Government of India

Ministry of Statistics and Programme Implementation

Using Administrative Data for Governance: Harmonising Departmental Data at the State/UT Level

Final Note

National Deliberative Summit, Bhubaneswar, Odisha

29-30 April, 2026

Table of Contents

Table of Contents 2

List of Abbreviations 3

Executive Summary 5

1. Introduction and Background 9

1.1 What is Administrative Data? 9

1.2 Why Harmonisation? 9

1.3 Steps towards data harmonisation 10

2. Current Situation and State/UT Overview 12

2.1 Summary of State Inputs 12

2.2 Institutional set up at States/UTs 14

3. Key Challenges and Gaps 17

4. Central Ministry and International Best Practices 20

4.1 Central Ministry and Private Sector Best Practices 20

4.2 International Best Practices 22

5. Potential Use Cases 24

6. Maturity Assessment and Actionable Roadmap and Timelines 25

6.1 Maturity Assessment and Good Data Practices 25

6.2 Roadmap for States (Four-Phase Strategy) 27

6.3 Activity List and Timeline 32

7. Conclusion and Way Forward 35

Annexures 36

Annexure 1: Glossary 36

Annexure 2: Government Guidelines 40

Annexure 3: Potential Use Cases 42

Annexure 4: Good Data Practices 45

List of Abbreviations

API	Application Programming Interface
BLADE	Business Longitudinal Analysis Data Environment
CDO	Chief Data Officer
CeG	Centre for e-Governance
CEMS	Continuous Emissions Monitoring Systems
DAU	Data Analytics Units
DBT	Direct Benefit Transfer
DES	Directorate of Economics and Statistics
DGFT	Directorate General of Foreign Trade
DoSEL	Department of School Education and Literacy
DPDP	Digital Personal Data Protection
DRD	Derived Record Depository
DST	Department of Science and Technology
DSU	Data Strategy Unit
EMIS	School Education database
EPFO	Employees' Provident Fund Organisation / Payroll data
ETS	Emissions Trading Scheme
FRUITS	Farmer Registration & Unified Beneficiary Information System (Karnataka)
FSI	Forest Survey of India
GI	Geographical Indications
GO	Government Orders
GODL	Government Open Data Licence
HMIS	Health Management Information System
IPD	In-patient Department

IRCTC	Indian Railway Catering and Tourism Corporation
ISA	Information Sharing Agreements
MDDS	Meta Data and Data Standards
MeiTy / MeitY	Ministry of Electronics and Information Technology
MoSPI	Ministry of Statistics and Programme Implementation
MSME	Micro, Small and Medium Enterprises
NDSAP	National Data Sharing and Accessibility Policy
NIEM	National Information Exchange Model
NMDS 2.0	National Metadata Structure 2.0
NRSC	National Remote Sensing Centre
NSO	National Statistics Office
ODOP	One District One Product
PIMS	Personnel Information Management System
PLIDA	Person Level Integrated Data Asset
RCH	Reproductive and Child Health
SDLE	Social Data Linkage Environment
SFDB	State Family Database
SQAF	Statistical Quality Assessment Framework
StatCan	Statistics Canada
TNeGA	Tamil Nadu e-Governance Agency
ULB	Urban Local Bodies
UMIS	Higher Education database
UT	Union Territory

Executive Summary

India's vision of Viksit Bharat@2047 places evidence-based policymaking at the core of public administration, demanding that policies and decision-making be driven by credible, high-frequency, and granular data. The government recognizes that transitioning administrative data from a functional by-product into a strategic national asset is essential for future governance. Achieving this requires overcoming existing barriers to unlock the true governance value of the massive volumes of administrative data generated across the country. By building foundational pillars such as data cataloguing, machine readability, interoperability, sharing protocols, and privacy by design, governments can dismantle silos to improve programme targeting, reduce duplication, and pave the way for AI-driven decision-making.

Despite the rich multi-source data ecosystem in India-which includes administrative, census, sample survey, geospatial, and alternate data-the core problem is that administrative datasets are inconsistently defined, fragmented, and inherently siloed. It is predominantly collected to fulfill departmental compliance rather than to facilitate cross-sectoral analysis. To repurpose the datasets, machine-readable datasets need to be accompanied by their metadata to ensure that datasets are semantically understandable to machines.

A **four-stage consultative process** was launched to build a shared vision and actionable roadmap for harmonisation of administrative datasets at the State/ Union Territory (UT) level and for making them interoperable and AI ready. The stages and the processes followed have been listed below:

- **Stage I:** National agenda-setting on the theme by the Ministry of Statistics and Programme Implementation (MoSPI) as a follow-up to the 5th National Conference of Chief Secretaries in December 2025, which established data reuse as a national priority. Accordingly, a Concept Note was prepared and shared.
- **Stage II:** To sensitize States/UTs about harmonization, and interoperability of datasets, a National Level Consultative Workshop was organized on 24 February 2026.
- **Stage III:** State/UT level internal workshops with respective key line departments (Health, Education, Agriculture, etc.) and field functionaries on the theme to prepare an agenda for the National Deliberative Summit were organised as the third step. Mapping of existing systems and best practices and documenting the grassroot level constraints were proposed to be captured through a structured template circulated during the Stage II consultation.
- **Stage IV:** The National Summit to consolidate inputs and finalize the implementation roadmap has been scheduled for 29-30 April 2026 at Bhubaneswar, Odisha. This note has been created to facilitate discussions scheduled during the National Summit.

Inputs from internal state workshops reveal a complex landscape where data exchange heavily relies on ad-hoc, manual methods through files and emails rather than established institutional frameworks. While enabling policies and guidelines exist at the national level, their systematic adoption remains uneven at the State/UT level, highlighting the urgent requirement for

coordinated efforts for harmonization to build data sharing ecosystems encompassing all departments.

The internal consultations organized by States/UTs brought to light several critical hurdles hindering data harmonization and integration. These include technical hurdles characterized by data fragmentation, a lack of machine-readable formats, and the absence of common unique identifiers, metadata catalogues etc prevent interoperability and further usability of datasets. These are exacerbated by human resource constraints, such as shortage of skilled data stewards. Underpinning these issues are institutional hurdles, like lack of unified conceptual definitions and established linkage frameworks, alongside a cautious approach to data sharing, shaped by compliance considerations to various acts and policies which together influence data shareability and consolidation across departments in a seamless manner.

MoSPI and the Ministry of Electronics and Information Technology (MeitY) have come up with guidelines for managing the whole life cycle of data management from design and generation stage to dissemination stage. However, transitioning to secure interoperable ecosystems requires robust institutional architecture to design administrative data reuse programs and to take the journey through various stages of maturity level of data discovery, harmonisation, machine readability and interoperability. The National Summit is being organised in this context to establish the shared understanding of data harmonisation and action plan with timelines.

Following the respective Internal Workshops, States/UTs have shared the structured feedback on the best practices and the challenges in furthering the theme of Using Administrative datasets at the State/UT level. These inputs are the basis for the preparation of this Final Agenda Note for discussion during the National Summit. Further, consultations were made with Central ministries/departments, regulators, several researchers and think tanks on the theme to prepare this note. Governments both at the Centre and States along with institutions have successfully operationalized cross-departmental data linkage proving the massive potential of responsible data reuse programmes. The few examples are as detailed below:

- **State Successes:** Karnataka utilizes the FRUITS farmer registry and Kutumba social registry with hashed Aadhaar identifiers for presenceless DBT subsidies and beneficiary targeting. Tamil Nadu integrates Higher Education and School Education databases via Application Programming Interface (API) for automated scholarship eligibility (Pudhumai Penn), and maps municipal data with GST/EPFO to augment revenue. Rajasthan's Pehchan portal automates voter list and ration updates upon death registration via Jan Aadhaar linkages.
- **Central & International Models:** Central APIs like VAHAN (vehicle registration linked with insurance), DigiLocker and Digi Yatra (biometric passenger processing) demonstrate highly efficient data harmonization. Internationally, Norway's Altinn portal, Australia's PLIDA, and Canada's Social Data Linkage Environment (SDLE) showcase the power of register-first policies and secure data linkages.

- **Strategic Use Cases:** Some of the examples of potential use cases that can revolutionise governance include, urban property tax monitoring, early childhood milestone tracking (linking health, nutrition, and education data), multidimensional women's empowerment indices, MSME formalization intelligence, and predictive climate-health mapping etc.

Based on feedback received from State/UTs, the capabilities of States/ UTs (and their departments) can be evaluated across three core dimensions of data harmonisation, namely, Data catalogue, Data Consistency and Standards, and Integration Readiness[1]. To facilitate discussion and for the States/UTs to develop an action plan, each of these dimensions have been classified into three maturity levels of **Foundational, Intermediate, and Advanced**.

To facilitate States/UTs towards data harmonization and ultimately seamless data sharing for data reuse, a systematic, roadmap is proposed with *four phases.* Discussions on these are expected during the Summit to arrive at an agreed roadmap with timelines.

- **Phase 1: Inventory of Datasets:** The Nodal agency in the States/UTs prepares a consolidated list of datasets of all Departments and organisations of the States/UTs with details as provided in the template shared with States/UTs.
- **Phase 2: Foundational Maturity.** Departments may appoint data custodians, conduct internal dataset audits for compliance with data harmonisation guidelines, and formulate a State Administrative Data Compendium using the National Metadata Structure 2.0 (NMDS 2.0) structure. States may mandate MoSPI/MeitY guidelines via Government Orders, digitize manual workflows, introduce common identifiers for priority use cases, and establish a Data Strategy Unit (DSU) to drive the effort across Departments and organisations.
- **Phase 3: Intermediate Maturity.** States/UTs have dynamic, web-based State metadata portals acting as repositories of metadata. Validation rules are embedded at the data capture stage to ensure quality. States/UTs have Data Exchange Platforms for seamless data sharing.
- **Phase 4: Advanced Maturity.** The focus shifts to proactive and presenceless governance, where eligible beneficiaries are identified automatically. States will leverage micro-geospatial linkages for precise infrastructure planning, implement AI-driven simulations, enforce strict data minimization, and create ecosystems where downstream applications continuously enrich central databases.

