progress and supports the rapid growth of data-driven emerging technologies like AI. The above steps will be fundamental to streamlining India's data sharing policy.

About Authors:

1. *Shyam Krishnakumar is Technology Policy Consultant with the Wadhwani Institute of Technology and Policy in New Delhi*

2. *Alok Gupta is Director Policy and Technology with the Wadhwani Institute of Technology and Policy in New Delhi and a Senior Fellow with the Wadhwani Chair in U.S.-India Policy Studies at the Center for Strategic and International Studies in Washington, D.C.*

Appendix 1

S. No.	Ministry	Scheme Name	Sector	Details	Shareable Data			Negative List
					Public (No PII or Sensitive Data)	Inter Ministry (After anonymization)	Intra-Ministry (After anonymization)	
1	Ministry of Rural Development	Mahatma Gandhi National Rural Employment Guarantee Scheme	Rural Wage Employment	Legal guarantee for one hundred days of employment in every financial year to adult members of any rural household willing to do public work-related unskilled manual work at the statutory minimum wage of Rs. 120 per day in 2009 prices.	1. No. and Types of Assets Created 2. No. of Beneficiaries and Amount Disbursed 3. No. of Days of Employment Provided	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details
2	Ministry of Rural Development	Pradhan Mantri Gramin Awaas Yojana	Housing, Rural	Provides financial assistance to rural poor for constructing their houses themselves.	1. No. of House Created 2. No. of Beneficiaries and Amount Disbursed	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details

3	Ministry of Rural Development	National Social Assistance Scheme	Pension	Public assistance to its citizens in the case of unemployment, old age, sickness and disablement and in other cases of undeserved want.	1. No. of Beneficiaries and Amount Disbursed under various programs	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details
4	Ministry of Women and Child Development	Anganwadi Services	Women and Child Development	The scheme aims to tackle nutrition and health problems in children below 6 years of age and pregnant women and lactating mothers by providing supplementary nutrition and other services upon registration at Anganwadi centres	1. Details of Anganwadi Centres 2. Infrastructure details of Anganwadi Centres 3. No. of Pregnant Women & Lactating Mothers 4. No. of Children between 0-6 months, 6-36 months and 37-72 months 5. Details of Take Home Ration Provided 6. Details of Hot Cooked Meals Provided 7. No. of Immunisation Carried out	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details

| 5 | Ministry of Women and Child Development | POSHAN Abhiyaan | Nutrition | The scheme aims to improve nutrition and health of children below 6 years of age and pregnant women and lactating mothers | 1. No. of Beneficiaries enrolled
2. No. of SAM Children
3. No. of Stunted Children
4. No. of Underweight Children
5. No. of Children with Low Birth Weight
6. No. of Anaemic Children
7. No. of Anaemic Pregnant Women and Lactating Mothers | Irreversibly anonymised Transactional Data | Beneficiary and Transactional Data without key identifiers like Aadhaar | Beneficiary's Personal Identifiable Information and Transaction Details |

6	Ministry of Women and Child Development	Pradhan Mantri Matru Vandana Yojana	Mother Care	Provides conditional cash incentive during pregnancy for compensation for wage loss ad to induce better health seeking behaviour	1. No. of Beneficiaries enrolled 2. No. of Beneficiaries provided maternity benefit 3. Amount Disbursed	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details
7	Ministry of Health and Family Welfare	Ayushman Bharat Yojana	Health	Ayushman Bharat National Health Protection Scheme of the Government of India aims to provide free access to healthcare for 50 crore people in the country.	1. No. of Beneficiaries Enrolled 2. No. of Beneficiaries provided benefit 3. Amount of Health Insurance provided 4. Details of Empanelled Medical Service Providers 5. Disease-wise no. of beneficiaries provided benefit	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details
8	Ministry of Health and Family Welfare	RNTCP	Health	Tuberculosis control initiative.	1. No. of Beneficiaries enrolled 2. No. of Beneficiaries provided benefit 3. Amount Disbursed 4. No. of dropouts 5. Mortality related information	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information and Transaction Details

