

OCTOBER 2026
OTT VELSBERG
VIVEK AGARWAL



From Digital Scale to Data Power: A Data Operating Model for India's States

Contents

3	Executive Summary
8	India's Digital Scale
10	Gaps in India's Government-Data Ecosystem
18	Building a State Data Operating Model
32	Building State-Level Data Governance
45	Preparing State Data for AI
48	Recommendations
61	Conclusion
62	Acknowledgments

Contributor: Ninad Sargar

Executive Summary

India has built government technology at a scale almost no other country can match. More than 1.4 billion citizens are enrolled in Aadhaar, the country's national identity system, while its Unified Payments Interface processes more real-time payments each month than most of the rest of the world combined, and a national data platform now hosts thousands of standardised data sets.^{1,2} Few countries have shown such capacity to build foundational systems that can reach hundreds of millions of people.

What India has not yet built consistently is a cross-government operating model that makes priority data reliably usable across existing systems, departments and levels of government. Indian states generate enormous volumes of administrative data every day: beneficiary lists, health records, land registers, school enrolments, tax filings. However, these data are collected to satisfy one department's reporting requirement, not to answer the cross-cutting questions that actually drive policy – which children are malnourished and also out of school? Or which beneficiaries are claiming the same scheme twice while others receive nothing at all?³ Answering those questions requires actionable quality data that are timely, reliable, holistic, secure and interoperable.

Data fragmentation is not an administrative inconvenience. It is a delivery failure, leading to fiscal leakage, exclusion from welfare schemes, misallocated resources, weak accountability and reduced trust in public institutions. For citizens, fragmentation can mean repeated registration and submission of documents, inconsistent eligibility decisions and difficulty correcting errors across systems. For frontline officials, fragmentation means collecting and reconciling the same information repeatedly while often lacking a usable view of the data needed to resolve a case or coordinate services. Fragmented data also limit the value of artificial intelligence. Without actionable quality data, AI will amplify the existing blind spots in government data and decision-making rather than solve them.

The Tony Blair Institute for Global Change's recommended approach is to establish a data operating model that enables more effective use of data across all layers of government. This model would define how government assigns mandate, ownership, standards, assurance, funding and decision rights. Within the model, a data operating layer provides the practical capabilities – the shared tools, standards and services – that allow authorised users and systems to discover, understand, trust, access, exchange, trace and correct data across government. The data operating layer may be implemented through centralised, federated or hybrid architectures; it is not synonymous with a single platform, data lake or exchange.

Several states are already operating real, working examples of harmonised data systems at meaningful scale: Karnataka's Kutumba Registry, Rajasthan's Pehchan Portal, Tamil Nadu's State Family Database and Odisha's Social Protection Delivery Platform, among others. Indian Prime Minister Modi has recognised data as a "national asset and a resource for the future" and called for "greater interoperability" and its "effective use". He has also emphasised that data silos are not "merely an organisational problem but also a problem of mindset" and urged officials to work with a "spirit of belongingness".⁴

The Ministry of Statistics and Programme Implementation (MoSPI) is currently spearheading a structured, time-bound national effort to address gaps in data quality, availability and use across government. Following the Fifth National Conference of Chief Secretaries in December 2025, MoSPI has run a multi-stage consultation – a national workshop, state-level internal workshops and a National Deliberative Summit in Bhubaneswar in April 2026. This culminated in a roadmap for states to identify and document their data sets, adopt common standards, and establish the responsibilities and systems needed to share and reuse data, with milestones through to December 2028. More recently, MoSPI released a Practitioner's Handbook on Data Harmonisation, providing guidance for officials managing government data sets on common definitions, documentation, quality and standards.⁵

Implementing the MoSPI roadmap should be treated as a governance and delivery reform, not as a technology programme. States should use MoSPI's framework as the national baseline, with clear responsibilities and everyday processes to ensure data are reliable, shared lawfully, and used to improve government decisions and services.