To ensure timely action, milestones for the immediate future have been outlined for the transformation of data harmonization and data sharing ecosystem. By *December 2026* (Short-

[1]**Dimension 1: Data Catalogue:** Assessing whether states have consolidated, digitally hosted compendiums of administrative datasets or if they rely on isolated departmental lists.
Dimension 2: Data Consistency and Standards: Evaluating the adoption of MoSPI/MeitY harmonization guidelines, standard definitions, and data lifecycle approaches.
Dimension 3: Integration Readiness: Measuring the transition from manual data exchange to fully automated, API-driven real-time integration.

term), States/UTs may establish a DSU in each of the departments, create data catalogue in the NMDS 2.0 metadata structure and harmonize these datasets as per MoSPI/MeitY guidelines, release publicly available datasets in machine-readable formats, create data sharing policy in alignment with Government guidelines and automate data sharing using APIs for priority legacy datasets, and initiate role-based training, etc. By *December 2027*, the States/UTs may prepare dynamic data catalogues, significantly increase the number of harmonized datasets as per MoSPI/MeitY guidelines and utilise platforms such as AI Kosh, NDAP, IUDX, State owned data exchange platforms, Open Data for data storage, discovery and exchange. Finally, by the *end of 2028* (Long-term), States/UTs may ensure that all future data management is institutionalized and follows best practices of the industry. They should aim to have a Data Exchange Portal at the State/UT level. In the long term, all datasets may be generated at the source, data systems may be fully automated from generation to sharing, and all staff can receive role-based training in data management.

To serve the objective of having a consensus around the roadmap, discussions during the Summit will focus on the States presentations on the current preparedness and roadmap to address challenges. Various dimensions of data harmonisation leading to interoperability of datasets and the various technological aspects which will be useful to drive administrative data reuse at State/UT level will also be discussed. There will be sessions on leveraging data reuse programmes for Governance which will showcase best practices of States/UTs. A session on Artificial Intelligence (AI) Readiness will facilitate States/UTs to take steps towards harmonisation and interoperability for AI readiness. The Sessions of the National Deliberative Summit will be concluded by a panel discussion to finalise the roadmap and way forward.

Thus, this summit is an initiative to formulate a comprehensive data harmonisation roadmap aligned with the global best practices. As a next step, the States/UTs may follow this roadmap. A coordination mechanism will be institutionalised to achieve these milestones in a time-bound manner. Ultimately, this shared vision will fundamentally transform administrative data architecture into a powerful engine for precise programme targeting, efficient public resource management, and outcome-driven governance.

1. Introduction and Background

India's vision of Viksit Bharat @2047 requires evidence-based policymaking to be at the core of public administration, ensuring that policies and public investments are guided by credible, high-frequency, and granular evidence. As the government seeks to leverage this information, it is also required to address the barriers in using multiple datasets seamlessly. The *core problem* is that while India generates massive volumes of administrative data, these datasets are inconsistently defined, and are inherently fragmented and siloed. Data is typically collected for departmental compliance purposes. To repurpose the information for cross-sectoral analysis, the data and metadata descriptions, common identifiers and classificatory codes need to be standardized to improve semantic understanding of datasets to the machines. Recognising these limitations and to unlock the data's true governance value, there is an urgent need for the harmonization and integration of these datasets. This realization formed the genesis of the current initiative, receiving its critical push during the *5th National Conference of Chief Secretaries* held in *December 2025*, which established using administrative datasets by enhancing interoperability as a national priority.

1.1 What is Administrative Data?

Administrative data refers to information collected as a matter of routine in the day-to-day management or supervision of a scheme, service, or revenue system. Unlike survey data, which is collected periodically and at a cost, administrative data is generated continuously as a by-product of governance. It covers beneficiaries, transactions, facilities, personnel, assets, and outcomes across every sector of public administration.

India generates one of the largest volumes of administrative data in the world ranging from health and nutrition registers, to welfare beneficiary lists, land records, school enrolment data, civil registration, and financial transactions. The challenge is not volume. It is coherence, consistency, and usability.

1.2 Why Harmonisation?

Administrative datasets have mostly been designed for a single purpose: operational management. A health department maintained immunisation registers to track coverage; a welfare department maintained beneficiary lists to process payments. The data served its purpose within the department. The reuse of data was not so much on the horizon especially when produced by a different department.

With the changing priorities, policymakers are now asking cross-cutting questions: Are the children who are malnourished also out of school and not receiving welfare transfers? Which districts show high vaccination coverage but still have poor child health outcomes and why? Which beneficiaries are enrolled in multiple schemes and which are excluded from all? These questions cannot be answered from any single department's data or a dashboard. They require linked data from multiple sources enabled for advanced analysis. And analytics readiness requires harmonisation.

A second, equally important shift is the arrival of AI and automated analytical systems. AI tools, unlike human analysts, cannot infer missing context, work their way through inconsistent definitions, or work around undocumented changes in data structure. They require explicit metadata, stable data models, and machine-readable provenance. Data that is perfectly adequate for a human analyst reading a report may be completely unusable by an AI system trying to answer a policy question.

The Cost of Not Harmonising

The consequences of fragmented, unharmonized data are not abstract. They show up as:

- Contradicting official statistics from different departments covering the same phenomenon, with no way to explain or resolve the difference
- Welfare beneficiaries being invisible in one register and duplicated in another, leading to both exclusion and leakage
- Inability to track the same person or household across health, education, and welfare systems which makes integrated outcome measurement impossible
- Significant duplication of effort, as each department conducts its own surveys to fill gaps that harmonised administrative data could close.

1.3 Steps towards data harmonisation

To address this problem, the theme *"Using Administrative Data for Governance: Harmonising Departmental Data at the State/UT Level"* was conceptualized to shift the perspective of administrative data from a functional by-product to a strategic national asset. Harmonising this data involves building foundational pillars-data discovery, interoperability, sharing protocols, curation, governance, and dissemination. By establishing trusted, machine-readable datasets with an overlay of interoperability layers with concepts of 'privacy by design' without compromising departmental autonomy, the Departments can overcome silos to improve programme targeting, reduce duplication errors, and prepare the ecosystem for AI-driven decision-making.

To systematically build a shared vision and an actionable roadmap for data harmonisation and integration, a *four-stage consultative process* was envisioned, starting with national agenda-setting by MoSPI following the 5th National Conference of Chief Secretaries in December 2025 (Stage I); advancing through a National Level Consultative Workshop on 24 February 2026 to sensitize state/UT stakeholders and establish technical frameworks (Stage II); cascading into State/UT-level internal consultative workshops to capture grassroot level realities, operational challenges, and priority use cases (Stage III); culminating in the consolidation of inputs for the National Summit in April 2026 to finalize the implementation roadmap during the summit (Stage IV).

As per the four-stage consultative process, the 24th February 2026 National Consultative Workshop was organised by MoSPI. It aimed to sensitize States/UTs on the scope of data harmonisation and to establish a shared understanding of how to transition from "data silos to data synergy". Through structured thematic sessions, the workshop explored real-world applications of data in decision-making, the legal and technical building blocks for

operationalizing data linkages (such as API frameworks, data anonymization, and machine readability), and successful state-level use cases from Karnataka, Rajasthan, Tamil Nadu, and Odisha. Ultimately, it provided clear guidelines and a roadmap for States/UTs to conduct their internal consultations and prepare for the agenda of the April 2026 Summit.

Following the National Workshop, States/ UTs were to conduct internal consultative workshops. These internal consultations were explicitly designed to engage line departments as well as field-level functionaries to ensure discussions were grounded in operational reality. The primary objectives were to map existing data systems, assess current readiness for data harmonization and document grassroots constraints (such as duplicate data entry and manual processes) in using administrative datasets as available with different departments. To prioritise cross-departmental datasets and use cases that could generate immediate value had been the other objective of the internal workshops.

MoSPI also participated in the internal consultations conducted by the States/UTs. This included in-person as well as virtual participation in certain states, ensuring that the discussions were well-guided and grounded in the national vision for data harmonization to ensure interoperability, giving clarity to the context of the summit and data harmonisation initiatives.

This final agenda note consolidates the inputs gathered throughout this consultative journey to present a clear snapshot of current status of administrative data harmonization. The note details the shared understanding on the critical grassroot level and institutional challenges hindering harmonization, showcase successful case studies, and propose an actionable roadmap for implementation with timelines. Ultimately, this roadmap aims to transition governments from fragmented compliance reporting to actionable, outcome-driven decision making. To ensure standardisation and clarity of the concepts discussed throughout this note, a detailed glossary of terms is placed at Annexure 1: Glossary.

Building upon this foundational understanding of why data harmonization and interoperability is necessary, the next section transitions directly into the "Current Situation and State/UT Overview," providing a detailed analysis of the documented state inputs, their technical readiness, existing data sharing practices, and the institutional setups currently driving data governance at the state/UT level.

2. Current Situation and State/UT Overview

India today operates within a rich and complex multi-source data ecosystem. Broadly, this ecosystem comprises: (i) administrative data generated by government departments through implementation of laws, regulations, and schemes; (ii) census data providing comprehensive but periodic snapshots of population, agriculture, and economic structures; (iii) sample surveys offering detailed analytical insights at defined intervals; (iv) geospatial data capturing the physical and spatial dimensions of development; and (v) emerging alternate data sources such as digital transactions, mobility patterns, and platform-based information, etc.

While each of these data sources is valuable in its own right, their true potential lies in their combined and complementary use. As governance increasingly demands responsiveness and real-time insights, interoperable administrative datasets across departments emerge as a critical pillar for strengthening evidence-based decision-making at the State /UT level.

Recognising the need, several foundational building blocks for a harmonised data architecture are already in place. These include enabling policy frameworks for data dissemination and sharing, nationally notified standards for metadata, standard codes and classifications, and identifiers, quality assessment frameworks, and multiple discovery and dissemination platforms. Collectively, these provide the legal, technical, and institutional scaffolding required for effective data use. A brief on Government guidelines and initiatives may be seen at Annexure 2: Government Guidelines.

Despite these enablers, the systematic linkage and use of administrative data remain uneven across states and sectors. Most administrative datasets are designed for operational purposes, with limited scope for interoperability, restricting their reuse for analysis and planning. Although states hold large volumes of high-frequency administrative data, mechanisms to routinely reuse datasets across sectors are limited, and existing efforts towards interoperability are largely sector-specific or through targeted initiatives rather than routine system-wide practice. A few illustrative examples of how administrative data can be made interoperable for the purposes of reuse across departments to support governance objectives are provided in Chapter 4.