9	Ministry of Health and Family Welfare	Covid 19	Health	Covid 19 Pandemic	1. No. of Covid Patients 2. No. of Active Cases 3. No. of Deceased 4. No. of Recovered Cases 5. No. of Tests Done 6. No. of Case Tested Positive 7. Details of Covid Hospitals 8. Details of Oxygen Banks 9. Details of Beds Available	Irreversibly anonymised Beneficiary Data	Beneficiary Data without key identifiers	Beneficiary's Personal Identifiable Information
10	Multiple Ministries	Scholarship Schemes	Education	Scholarships are being provided by different Ministries to students at various level of education so that they can pursue their studies	1. Scholarshipwise name of beneficiaries enrolled during a period 2. Status of application - accepted/rejected/in process 3. Amount of Scholarship provided	Irreversibly anonymised Transactional Data	Beneficiary and Transactional Data without key identifiers like Aadhaar	Beneficiary's Personal Identifiable Information
11	Ministry of Electronics and Information Technology	Digidhan Mission	Digital Payments	Digidhan Mission intends to promote digital payment transactions across the country	1. Details of POS along with acquirer 2. Details of transactions which include mode, amount, location, etc. 3. Aadhaar Seeding details 4. Mobile seeding details	As per public data	Irreversibly anonymised beneficiary data	Beneficiary's Personal Identifiable Information (PII)

| 12 | Department of Revenue, Ministry of Finance | GST | Finance | GST System gets validated data on taxpayers, data on returns, taxes paid, HSN level data for external trade as well as internal trade. It also has data on e-way bills generated for good showing transfer of goods from one place to another at PIN code level where HSN code of goods are shown at 4, 6 or 8 digit levels | Aggregate data on return filing States-wise for each month/quarter Aggregate data on tax payment under each tax for all States for each month Aggregate data on goods supplies based on e-way bill at HSN level from PIN code to PIN Code for each month. Aggregate data on export and import of goods at HSN level for each month. | GSTIN validation and basic data as made available on GST portal using APIs Aggregate Annual Turnover range at PAN level HSN dealt by Taxpayer based on GSTIN given by the other ministry (thru APIs) | Return data with Income tax and CBIC | PIIs of promoters of a Firm |
| 13 | Ministry of Power | Multiple Schemes | Power | Power schemes and issues relating to power supply/development schemes/ programmes/decentralized and distributed generation in the States and Union Territories | 1. Details of Categorywise Installed Capacity 2. Details of Daily Power Generation 3. Details of state, utility and number of circles sanctioned, awarded, and completed 4. Details of sanctioned amount, amount awarded, and amount disbursed | Irreversibly anonymised Transactional data | Irreversibly anonymised Consumer and Transactional data | Consumer Personal Identifiable Information (PII) |

#	Ministry			Description	Data			
					5. Total households 6. Electrified households as of date 7. Unelectrified households as of date 8. Details of Power transmission 9. Aggregated Technical and distribution losses 10. Cost of supply and aggregated revenue realization gap 11. Operational indicators such feeder and DT metering 12. Smart metering (%) 13. Feeder Segregation			
14	Ministry of Transport	Road Transport	Transport	Vehicle Registration and Driving License related activities in transport authorities of country	1. Details of Category-wise Vehicles Registered 2. Details of Vehicle type-wise Vehicle Registered 3. No. of Vehicles which are beyond the stipulated age 4. Details of Vehicle Registration Cancelled	Irreversibly anonymised Owner data	Owner and Transactional Data without key identifiers like Aadhaar	Owner's PII

WADHWANI
FOUNDATION

www.wfglobal.org

For more information reach us at: witp@wfglobal.org

ASIA | AFRICA | LATIN AMERICA