This paper sets out seven recommendations for building this kind of state data operating model. States should:

1. **Establish an empowered cross-government data function to mitigate the risk of fragmentation.** The central function would set rules, convene institutions, support implementation and maintain accountability.
2. **Build governed data capabilities around priority decisions.** This would produce reliable data for targeting, delivery, planning, monitoring, service improvement, accountability and safe AI adoption, and avoid the risk of being led by technology reform.
3. **Make ownership, stewardship and authoritative sources operational.** This would have the dual benefits of establishing accountability and turning it into continuous practice, mitigating the risk of data being shared, linked or used without clarity on meaning, quality, permission or accountability.
4. **Create independent data-quality assurance and safe external access.** This is particularly important in helping government to retain credibility where data are used for welfare eligibility, targeting, enforcement, health planning, education, disaster response or AI-enabled decision support.
5. **Establish legal, privacy and procurement safeguards for reuse, including clear legal gateways, retention rules, audit logs and procurement standards.** This would mitigate the risk of unlawful or unsafe reuse, cyber-exposure and vendor lock-in.
6. **Use a State Data Balance Sheet to manage critical data assets, risks and debt to make data governance visible to leaders and show where better use of data creates opportunities.** This would reduce government's tendency to fund new initiatives while the underlying data assets remain unreliable or outdated.

7. **Sequence reform around priority decisions and delivery use cases.**

This mitigates the risk of premature automation before state institutions can trust the underlying data.

These steps give states a practical route to better services, more effective use of public resources and accountable AI adoption. Progress should be measured by improvements in decision-making and in the outcomes delivered for people.

What This Paper Means by a State Data Operating Layer

The data operating layer is defined by function, not by a prescribed technical architecture. It should make priority data discoverable through common catalogues and metadata; establish authoritative sources and shared semantics; apply quality and validation rules; support secure and purpose-limited access and exchange; preserve lineage and auditability; and enable correction when source data are wrong. Different states may deliver these capabilities through different combinations of existing departmental systems, shared platforms, exchange services, analytical environments and federated components. The objective is not to move all data into one place, but to make the right data reliably usable for an authorised purpose.

Technical patterns can vary: centralised repositories, lakehouse approaches, federated exchange and hybrid architectures each solve different problems and can coexist. A **data lake** brings together large volumes of data from

different systems in a shared environment. A **lakehouse** combines this flexibility with stronger structure, quality controls and governed access. A **federated model** leaves data within departmental systems and connects them through application programming interfaces (APIs), catalogues, common identifiers and access rules.

The appropriate architecture will also vary by state. The policy test is the same: departments remain accountable for authoritative source data, while common standards and controls allow priority data to be reused safely for delivery, accountability, research and AI.

India's Digital Scale

India's digital public infrastructure is real and it works. Aadhaar has enrolled more than 1.4 billion residents and supported more than 24.5 billion e-KYC (electronic Know Your Customer) transactions. Linked to bank accounts and mobile numbers through the Jan Dhan-Aadhaar-Mobile (JAM) Trinity, it helped raise formal banking inclusion from roughly 25 per cent of the population in 2008 to more than 80 per cent by 2023.^{6,7} The Unified Payments Interface (UPI) now processes more than 20 billion transactions a month – an estimated 49 per cent of all real-time payments globally – and in 2024 alone handled roughly 172 billion transactions worth about \$3 trillion.⁸ Direct Benefit Transfer, built on this payments rail, is estimated to have saved more than \$60 billion by March 2025 through reduced leakage and fewer intermediaries.⁹

The data layer has grown alongside it. NITI Aayog's National Data and Analytics Platform (NDAP) hosts more than 5,000 data sets across 31 sectors with cross-data-set merge tools.^{10,11} The Ayushman Bharat Digital Mission (ABDM) has issued more than 960 million health IDs, with more than 1.1 billion health records digitally linked.¹² DigiLocker has issued and verified nearly 10 billion documents for more than 685 million citizens, and now operates as a standardised API gateway connecting thousands of institutional issuers and requesters – banks, passport offices, universities – without requiring every government agency to build its own document infrastructure.¹³

Efforts to make this data layer linguistically inclusive are also underway: Bhashini, the National Language Translation Mission, provides translation, speech-recognition and text-to-speech services across all 22 scheduled languages. Through its BhashaDaan initiative, it crowdsources voice and text contributions from citizens to build open data sets for historically undersourced languages such as Konkani and Bodo.^{14,15} As of March 2026, more than 10,000 contributors had joined its community platform, Bhashini Samudaye.¹⁶

Compared with the rest of the world, these achievements are not marginal. The World Bank estimated that a financial inclusion gain that took about six years would have taken close to five decades through conventional banking expansion alone.¹⁷ This remarkable speed of progress is clear evidence that identity, payments and data reinforce one another when built as shared public infrastructure rather than as isolated agency systems.