The following sub-section draws upon the insights, self-assessments, and use cases provided by States/UTs during their internal consultative workshops as well as during the national consultative workshop, highlighting the current landscape of administrative data linkage, institutional readiness, and operational challenges.

2.1 Summary of State Inputs

Inputs from internal State-level workshops and discussions during the State/UT Workshop held on 24th February 2026 reveal a complex data ecosystem where administrative data is largely maintained in departmental silos. While most states possess rich, high-frequency digital data, the lack of common standards, metadata practices, and interoperability prevents routine cross-departmental data reuse programmes.

In most states, data exchange continues to rely on ad-hoc, manual methods such as files or emails, heavily dependent on personal relationships rather than institutional frameworks. Concerns regarding data privacy, confidentiality, and a lack of role-based access control frequently result in "institutional resistance" to sharing data across departments.

Some states report that data is often maintained in fragmented formats, standalone software, or physical registers, and is updated annually rather than in real-time. Poor data quality due to manual entry errors, inconsistent definitions, and missing identifiers remains a widespread challenge at the grassroots level.

However, several states have successfully operationalised cross-departmental data interoperability to improve targeting, monitoring, and efficiency and initiated robust dashboards and registries. Karnataka utilizes the Kutumba social registry and FRUITS farmer registry to unify beneficiary data. Tamil Nadu leverages the State Family Database (SFDB) and Makkal Portal to track scheme eligibility. Rajasthan utilizes the Pehchan Portal for civil registration. Bihar has launched the Viksit Bihar Strategy Room to consolidate data. Some of the examples are detailed below:

- **Karnataka (Kutumba & FRUITS):** The state developed *FRUITS*, an integrated farmer registry that connects with the Bhoomi land records system and digital crop surveys. This linkage allows multiple departments (Agriculture, Horticulture, Fisheries) to provide paperless, cashless DBT subsidies to farmers without repetitive manual verifications. Additionally, the *Kutumba* social registry links individual records using a hashed Aadhaar identifier to proactively identify eligible beneficiaries for housing and welfare guarantee schemes while weeding out the ineligible beneficiaries.

- **Tamil Nadu (State Family Database & Revenue Augmentation):** Tamil Nadu established the State Family Database (SFDB), integrating the PDS database with 336 schemes across 62 departments. For scholarship schemes like *Pudhumai Penn* and *Tamil Pudhalvan*, the state successfully integrated Higher Education (UMIS) and School Education (EMIS) databases via API for automated eligibility verification without physical applications. Additionally, the Greater Chennai Corporation cross-mapped municipal property data with GST, EPFO, and electricity (TNEB) databases, successfully identifying unassessed commercial properties and boosting Profession Tax demand by over 10%.

- **Rajasthan (Pehchan Portal and Jan Aadhar):** The Civil Registration System (Pehchan) is actively linked to the Jan Aadhaar family database, Food & Civil Supplies, and Social Security systems. Real-time API linkages ensure that when a death is registered, the deceased individual is automatically removed from voter lists, PDS ration distribution, and pension beneficiary lists, saving government resources and reducing fraud.

- **Odisha (SPDP & Krushak Odisha):** Odisha implemented the Social Protection Delivery Platform (SPDP), which acts as a "Single Source of Truth" by integrating 75

social sector scheme databases. This unified registry enables cross-scheme de-duplication and automated exit for households crossing eligibility thresholds. The state also created *Krushak Odisha*, a comprehensive agricultural intelligence database to facilitate transparent DBT disbursements for farmers.

- **Uttar Pradesh (LAKSHYAM, FactoryID, & FamilyID):** Uttar Pradesh is utilizing integrated data systems to solve administrative bottlenecks. This includes *LAKSHYAM*, a unified performance system assessing frontline workers (Lekhpals) using data from multiple portals; *FactoryID*, a dynamic database merging industrial data (Factories Act, Udyam, etc.) to capture unregistered manufacturing units; and *FamilyID* to auto-enroll eligible citizens for welfare.

- **Nagaland (PIMS):** The Personnel Information Management System (PIMS) tracks all state government employees and acts as a central data hub. Through API integrations, PIMS shares verified service records with the Health Department for the CM Health Insurance Scheme, the Finance Department for automated ePaybill salary disbursals, and the Vigilance Department for online clearances, radically reducing clerical errors and manual paperwork.

- **Uttarakhand (Devbhoomi Parivar Yojana):** The state's Planning Department is developing a Family ID system to seamlessly link beneficiary data across schemes. This identifier will act as the foundational layer to consolidate information, reduce duplication, and enable targeted service delivery.

- **Tripura (VAHAN, e-Way Bill & Scholarship Management):** The state effectively tracks vehicle and goods movement in real-time by linking VAHAN (vehicle registration), SARATHI (driving licences), and e-Way Bill data across the Transport and Taxes departments, reducing manual processing and improving tax enforcement. Additionally, for scholarship management, Tripura uses the NSP 2.0 and BMS portals to cross-verify Aadhaar and ration card records across Tribal, SC, and OBC Welfare departments through a hierarchical digital workflow, preventing ghost and duplicate beneficiaries.

- **Other State Initiatives:** States like **Andhra Pradesh** have utilized data integration for urban youth surveys and revenue monitoring, while **Maharashtra** is building a District-Level Economic Intelligence Platform combining official, administrative, and spatial data to track formal and informal economic production. **Punjab** is focusing on operationalizing its State Data Policy through a coordinated data standards ecosystem for seamless transactional and analytical data exchange.

2.2 Institutional set up at States/UTs

Transitioning from ad-hoc, relationship-driven data sharing to a secure, interoperable ecosystem requires robust institutional architecture. Based on the State/UT consultations, the institutional mechanisms required to champion administrative data reuse programmes span

policy mandates, multi-tiered governance committees, specialized data units, and dedicated personnel roles.

1. **Empowered Nodal Agencies and the Evolving Role of DES:** Traditionally, data collection has been fragmented across departments. There is a need to position the Directorate of Economics and Statistics (DES) and Planning Departments not just as statistical agencies, organising Sample Surveys and Censuses, but as central data authorities. Technical integration requires specialized IT infrastructure, the most successful institutional setups pair the DES with empowered e-Governance agencies.

    - For example, Tamil Nadu has designated the Tamil Nadu e-Governance Agency (TNeGA) as the nodal empowered agency under its State Data Policy, while Karnataka relies on the Centre for e-Governance (CeG) to technically anchor its social registry.

2. **Multi-Tiered Data Governance Committees** To address departmental fears regarding loss of control, privacy, and accountability, some states have institutionalized structured Data Governance Committees to oversee access and resolve disputes:

    - Karnataka instituted a comprehensive Data Governance Structure comprising:
        1. A Data Governance Steering Committee acting as an appellate body for policy and government-level decisions regarding data governance involving multiple departments.
        2. A Data Governance Executive Committee to handle decisions around data ownership, structural changes, and impact.
        3. A Data Governance Working Group at the operational level to examine the purpose of specific data requests and authorize sharing.
    - Need for Central Mandates: Bihar, however, reported a need for formal State-level policy or central administrative order for data dissemination highlighting sharing protocols, privacy requirements and legal safeguards, as few departments remain hesitant in releasing their data due to privacy concerns.

3. **Defining Accountability: Chief Data Officer (CDOs), Custodians, and Stewards**: A persistent challenge is the lack of clarity on data custodianship versus access. To establish clear lines of responsibility, States/UTs are creating specialized data roles within individual departments:

    - Tamil Nadu has mandated the designation of a CDO in each department to authorize access and ensure adherence to governance protocols.
    - Tripura identified the appointment of a designated "Data Officer" in each line department as a short-term actionable step to maintain the departmental data inventory and catalogue
    - Bihar recommended appointing Departmental Data Custodians equipped with the skills to handle data standardization, access control, and inter-departmental coordination.

- Himachal Pradesh outlined actionable steps to designate Data Stewards who will be directly responsible for maintaining the accuracy and completeness of departmental data catalogues.

4. **Capacity Building and Contractual Safeguards on Harmonization:** Finally, institutional setups remain weak without skilled human resources. Across the board, states highlighted severe shortages in technical capacity and digital proficiency among grassroots functionaries. States have requested sustained support for regular training programs on data lifecycle management, privacy protocols, and AI applications.

    - Bihar recommended making appropriate arrangements for "regular training and capacity building on multidimensional aspects of data governance, including data quality assurance, harmonization standards, and security protocols". They explicitly mentioned training for department officials on interoperability standards and digital platforms.
    - Tripura explicitly requested support for the "capacity building of departmental data managers, IT teams and field-level functionaries on NMDS 2.0 metadata documentation, Statistical Quality Assessment Framework (SQAF) quality assessment, common identifiers and data interoperability standards through structured training programmes.

Thus, to further the objective of ensuring harmonization among administrative datasets, a comprehensive institutional set up is important in each of the States/ UTs.

3. Key Challenges and Gaps

The effective use of administrative data through cross-departmental linkage requires focused deliberation on several interrelated issues. During the internal consultative workshops, the States/ UTs highlighted the following key challenges:

Technical Challenges:

1. **Uneven Adoption of Standards and Metadata Practices**: Although nationally notified standards for metadata, classifications, etc are available, their adoption across States/UTs and departments remains uneven, thus limiting interoperability. Some examples from State consultations are,
    - Karnataka noted inconsistent definitions across its systems, such as the definition of "In-Patient Department (IPD)" differing between the Health Management Information System (HMIS) and hospital-level systems.
    - Uttarakhand highlighted that if one department defines a "family" or "beneficiary" differently from another, consolidated data usage becomes impossible.
    - Kerala noted that metadata and documentation are not uniformly available, making it difficult for external users to interpret datasets.

2. **Limited and Inconsistent Use of Common Identifiers**: Limited and uneven use of identifiers for people, enterprises/assets and locations restricts linkages across datasets. It prevents cross-sectional insights and longitudinal analysis of outcomes at the beneficiary, household, or enterprise level.
    - Bihar flagged "missing identifiers" as a prominent challenge across almost ten major departments.
    - Kerala, Nagaland and Sikkim also reported that the absence of common identifiers prevents the mapping of welfare schemes across departments.
    - Without common identifiers, states like Karnataka struggle to track beneficiaries or cases seamlessly across health and education portals.

3. **Absence of a Data Linkage Framework**: In practice, unique identifiers are often personal identifiers, leading to inhibitions within departments to share datasets due to privacy concerns and accountability for any misuse. However, to answer policy or programme questions, appropriate safeguards can be introduced as per the Digital Personal Data Protection (DPDP) Act. The absence of a guiding framework on data sharing results in inconsistent practices, delays in data sharing, and underuse of administrative data for analysis and governance.
    - Bihar highlighted that there is often a lack of clear communication regarding which specific fields of a dataset need to be shared, resulting in poor quality data being transferred and further confusing the receiving departments.

4. **Data quality and credibility concerns**: If issues relating to data quality are adequately addressed in administrative datasets, their credibility for use in statistical or policy

analysis can be greatly enhanced. Although data quality frameworks such as the SQAF have been notified, their systematic and sustained application across administrative datasets remains uneven. State inputs indicate some grassroots quality issues:

- Karnataka reported that errors are largely caused by manual entry and a lack of validation checks.
- Nagaland and Himachal Pradesh cited poor data quality stemming from inaccurate responses, manual updating, and inconsistent record-keeping.
- Uttarakhand also noted that data collected at the field level varies in quality, causing discrepancies when tracking indicators for aspirational districts.

5. **Lack of Machine Readability**: There is a lack of data availability in machine-readable formats, and even where they are available, they are not shared through standardised APIs.
    - Himachal Pradesh reported that departments maintain data in different formats, including paper registers, standalone software, and Excel sheets, which makes interoperability nearly impossible.
    - Bihar echoed this, pointing to improper field and file formats preventing digital handshakes.

Institutional/Legal Challenges:

6. **Data Fragmentation and Siloed Systems**: Administrative data is spread across multiple departmental systems and portals, often without a unified mechanism for discovery, interoperability, or use. Most datasets are designed to support programme implementation within a department, rather than analysis across sectors such as education, health, nutrition, livelihoods, or urban services. As a result, even where data exists digitally, cross-departmental linkage requires manual reconciliation or special one-off exercises, limiting routine and scalable use of administrative data for governance. State consultations heavily validate this -
    - Kerala and Sikkim reported that data is kept in silos.
    - Himachal Pradesh observed that multiple databases are created for similar purposes, with department-specific MIS platforms lacking any interoperability.
    - Bihar specifically identified the absence of a single unified platform and unstructured data schemas as primary barriers across its line departments.

7. **Data Sharing, Privacy, and Governance**: Balancing the legitimate need for data sharing and reuse with privacy protection, security, and accountability remains a key governance challenge. Current access mechanisms are often manual, ad hoc, or relationship-driven, limiting scalability and trust. In the context of increasing data volumes and sensitivity, the absence of clearly articulated and consistently applied safeguards - such as role-based access controls, audit trails, encryption, and secure data environments - further constrains responsible data access and reuse.

- Bihar and Kerala cited data privacy and confidentiality concerns, noting that line departments are reluctant to share data because of fear of violating privacy norms.
- Rajasthan also highlighted strict compliance with the DPDP Act and heavy penalties as a reason department are hesitant to share data without clear legal exemptions.
- Nagaland confirmed that data sharing between departments remains an entirely manual practice.

8. **Organisational Priorities**: Data management and dissemination are not core functions of administrative departments, affecting the reuse of data as a strategic asset for evidence-based governance.

Human Resource Constraints

9. **Institutional and Capacity Constraints**: Dedicated institutional arrangements and skilled human resources for data governance and management are unevenly developed across States/UTs departments. Almost all consulted states reported a severe lack of technical capacity.
    - Himachal Pradesh, Nagaland, Sikkim and Kerala cited critical shortages of digitally proficient field functionaries, trained IT personnel, and basic digital infrastructure (such as hardware and internet).
    - Bihar also noted language barriers (Hindi-English data entry issues) as an operational constraint.

These challenges were also reiterated by knowledge partners working with States on administrative data linkage. They highlighted that the technical barriers they faced primarily involved highly fragmented and siloed data systems that lack common identifiers, standardized metadata, and API integrations, requiring extensive harmonization to reconcile disparate schemas, definitions, and timeliness mismatches. Institutional and legal hurdles critically impeded the progress, as organizations struggle to access granular administrative or unit-level datasets due to the absence of unified data-sharing agreements, evolving privacy regulations, and decentralized data governance. Consequently, engaging individually with multiple data custodians or bypassing access restrictions on public portals is often a time-consuming process with limited success. Furthermore, resource constraints exacerbate these difficulties, with critical skills gaps in data curation and modelling, limited computational infrastructure for handling large datasets, and a heavy reliance on tedious manual effort and external consultants to clean data and maintain pipelines.

While many States have functioning platforms which work with linkages of multiple datasets, there are no effective data harmonization programmes across Departments/Organisations. Concerted efforts towards having a data catalogue for each of the States/UTs and graduating to a maturity level of data curation with data modelling and API driven data sharing are envisaged in near term to keep up with the changing landscape of data usage for decision making.

4. Central Ministry and International Best Practices

This section brings out the initiatives on administrative data linkage-ranging from Central Ministry API integrations like Digi Yatra, to highly efficient international models. This can help in identifying actionable strategies that reduce duplication, improve targeting, and enhance citizen-centric service delivery.

4.1 Central Ministry and Private Sector Best Practices

4.1.1 IRCTC-Food Delivery Platform Integration (Indian Railways)

The Indian Railway Catering and Tourism Corporation (IRCTC) enables authorised food service platforms such as Zomato to integrate with railway reservation systems through APIs. The integration uses datasets from the Passenger Reservation System and using PNR, it provides information on train schedule, coach and seat details. These datasets are combined with vendor-side information such as restaurant (on station city), order processing time, and delivery logistics to coordinate food delivery at selected stations during scheduled halts. The collaboration demonstrates how operational railway datasets can be harmonised with private service platforms to enhance passenger services.

4.1.2 Forest Survey of India Integration with NRSC Satellite Data

The Forest Survey of India (FSI) integrates satellite imagery and geospatial datasets generated by the National Remote Sensing Centre (NRSC) with its own field-based forest inventory data. These datasets are processed using GIS tools to map forest cover and monitor changes over time. The integrated information supports the preparation of the India State of Forest Report and informs forest management and conservation planning.

4.1.3 Motor Insurance Verification and VAHAN Vehicle Registration Data Platform

The VAHAN platform, maintained by the Ministry of Road Transport and Highways in collaboration with State Transport Departments, consolidates vehicle registration data from Regional Transport Offices across the country. The system integrates datasets such as vehicle registration number, vehicle class, fuel type, manufacturer and model, engine specifications, registration date, and jurisdiction of registration. At the same time, motor insurance platforms such as Acko General Insurance and Go Digit General Insurance retrieve vehicle and policy information by querying authorised national databases using the vehicle registration number as the reference key. Information on existing or past insurance policies is accessed through the central motor insurance repository maintained by the Insurance Information Bureau of India. The harmonisation of vehicle registration and insurance datasets through authorised API access enables insurers to auto-populate vehicle details, verify insurance status, and generate renewal or quotation services efficiently.

4.1.4 Digi Yatra Biometric Passenger Processing System

Digi Yatra enables seamless passenger processing at airports by integrating Aadhaar-based identity verification, boarding pass data, and AI-powered facial recognition systems. Interoperability is enabled through digital identity frameworks such as DigiLocker and Aadhaar, allowing passenger identity and travel information to be securely linked for authentication at airport checkpoints. The system follows a consent-based, voluntary participation model, with privacy-by-design safeguards including encryption standards and automatic deletion of biometric data within 24 hours after departure. Digi Yatra is currently used for domestic travel and is being developed to integrate with airline systems, airport infrastructure, and international identity frameworks to enable seamless future expansion to international travel.

4.1.5 DigiLocker API Integration Journey

Launched as a flagship initiative under MeitY in 2025, DigiLocker is a prime example of successful administrative data linkage and interoperability at scale. Instead of relying on strict formal mandates, the platform expanded through strategic partnerships, with the DigiLocker team acting as an active implementation partner for government agencies.

The platform's architecture is built around an API gateway that enables secure, real-time document exchange between verified issuers (such as school boards and road transport departments) and requesters (such as banks or passport offices). Crucially, for government agencies that lacked the technical infrastructure to issue and store digital records independently, DigiLocker provided a dedicated repository layer to host their documents. By standardizing APIs, prioritizing high-frequency documents like academic transcripts and driving licenses, and assisting states with technical capacity gaps, DigiLocker successfully dismantled departmental data silos. Today, it acts as a trusted digital bridge, seamlessly linking thousands of institutional entities to issue and verify nearly 10 billion documents for over 685 million citizens.

4.1.6 Model Context Protocol (MCP) Server

The most recent addition to MoSPI's framework is the Model Context Protocol Server which is a technical component that enables standardised, machine-actionable context exchange between data repositories and AI systems. In plain terms, the MCP Server is what allows an AI tool to ask 'what does this indicator mean, where does it come from, and how should I interpret it?' and receive a reliable, machine-readable answer. This is the final link in the chain from data to AI-enabled policy insight and it enables users to get answers from their own environment without going through the search and access route.

4.2 International Best Practices

4.2.1 The Nordic Register-Based Statistical System (Norway, Denmark, Sweden, Finland)

The Nordic countries represent a mature "register-first" model where costly traditional surveys have been widely replaced by harmonised administrative data. This system relies on universal identification numbers for persons, businesses, and dwellings, enabling seamless, deterministic linkage across all government registers. Administrative systems are intentionally designed upfront for statistical reuse to ensure harmonisation at the source. For example, Norway utilises the Altinn portal, a centralised digital platform that ensures businesses and agencies report data once in strictly standardised formats. Additionally, Norway maintains KLASS, a centralized source for standard classifications that is integrated directly into administrative systems via open-access APIs. A major milestone for Norway was the 2011 Register-based Census, which replaced costly traditional surveys with statistics derived entirely from existing administrative data, saving time and public funds.

4.2.2 Australia: Whole-of-Government Data Integration and Modernisation

Between 2015 and 2019, Australia embarked on a comprehensive modernisation project to transform its statistical system into an integrated national data ecosystem. Led by its national statistical office, Australia systematically integrated administrative datasets alongside survey data to create massive, interlinked cross-departmental systems, most notably the Person Level Integrated Data Asset (PLIDA) and the Business Longitudinal Analysis Data Environment (BLADE). Australia also pioneered the Statistical-Geospatial Framework (SGF), seamlessly embedding geospatial integration across the entire data production cycle.

4.2.3 United States: Economic Measures and Semantic Interoperability

Operating within a highly decentralised federal structure, the US frequently leverages "Fiscal Federalism"—tying federal grant funding to strict data standard compliance—to harmonise administrative data across states, as seen in the T-MSIS for healthcare and the Highway Performance Monitoring System. To achieve data linkage across disparate administrative domains (such as Justice, Health, and Public Safety), the US implemented the National Information Exchange Model (NIEM). This semantic interoperability framework ensures that different departments share common, standardized definitions for critical information, allowing disparate systems to seamlessly exchange data.

4.2.4 Canada: Quality Assurance and Secure Data Linkage

Canada maintains a highly integrated data system driven by Statistics Canada (StatCan), with standardisation enforced through strict Conceptual Standards (like the Standard Geographical Classification and standard definitions for complex metrics) and Technical Standards (such as machine-readable formats, SDMX exchange protocols, and ISO 11179 metadata standards). A key operational mechanism is the use of Information Sharing Agreements (ISAs), which act as technical gatekeepers by mandating compliance with data structuring and cleaning protocols before departmental data can be linked. Furthermore, Canada operates a Social Data Linkage Environment (SDLE) that utilises a Derived Record Depository (DRD)-a secure national

repository of unique individuals compiled from tax, birth, and immigration records—to effectively audit, verify, and link incoming provincial administrative datasets.

4.2.5 Mexico: Institutional Integration and Geospatial Information

Mexico provides a distinctive model where statistical and geospatial functions are institutionally unified within a single organisation, INEGI. In this system, geospatial data is embedded across the entire statistical lifecycle, with all major data collection exercises and administrative records redesigned to capture geospatial coordinates right at the source. Mexico also created a centralized National Spatial Data Infrastructure (SDI) to host, integrate, and disseminate interoperable geospatial and statistical data.

These case studies prove that secure, interoperable data ecosystems can be successfully institutionalized at scale, yielding massive efficiency gains and public savings. Extracting the core principles from these successful implementations provides the necessary evidence base to formulate a structured, phased approach, which is detailed in the actionable roadmap for States/UTs.

5. Potential Use Cases

To transition from fragmented compliance reporting to integrated, outcome-driven decision-making, it is essential to build high-impact use cases that bridge existing departmental silos. Based on the inputs from think tanks and non-state organisations, some of the use cases have been listed at Annexure 3: Potential Use Cases, where data harmonisation can create immediate, practical value for state/UT-level planning, targeting, and service delivery.

While these potential use cases can be taken up depending on the requirement on ground, the importance of harmonisation of administrative datasets is the preliminary step. Basically, harmonisation through standards ensures how datasets are documented, processed, and are made reusable to enable reuse across departments and states. To detail further, harmonisation will ensure that the datasets being created should be able to be replicated based on the associated metadata documentation. Since the outcome of any potential linking of datasets depends heavily on having clean, standardised, and aggregated data beforehand, guidelines on SQAF, Operational guidelines for managing the data lifecycle, etc. assume high importance. While multiple datasets are used for generating an indicator, there should be enough documentation delineating the logic and steps involved along with clarity on definition of measures/indicators.

These use cases are provided only to prompt discussion within States/UTs to prioritise some of the use cases depending on the priority in respective states/UTs. To demonstrate that these conceptual frameworks are not only viable but already yielding tangible results, the next section examines established best practices and successful real-world case studies.

6. Maturity Assessment and Actionable Roadmap and Timelines

To enable a systematic transition from fragmented, department-centric data systems to a harmonised and interoperable data ecosystem, this section outlines a structured approach combining maturity assessment, a phased roadmap, and clearly defined timelines.

6.1 Maturity Assessment and Good Data Practices

6.1.1 Maturity Assessment

Institutionalizing data harmonization for ensuring interoperability in different States/UTs requires a clear and objective understanding of current operational capabilities of each State/UT. To establish the baseline insights, the feedback gathered in the shared template during the respective internal State/UT workshops were used and were synthesized into a maturity assessment. This framework evaluates states/UTs across three core dimensions-Data Catalogue, Data Consistency and Standards, and Integration Readiness-categorizing their current data ecosystems into *Foundational, Intermediate, or Advanced maturity levels*, as described in *Table 1.*

Table 1: Maturity framework for administrative data harmonization and integration

Dimensions	Stages leading to maturity		
	Foundational Maturity	Intermediate Maturity	Advanced Maturity
Data Catalogue	There is no consolidated State/UT-level compendium of administrative datasets. At best, individual departments maintain their own isolated, internal lists of datasets.	A formally notified State/UT-level compendium exists and serves as a single consolidated reference across departments. It is periodically updated, prioritises data-sharing, and follows the NMDS 2.0 metadata structure.	The dataset compendium is hosted on a digital platform that includes specific release dates, and individual departments are responsible for regularly updating their own entries. The details are exposed through APIs.

Data Consistency and Standards (harmonization)	While standard definitions (like geography or beneficiaries) might exist, they are applied inconsistently. Only a few systems follow the MoSPI/MeitY harmonisation guidelines for classifications and unique identifiers.	The MoSPI/MeitY harmonisation guidelines and data quality frameworks are adopted by most departments. The MoSPI guidelines for managing statistics across the data lifecycle are integrated as an ongoing programme by States/UTs.	A comprehensive data life cycle approach is embedded directly into system design and data capture processes. New systems automatically adopt common definitions, and validation rules are applied digitally to ensure consistent data coding across all departments over time.
Integration Readiness	Systems operate completely independently. Data exchange is either entirely manual (via files or emails) or involves data being digitally extracted but transferred and processed manually.	Most major systems share standard classifications. Data is exchanged through defined technical interfaces (or basic electronic exchange for specific use cases), though record matching still requires some manual reconciliation.	Systems are deliberately designed around common identifiers for real-time integration. Data schema documentation is done and data exchange is fully automated (e.g., utilizing APIs), allowing records to be continuously matched across multiple systems without manual intervention.

Maturity assessments of departments and organisations of the respective States/UTs in various dimensions of data harmonisations were also part of their internal workshops. A template with various levels and their descriptions was shared with States/UTs to prepare documentation on the maturity levels during the respective internal workshops that were scheduled in the month of March and April 2026 on the theme of using administrative datasets at the States/UT level. Based on these internal workshops, it was expected that the self-assessments by States/UTs would pave the way for identifying and prioritising State/UT specific action plans in furthering the objective of the theme.

6.1.2 Good Data Practices

The States/UTs may adopt good data practice principles, which go beyond statistical accuracy and encompass the full data lifecycle - ensuring that data is discoverable, interoperable, and interpretable. It can be done through clear definitions, robust metadata, and contextual documentation. It requires end-to-end traceability from source to output, availability in

machine-readable formats, and adherence to rigorous quality assurance standards. The list of good data practices is mentioned in Annexure 4: Good Data Practices.

6.2 Roadmap for States (Four-Phase Strategy)

Transitioning from the current fragmented landscape to an advanced, AI-ready data ecosystem requires a systematic, *phased approach* tailored to varying levels of institutional capacity. This actionable roadmap outlines a strategic four-phase strategy, guiding States/ UTs from establishing foundational data catalogues and standard definitions, to embedding automated consistency, and ultimately achieving real-time integration and proactive governance.

Phase 1: Inventory of Datasets

The prerequisite for the States/UTs to begin moving up the maturity ladder is having an **inventory of datasets**. The Nodal agency in the States/UTs should prepare a consolidated list of datasets of all Departments and organisations of the States/UTs with associated characteristics (like, accessibility, storage format, metadata availability, API readiness etc.). A template in this regard has been shared with States/UTs.

Phase 2: Foundational Maturity

Goal: *Establish basic digital inventories, and introduce standard definitions.*

Dimension 1: Establishing the Data Catalogue

- **Prepare a Centralized Compendium at State-level:** The State Directorate of Economics and Statistics (DES) or equivalent nodal agency to circulate the NMDS 2.0, the standard metadata template for referential metadata. Departmental lists must be consolidated into a formally notified State Administrative Data Compendium to provide immediate visibility into what data exists. Departments can compile an internal inventory of all existing datasets, documenting the fields, storage formats, update cycles, and data owners.
- **Designate Data Custodians/Stewards:** Every line department can appoint a dedicated Nodal Officer or "Data Custodian/Steward" responsible for data quality, standardisation, and intra-departmental coordination.

Dimension 2: Enforcing Data Consistency and Standards

- **Adopt National Guidelines:** State governments may issue formal Government Orders (GOs) mandating the adoption of MoSPI/MeitY harmonisation guidelines (such as NMDS 2.0, SQAF, standard classificatory codes, unique identifiers, operational guidelines on data management lifecycle, etc).
- **Improving Data quality:** Data quality, documentation, and clarity on limitations are critical enablers of responsible linkage and reuse. To support informed use, data quality rating for administrative datasets may be considered, drawing on existing frameworks such as SQAF. Such a signal would help users assess fitness-for-purpose when linking

datasets, for example, distinguishing exploratory use from datasets suitable for high-stakes policy or financial decisions.

- **Establish Change Management through versioning:** Departments may document the current definitions and formats they use for datasets to identify where harmonization is needed. It is also required to document any changes to definitions, coverage, or methodology (with dates from which it is applicable) that are being introduced in the datasets as part of the harmonisation initiatives to help the data users.
- **Document Data Schemas:** Formally document the data structure for each priority dataset: what the entities are, what the variables are collected, what the valid values are, and how the schema has changed over time.

Dimension 3: Building Integration Readiness

- **Identify Priority Use Cases:** Rather than attempting full state-wide data interoperability, States/UTs may identify high-impact, specific datasets based on priority use cases (e.g., linking education and health data for child tracking, or unifying farmer registries) to pilot basic electronic data exchange.
- **Digitize Manual Workflows:** Departments may migrate any remaining paper-based or email-based data exchange processes to basic digital formats (like csv, json or xml).
- **Introduce Common Identifiers:** Begin integrating established common identifiers-such as Aadhaar, Family IDs, or standard location codes-into the priority departmental systems.
- **Data Sharing policy:** States/UTs can publish a Data Sharing Policy.

Cross-Cutting Enablers:

- **Vendor Mandates:** To prevent future silos, any system integrators or IT vendors engaged by departments can be contractually bound to follow the State-approved data harmonisation Policies when building new systems.
- **Establishing Data Boards/Governance mechanism:** The States/UTs may form a cross-departmental committee comprising IT, planning, and legal experts to resolve data-sharing disputes, oversee privacy compliance under data protection laws, enforce activity deadlines, review and approve the annual State/UT Data Readiness; resolve inter-departmental disputes on standards and definitions; commission technical working groups on specific harmonisation challenges; and report progress on an annual basis.
- **Establish Data Clinics:** States/UTs may establish Data Clinics (small mobile team) which may comprise of experts from various institutions and organisations. These experts might support State/UT departments in creation of metadata records, or support an IT team in exposing a dataset via API etc.
- **Establish Data Strategy Units:** To promote a culture of data management in the evolving data governance ecosystem, MoSPI advocates the NITI Aayog initiative of establishing DSUs for effective coordination within Departments. These cross-functional units are envisioned to oversee data governance, and champion

harmonization. The Nodal Agency in the State/UT to furthering the objective of the theme may be given the with following responsibilities in the respective States/UTs:

1. Nodal Agency can clarify coordination structures across departments and levels of government.
2. Assess the maturity of each of the departments and organisations in the States/UTs in the dataset generation, harmonisation and in ensuring integration readiness and to encourage continuous discussion.
3. Sector- or theme-based working groups can be set up to decide on the requirement of sector specific data reuse programmes.
4. To assess the resource needs in the respective States/UTs.
5. Identification of knowledge partners for advisory and handholding support. Making provision for a shared pool of technical expertise also may be thought of.
6. A framework for states to assess their own resource needs.
7. Mechanisms for sharing tools, templates, and implementation experiences within the States/UTs and also to report back to the relevant Central Ministries/Departments and also to MoSPI.
8. To take up the action points emanating out of this Summit.

Phase 3: Intermediate Maturity

Goal: *Transition from static lists and partial data exchange to a fully automated, API-driven, and "real-time" interconnected ecosystem.*

Dimension 1: Dynamic Data Catalogue

- **Deploy a Unified Digital Platform:** Migrate the static state compendium to a dynamic, web-based cataloguing through State Metadata Portals. The data catalogue is exposed through APIs for seamless sharing.
- **Decentralized Updation with Accountability:** Departments can be mandated to update their metadata and data entries on the digital platform at defined frequencies.
- **Link Data to Funding:** To enforce compliance, States may link the timely production and updating of digitalised administrative datasets following harmonisation guidelines with the release of departmental resources or funds.
- **Annual Audits:** Institutionalize annual audits of the data catalogue to ensure completeness, quality, and continued alignment with national metadata frameworks.

Dimension 2: Embedded Data Consistency and Standards

- **Redesign Legacy Systems:** Departments should overhaul legacy IT systems to embed harmonized definitions, standardized codes, and validation rules directly into the initial data capture processes. Data Modelling for each of the dataset curation can completely automate the process.

- **Automate Data Quality Checks:** Implement a State-wide Data Quality Framework (following the principles of SQAF released by MoSPI) that utilizes digital, automated validation checks to flag errors and ensure consistent interpretation of data across all departments over time.
    - *Semantic Interoperability:* Develop a central data dictionary or ontology that ensures different IT systems computationally understand the exact context of the data being shared (e.g., ensuring "income" means the exact same metric mathematically across the housing and welfare departments).

Dimension 3: Real-Time Integration Readiness

- **Data Exchange through National Platform:** Platforms such as AI Kosh, NDAP, IUDX, Open Data, and the Account Aggregator (from Private) can be actively utilised for data storage, discovery, and sectoral exchange in accordance with their respective mandates.
- **Automate Record Matching:** Transition from manual reconciliation to automated cross-verification mechanisms.
    - *Establish Data Sandboxes:* Before live API connections are established between sensitive databases, IT departments are provided with a "sandbox" (a secure, isolated testing environment) to test data models, security protocols, and integration stability safely.

Cross-Cutting Enablers for Long-Term Success:

- **Implement Clear Dissemination Policies:** Issue a central administrative order clearly defining what data can be shared, what must be restricted for privacy, and who is authorized to share it, thereby removing the ambiguity that causes institutional resistance.
- **Continuous Capacity Building:** Establish regular, institutionalized competency-based capacity building programs for field functionaries and IT staff on data management, quality assessment, metadata standards, and security protocols etc.

Phase 4: Advanced Maturity

After having functional data exchanges, the focus may shift from basic interoperability to proactive governance, predictive analytics, and highly secure architectures. Provided below are certain initiatives that advanced states can undertake to elevate their data ecosystems for evidence-based governance:

1. Interoperable data systems through State data exchange platforms: Advanced States/UTs may develop a secure, centralized State Data Exchange Platform designed with API-based interoperability, allowing systems to "talk" to each other continuously. This will help bring together data across departments into a unified, interoperable data system. Rather than replacing existing departmental databases, it will complement them by enabling seamless data exchange through APIs, ensuring that updates in source systems are automatically

reflected. In this process, the ownership of data shall remain with the respective departments. States/UTs can adopt different modes of data exchange, depending on technical feasibility and policy needs.

2. Create Data Compounding Ecosystems: Advanced states/UTs should ensure that data linkage operates as a continuous loop where downstream applications constantly enrich the core database. This data compounding effect continually increases the long-term value, accuracy, and richness of the central database.

3. Make datasets AI ready: In the context of the use of Artificial Intelligence for decision making, the data harmonisation and integration readiness can be taken to the next level for making the datasets AI ready. The technical requirements for AI readiness will be a little more extended. One conceptual view is given as below:

Layer	What it does
Discovery	A machine-readable data catalogue allows AI systems to find available indicators and datasets automatically, without human assistance.
Semantics	NMDS 2.0 combined with a National Statistics Office (NSO) Vocabulary expresses the authoritative definition of every statistical concept in a form machine can read.
Delivery	Data structure definitions clearly separate dimensions, measures, and attributes enabling machines to query data correctly through APIs.
Integration	Context files provide semantic mappings and formal links between field names and their statistical meanings so machines could do meaningful integration.
Governance	NSO-led, NMDS-centric version control ensures that as definitions evolve, machines always access the correct version.

4. Transition to Proactive and Presenceless Governance: Advanced States/UTs can move beyond requiring citizens to manually apply for services. By leveraging API-linked ecosystems, States/UTs can proactively identify eligible beneficiaries and deliver entitlements automatically. For example, Karnataka's FRUITS registry enables paperless, cashless, and presenceless service delivery by integrating farmer data directly with Direct Benefit Transfer (DBT) and financial management systems.

5. Leverage Micro-Geospatial Linkages for Infrastructure Planning: States can refine their geographical data linkages to a hyper-local level (with privacy by design principles) to enable highly precise, data-driven policy decisions. Rajasthan has proposed linking its unique family identifier (JanAadhaar) with "digipin" technology to achieve a highly accurate 4-by-4 metre grid of location coordinates. This allows the state to map the exact spread of the population

against existing infrastructure (like roads and streetlights) to precisely determine the optimal placement for new hospitals or educational institutes.

6. Implement AI-Driven Simulations: Instead of merely using data for retrospective monitoring, advanced states can utilize AI-based tools for proactive policy simulation and forecasting. Furthermore, States/UTs can establish dedicated Data Analytics Units (DAUs) to integrate internal administrative data with external data sources - such as survey data or alternate data-for a more well-rounded and robust statistical system.

7. Enforce Data Minimisation and Advanced Privacy Architectures: As data exchange deepens, advanced States/UTs may adopt highly secure data architectures that protect citizen privacy. Karnataka's "Kutumba" social registry achieves this by maintaining a strict emphasis on data minimisation and relying only on single "authoritative sources of truth". Rather than storing actual Aadhaar numbers across multiple databases, the system automatically converts the Aadhaar into a secure 64-character hashed value. The original number is kept in a secure vault without the possibility of backtracking, while the hashed reference number safely links records across more than 30 state portals.

By following this roadmap, states/UTs can systematically eradicate isolated manual processes and create secure data ecosystems that continuously enrich public administration. To ensure accountability and track progress against this ambitious roadmap, the next section outlines the specific deliverables, timelines, and expected outcomes from the national summit.

6.3 Activity List and Timeline

To ensure accountability and operationalize the four-phased strategy discussed in the previous section, it is essential to establish clearly defined action points and milestones for States/UTs and nodal departments. The following table outlines this specific activity list, setting targeted objectives to guide the ecosystem's transformation over the next three years.

Table 3: Activity list and timelines for States/UTs

Timeline	Activity List
December 2026 ***(To be achieved in the next 6 months)***	**Institutional Setup** 1. Establish a governance system for data ecosystem management like data boards, DSUs, coordination committees, data clinics etc. to strengthen the use of administrative data for policy and service delivery. 2. Develop and adopt data-sharing policies aligned with the Government Guidelines. 3. While developing new applications and databases, the standardization of datasets based on guidelines of MoSPI/MeitY may be ensured. https://egovstandards.gov.in may be referred for

	standards related to MDDS, API design standards, and certificate formats. A clause to this effect be put in RFP as well as in work order, wherever, any vendor is onboarded for application development. **Harmonization Activities - Data Cataloguing** 4. Create an inventory of all datasets by States/UTs covering all departments and organizations. 5. Initiate the formulation of a department-wide data catalogue in the NMDS 2.0 format for priority datasets. **Harmonization Activities - Data Classifications and Standards** 6. Ensure harmonization (metadata, unique identifiers, standards and classifications and quality assessment following the MoSPI/MeitY guidelines) of priority datasets. 7. Identification of priority datasets (including legacy datasets) as per the identified use case. 8. Undertake quality assessment of priority datasets using SQAF to ensure continuous improvement in data management practices 9. Release Advance Release Calendar for all data outputs **Harmonization Activities - Integration Readiness** 10. Release all publicly available datasets in machine readable formats. 11. Document Data Schemas for priority datasets 12. Initiate automated data sharing through APIs **Capacity building** 13. Initiate role-based training programme and competency-based capacity building programmes based on identified priorities.
December 2027	1. Continue implementation of activities to strengthen data management systems **Harmonization Activities - Data Cataloguing** 2. Develop and operationalise a dynamic web-based data catalogue and data schemas for all datasets. **Harmonization Activities - Data Classification and Standards** 3. Increase the number of harmonised datasets to include all datasets in a phased manner, aligned with MoSPI/MeitY guidelines and evolving data needs. 4. Ensure change management practices are implemented for all datasets. 5. For legacy systems, the standardization of data sharing should be established through lightweight data wrappers or adapters. **Harmonization Activities - Integration Readiness**

	6. Each State should publish personal documents (including citizen-facing certificates, records, and credentials) on DigiLocker and provide meta data in a standard format. Similarly, Entity Locker on business related documents be published with metadata. 7. Platforms such as AI Kosh, NDAP, IUDX, Open Data, and the Account Aggregator (from Private) should be actively utilised for data storage, discovery, and sectoral exchange in accordance with their respective mandates. **Capacity Building** 8. Continue role-based training programme and competency-based capacity building programmes based on identified priorities.
December 2028	**Harmonization activities** 1. Sustain and institutionalise improvements in data management practices. 2. Ensure all data management programmes align with established best practices in industry. 3. All datasets required in regular departmental functioning are identified and made interoperable, including the legacy datasets. 4. Ensure datasets are generated at source and systems are fully automated from data generation to sharing. 5. Each State can have a Data Exchange Platform of its own or it can choose to ride exclusively on the National Data Exchange Platform. 6. Private Data Exchange Platforms should be encouraged to bring private investment, utilise the Government Data Exchange Platforms and offer value-added services. **Capacity Building** 7. Institutionalise role-based training and competency-based capacity building for all staff involved in data management.

States/UTs may take necessary action to adhere to these targeted milestones-ranging from formulating initial data catalogues and automating data sharing via APIs by December 2026, to achieving fully automated, source-generated data ecosystems by December 2028.

7. Conclusion and Way Forward

The transition towards evidence-based, outcome-oriented governance is a fundamental prerequisite for realising the national vision of Viksit Bharat @2047. Currently, States/UTs face persistent challenges in the harmonisation, integration, and use of administrative data, primarily because this information is maintained in departmental silos characterised by inconsistent definitions, manual processes, and fragmented IT systems. To unlock its true governance value, it is imperative to shift the perspective of administrative data from a mere by-product of departmental compliance to a strategic national asset.

The data harmonisation roadmap formulated for States/UTs by MoSPI is anchored in global guidelines established by multi-national agencies such as the UN, World Bank, and IMF. Specifically, the roadmap and its action points are closely aligned with the latest UNECE Guide to Data Integration for Official Statistics. By adopting these international best practices, the framework ensures that India remains synchronised with evolving global trends in data integration.

As States/UTs progress along this roadmap, a critical next step will be the transition towards unified data systems and making data AI ready. Such systems enable the convergence of data across departments, levels, and sources into a unified yet interoperable architecture. By leveraging API-based linkages and appropriate integration approaches, a unified data system can reduce data silos, improve data quality, support a more comprehensive, real-time view of governance outcomes and build a decision support system - thereby strengthening evidence-based decision-making.

In parallel, Central Ministries may lead by example by leveraging existing administrative data to demonstrate high-impact use cases of data integration. By institutionalising data-driven decision-making within central schemes and programmes, they can create replicable models and standards for States/UTs, thereby accelerating adoption and ensuring alignment across levels of government.

To ensure the effective implementation of this roadmap, MoSPI will institutionalise a coordination mechanism to systematically track and facilitate the progress of States/UTs against their time-bound activities.

Ultimately, the diligent execution of this shared roadmap will eradicate isolated manual processes, bridge existing departmental silos, and establish a highly secure, AI-ready data compounding ecosystem. This framework will fundamentally transform India's administrative data architecture into a powerful engine for precise programme targeting, efficient public resource management, and outcome-driven governance.

Annexures

Annexure 1: Glossary

Glossary Terms	
Administrative Dataset	Information collected by government departments during their daily operations such as beneficiary registries or health records etc. rather than through specialized surveys. These are live datasets that provide a low-cost, real-time look at how public services are reaching citizens on the ground.
AI Readiness	A state where data is harmonized and machine-readable, allowing AI systems to understand context and make responsible, reliable decisions with safety guardrails.
Attribute	A descriptive label or characteristic attached to a data point that explains the quality, status, or nature of that data, but does not identify it.
Comparability	The ability to compare data across time, regions, or countries using standardized concepts and methodologies.
Consistency	The degree to which data values are logically aligned and non-contradictory within a single system or dataset. It ensures that related pieces of information (like a total and its parts) are identical across different tables or storage locations.
Context File	A special companion file published alongside a dataset that tells computers what all the codes and references in the data mean, and where to find related information. It is a machine-readable guide to the dataset.
Data Asset	Any collection of data that has value and needs to be managed, tracked, and governed. Eg. - survey, a dataset, a report, a database table
Data Catalogue	An organised inventory of all data assets held by an organisation that helps users find the data they need. It contains information about the source, content, quality, access conditions, and update frequency of each dataset.
Data Coherence	Data Coherence means ensuring that statistical information remains logical when compared over time or across different sources.
Data Documentation	The recording of information about a dataset, including its content, structure, collection methodology, concepts and definitions used, data quality, and conditions of use, in a manner that enables users to interpret and use the data correctly.

Glossary Terms	
Data Governance	A formal system of rules, responsibilities, and processes that decides who can define, change, access, or approve data and its definitions across an organisation. It ensures everyone uses the same definitions and standards.
Data Harmonization	Data harmonization is the process of ensuring that data collected from different sources, departments, regions, or countries are: • Comparable • Consistent • Coherent • Compatible with international standards
Data Harmonization Importance	• Enables international comparability • Supports SDG monitoring • Improves data credibility • Facilitates policy coordination • Reduces duplication and inconsistencies across ministries
Data Interoperability	The ability of different government systems to "speak the same language" and exchange data seamlessly using common standards.
Data Modelling	The process of designing how data will be organised and structured, deciding what the rows, columns, and variables of a dataset will be. A good data model makes data easier to share, compare, and analyse across systems.
Data Provenance	The complete record of where data came from and every change made to it including sources, collection methods, processing steps, and any modifications that allows users to assess the reliability and fitness for use of the data.
Data Quality	A comprehensive measure of how well data serves its intended policy purpose. High-quality data must meet several critical standards simultaneously: • Core Dimensions: It must be relevant to the problem, timely enough to act upon, and complete enough to represent the whole population. • Reliability & Clarity: Information must be gathered through consistent methods and presented in a way that is easily understood by decision-makers.

Glossary Terms	
	• Coherence & Comparability: The data must be linkable across different departments and consistent over time to track long-term progress. • Machine Readability: Data must be stored in formats (like CSV or JSON) that computer systems can process automatically.
Data Revision	An official, documented update to a previously published data value, made because better information became available or an error was corrected. A revision policy specifies in advance the schedule and circumstances under which revisions will be made.
Data Standardization	The process of developing, establishing, and applying consistent rules, definitions, formats, and codes for data across systems and organisations, to enable coherent interpretation, comparison, integration, and exchange of statistical information.
Dimension	The identifying characteristics that together uniquely pinpoint a single data value in a dataset. Dimensions form the key of a dataset. Common dimensions include time period, geography, sex, age group, and sector.
Machine-Readable Format	A file format that a computer programme can automatically read, process, and analyse without human intervention. A PDF or scanned image is NOT machine-readable. CSV, JSON, or XML files are machine-readable.
Metadata	Data that define and describe other data. Metadata includes information about the concepts, definitions, classifications, data sources, collection methods, data processing, and quality of statistical data.
Official Statistics	Official Statistics are numerical data produced and disseminated by government agencies or authorized national bodies. It provides a reliable, impartial, and scientifically rigorous evidence base for a society's economic, social, and environmental situation. Unlike raw departmental data, Official Statistics undergo rigorous quality checks to ensure they are a trustworthy "public good" for making high-stakes decisions.
Open Licence	A licence that grants permission to access, use, modify, and redistribute data or other resources with minimal or no restrictions so that anyone can legally reuse the data without seeking separate approval.
Programmatic Access	The provision of data in a manner that allows automated software applications to query, retrieve, and process the data directly through a defined interface (such as an API), using standardized protocols and formats, without requiring manual intervention.

Glossary Terms	
Quality is Institutionalized and Transparent	Institutionalized means quality standards, checks, and processes are formally embedded into the rules, systems, and procedures of every organisation that produces statistics so that quality assurance happens consistently and automatically at every stage, regardless of who is responsible at any given time. Transparency means that when data is published, information is provided of how it was collected, what methodology was used, what the coverage is, and what the known limitations are.
Reliability	The degree to which data is accurate, consistent, and verifiable at the source, ensuring it remains a trustworthy basis for automated decision-making and policy.
Traceability	The ability to follow the complete journey of a data value from the original source to the final published output, so that at any given stage the lineage of every published value can be established and verified.
Unique Identifier	A permanent, unique code assigned to each data asset, variable, or concept which enables seamless integration by accurately linking individual records across different departmental datasets.

Annexure 2: Government Guidelines

A. Guidelines on harmonization of datasets by MoSPI:

1. **National Metadata Structure (NMDS):** Enhanced discoverability of Government Datasets: MoSPI has designed National Metadata Structure (NMDS 2.0) with the objective to bring in improved understanding of various data sets in the public domain, such as census, surveys, administrative and those from other sources. Also, there are standards and policy prescriptions from MeitY for Meta Data and Data Standards (MDDS).
2. **Compilation of national and international classification:** MoSPI has compiled a list of 87 major global standards and 10 national standards to progress towards the Global level standardisation.
3. **Compilation of unique identifiers:** In order to enhance the interoperability of government databases, a set of common identifiers used frequently in the collection and compilation of administrative data has been identified. Till now, 17 such items have been identified.
4. MoSPI has developed a **mechanism for the reconciliation of divergent datasets**. The said mechanism, along with a detailed Standard Operating Procedure (SOP) for referring data divergences to the NSO for their harmonious resolution, has been established and placed in the public domain.
5. A **Statistical Quality Assessment Framework (SQAF)** has been finalised and circulated to all Ministries/ Departments and States/UTs after taking up a pilot exercise by the states of Rajasthan, Odisha, and central Ministries, namely, DoSEL and MSME, and in consultation with NITI Aayog and MeitY. It provides a comprehensive framework, having 19 Principles and 85 requirements, covering possible aspects of the coordination of the statistical system, i.e., the legal support and provisions, the institutional arrangements and mechanisms.
6. **Compendium of Datasets and Registries in India:** MoSPI has published the latest edition of the Compendium of Datasets and Registries in India, 2024, a key initiative aimed at strengthening data accessibility and informed decision-making. As part of the ongoing modernisation of the National Statistical System, this compendium ensures that government data is easily accessible for policymakers, researchers, academicians, students, analysts, businesses, and the general public.
7. **Operational guidelines on managing official statistics across the data lifecycle:** It provides guidelines to ensure responsible, consistent and effective use of government data while protecting privacy and confidentiality in alignment with international standards.
8. **eSankhyiki Portal and API Framework:** eSankhyiki is MoSPI's integrated data portal, providing a centralized interface for accessing official statistics. For AI-readiness, the portal's data structure definitions which distinguish clearly between dimensions, attributes and measures (the values) and align with international SDMX standards . It provides a replicable model for administrative data departments as without

structured and standardized data model it would not be possible to make data analytics ready, leave aside AI readiness.

MoSPI has also released API guidance and working APIs with documentation, enabling controlled, machine-readable data access. The goal is for administrative datasets, once harmonised, to be accessible via standardised APIs that any analytical system can query.

B. Guidelines on Data Sharing and Privacy by MeitY:

1. **National Data Sharing and Accessibility Policy (NDSAP)** - is fundamentally an "Open Government Data" instrument designed to transform the legal status of non-sensitive government information from proprietary assets into "public goods." It mandates that all data generated using public funds must be accessible to the public by default. Legally, it employs a "Negative List" approach, where the onus is on individual ministries to justify confidentiality; if data does not fall under restricted categories (like national security or personal privacy), the policy compels its release under the Government Open Data Licence (GODL).
2. **Standardization of Data, Structural Metadata, APIs, Certificates and Registries:** Data and metadata standards (MDDS), API design standards, and certificate formats have been prescribed at https://egovstandards.gov.in for all applications and databases. At present, there are 23 Standards (viz. MDDS – Demographics (Person Identification and Land Region Codification), MDDS for Panchyati Raj Domain, MDDS for Health Domain, MDDS for Rural Drinking Water and Sanitation, etc), 25 Guidelines and 19 Frameworks, which have been notified. For legacy systems, the standardization of data sharing should be established through lightweight data wrapper or adapters as per MeitY guidelines.
3. **Open API Policy for eGovernance applications in India 2015:** A manual for creating standard APIs has been circulated among data producing entities to promote uniformity in data dissemination. The API policy of MeitY and NDSAP of DST can act as enablers for future AI-ready data ecosystems.

Annexure 3: Potential Use Cases

This annexure outlines a set of illustrative use cases that demonstrate the potential of harmonised administrative and alternative data systems to enable more responsive, data-driven governance across sectors at the State/UT level.

1 Climate, Health, and Adaptive Behaviour

- **Use Case:** Linking climate exposure, energy use, and health outcomes for extreme heat and climate risk management.
- **State/UT-level Utility:** States/UTs and districts can move towards proactive heat risk management by identifying the geographies and populations most vulnerable to heat stress. Integrated dashboards combining signals like temperature spikes, rising electricity consumption, and health system utilisation enable early warning systems and allow frontline workers to deploy targeted interventions during heat events.

2 Urban Governance and Revenue Systems

- **Use Case:** Property tax and revenue monitoring by integrating municipal property records, land ownership systems, valuation benchmarks, and spatial/GIS data.
- **State/UT-level Utility:** Enables Urban Local Bodies (ULBs) to identify inconsistencies in property valuation and detect revenue leakages. States/UTs can use this harmonised data to implement risk-based property re-assessment strategies, improve tax compliance, and strengthen municipal revenue planning and analytical capacity.

3 Human Development and Early Childhood Outcomes

- **Use Case:** Harmonising early childhood health (RCH), nutrition (Poshan Tracker), and education (UDISE+/APAAR) datasets to track shared child developmental milestones.
- **State/UT-level Utility:** Institutionalising these linkages allows states/UTs to proactively flag compounded vulnerabilities in children and automate service delivery-such as immediate Anganwadi enrolment at birth or paperless school admissions. Examples like Karnataka's Kutumba and Telangana's Samagra Vedika demonstrate how states can shift to a precision-based policy framework to ensure no child falls through the gaps.

4 Measuring Women's Empowerment and Economic Agency

- **Use Case:** Building a Multidimensional Women's Empowerment Index and generating deep insights into women's work outcomes by linking datasets like NFHS, PLFS, SRS, and NCRB.
- **State/UT-level Utility:** States can identify specific districts where low agency coincides with poor labour force participation or adverse demographic outcomes. This enables highly targeted scheme design—for instance, combining financial inclusion

with digital skills training in specific areas—and allows states to monitor the long-term impact of their interventions on women's employment outcomes.

5 Economic Measurement, Productivity, and Local Development

- **Use Case:** Linking Administrative and Alternative Data for District-Level Economic Intelligence. This involves integrating employment and productivity survey data (PLFS, ASI, ASUSE) with high-frequency administrative datasets (GST turnover, e-Way bills, Udyam MSME registrations, electricity consumption) and alternative spatial indicators (night-time lights, freight movement).

- **State/UT-level Utility:** States/UT can move away from relying on static estimates to establish continuous, high-frequency economic monitoring systems at the district level. With the help of district-level economic dashboards and spatial mapping, state and district administrators can identify high-growth clusters, spot lagging regions, and understand local sectoral composition. This allows planning bodies to prioritize investments, design sector-specific interventions, and actively monitor local economic performance and structural changes over time.

6 Integrated Measurement of Financial Inclusion

- **Use Case:** Building a composite Financial Exclusion Index using data from the RBI, NPCI, MoSPI, and other socio-economic sources.

- **State/UT-level Utility:** States/UTs can shift from measuring mere account ownership to capturing true financial participation. Harmonised data allows states to distinguish between infrastructure availability and behavioural gaps, helping them detect financially stressed populations despite high participation and enabling better targeting of financial literacy programmes.

7 MSME Growth, Employment, and Trade

- **Use Case:** Creating an employment intelligence layer by combining enterprise registration (Udyam), payroll (EPFO), and labour surveys (PLFS), alongside trade compliance data.

- **State/UT-level Utility:** States/UTs can generate a live, district-level "formalisation ratio" to reveal where MSMEs are growing but formal hiring is lagging, allowing them to precisely target skilling and labour compliance drives. Furthermore, states can collaborate with the DGFT's Trade Connect portal by feeding their Geographical Indications (GI) and One District One Product (ODOP) databases to attract global buyers to local MSMEs.

8 Environmental Regulation and Policy Evaluation

- **Use Case:** Utilizing real-time Continuous Emissions Monitoring Systems (CEMS) to run market-based pollution control mechanisms, like the Emissions Trading Scheme (ETS) in Gujarat.
- **State/UT-level Utility:** States/UTs can link real-time emissions data with permit registries and plant-level economic data to monitor compliance, trigger regulatory actions, and reduce overall pollution at the lowest economic cost. This architecture can be easily adapted by states for other pollutants and industrial sectors.

9 Understanding Movement and Migration

- **Use Case:** Using unique identifiers across administrative systems (railways, health, telecommunications, UPI) to plot the geographic movement of populations.
- **State/UT-level Utility:** Equips state/UT departments to understand the specific needs of incoming, outgoing, and within-state migrants. This granular, high-frequency map of people's movements allows states to plan infrastructure, dynamically adjust schemes, and allocate funds more efficiently.

10 Tourism Statistics

- **Use Case:** Estimating domestic tourist arrivals using alternative tracking methods rather than traditional accommodation or ticket data.
- **State/UT-level Utility:** States/UTs can leverage Big Data, AI, GIS, mobile phone locations, and transportation data to arrive at highly accurate, short-frequency estimates of state-level domestic tourism, successfully capturing social trips that account for the vast majority of total travel.

Annexure 4: Good Data Practices

A structured framework of "Characteristics of Good Data Practices", organised into:

- Core Attributes are foundational, non-negotiable for modern statistical systems for making data analysis, linking and AI ready. These are foundational requirements for a modern, credible, and internationally aligned data system.
- Advanced Attributes are maturity-enhancing, enabling automation, interoperability, and comprehensive AI readiness. These represent higher-order capabilities that position data as comprehensively AI ready.

The list of attributes includes:

I. Core Attributes

These are foundational requirements for a modern, credible, and internationally aligned data system

1. All data products are documented using recognized metadata standards.
2. The statistical concepts, classifications, and variables are available in a centralized, machine-readable repository.
3. A centralized data catalogue covering Indicators and Microdata is available
4. All catalogued datasets and their versions are assigned globally unique IDs.
5. Each dataset has a link to its full metadata record which prevents separation of data from documentation
6. Data modelling involves structuring data through well-defined dimensions and attributes that together uniquely identify each observation
7. Comprehensive public API with Access to data and metadata and Clear documentation are provided
8. All public data products are released under a clear, recognized license for access clarity.
9. Quality Assurance Framework aligned with international standards is available.

II. Advanced Attributes

These represent higher-order capabilities that position data as comprehensively AI ready

1. The catalog supports search by concept not just keywords for semantic understanding.
2. The catalog provides detailed, machine-readable data lineage tracing.
3. The data catalog provides for Advance release calendar.
4. A context file to route AI systems through URLs for machine understanding is available
5. Data structure file to provide information about indicators and their dimensions and filters are available in machine readable format.

*****