The same holds for payments. According to the International Monetary Fund's June 2025 report on retail digital payments, UPI is the world's largest real-time payment system by transaction volume; ACI Worldwide put its 2024 volume at roughly 172 billion transactions, a 49 per cent share of global real-time payment activity – more than Brazil, Thailand, China and South Korea combined.¹⁸

That transaction trail has itself become a policy lever: lenders now treat UPI payment histories as a form of digital collateral, extending cash-flow-based credit to micro-merchants, street vendors and gig workers, who had no conventional credit file for banks to underwrite against.¹⁹

This is also why the model travels. The World Bank, the United Nations and the G20 have each cited India's stack – digital identity, interoperable payments and emerging data-exchange layers such as the Account Aggregator framework – as a template other developing economies are now adapting, from Sri Lanka and Morocco to the Philippines and Ethiopia. This is alongside a growing range of international UPI and cross-border payment arrangements.

These figures demonstrate extraordinary reach and infrastructure adoption, but scale is not the same as effective service delivery. Measured against the markers of actionable quality data, this ecosystem speaks to only one – holistic – and even there it delivers scale rather than the full range of inputs, outputs and services that holistic data require. Enrolment, transaction volumes and digitally linked records do not themselves show whether data move effectively between institutions, whether frontline processes improve, or whether citizens can consistently access the services built on top. The next challenge is not simply to extend digital coverage but to make the data generated through these systems usable across government.

02

Gaps in India's Government-Data Ecosystem

Despite the sprawling data infrastructure at a national scale, India still lacks trustworthy decision-ready data for a truly data-driven policy ecosystem.

Data collection must be **holistic**, capturing the people, inputs (such as projects, auctions), outputs (such as human and development indicators) and services (such as certifications and subsidies). It must be **timely**, reflecting the current population, events and circumstances. It must also be **reliable**, providing an accurate account where incentives are aligned and data are collected and reported by skilled individuals. Verification mechanisms should be in place to confirm that what is reported accurately reflects what is happening, and whether inconvenient findings are challenged or withheld. It must be **interoperable**, understandable and reusable across departments and levels of government. And it must be **secure**, protected end to end across its lifecycle. This requires clear guidelines on who can access data and under what authority.

Failure points arise throughout the chain that transforms administrative information into trustworthy, decision-ready evidence, and a weakness in any dimension limits what can be achieved in the others. Below we outline the issues raised most consistently and how these gaps impact the design of the operating model set out later in this paper.

First, data are not holistic – they can exclude as well as include. Linking databases, identities and eligibility systems can reduce duplication and fraud, but it can also magnify the effects of incomplete records or incorrect classifications. Fragmentation can also impose additional time and administrative burdens on citizens even when they are not denied access to a public service or benefit: citizens may need to register separately for related programmes, resubmit information government already holds, prove the same eligibility criteria repeatedly, or navigate different correction processes when the underlying record changes.

The risk is particularly high where correction mechanisms and grievance routes are weak. A randomised trial in Jharkhand found that linking food rations to biometric authentication caused a 10 per cent reduction in benefits among the 23 per cent of beneficiaries who had not completed Aadhaar linkage, with 2.8 per cent losing access entirely. In Odisha, the state government acknowledged in a High Court affidavit that database errors had cut off food-security benefits for more than 2 million people.^{20,21}

Second, some foundational data are outdated. Where population baselines, household information or migration patterns are no longer current, welfare design and resource allocation become less accurate. People who have moved, become poor or out of work, entered a vulnerable category or changed household status may not be visible in the data used to serve them.

Census 2027 is now underway, with house listing scheduled during 2026 and population enumeration in 2027. Until the new results become available, many planning and entitlement baselines continue to rely on population data from Census 2011. Every major welfare programme, from the National Food Security Act to Pradhan Mantri Garib Kalyan Anna Yojana, allocates benefits using 2011 population figures, which means anyone who was born, migrated or became eligible since then risks being missed.²² In addition, millions of internal migrants are structurally undercounted because major surveys like the Periodic Labour Force Survey use house-based sampling that misses people in transit or temporary settlements.²³

Third, weak independence and transparency undermine data reliability. Data quality is not only a matter of accuracy at the point of collection. It also depends on whether methodologies, revisions, limitations and discrepancies can be scrutinised. High-impact statistics and administrative data sets need transparent methods, published quality statements and independent challenge. Without these, even technically accurate data may not command sufficient confidence to inform difficult public decisions.

The starkest case is Covid-19 mortality: India officially recorded roughly 530,000 deaths, but Civil Registration System data – delayed for four years before its release in May 2025 – implies approximately 2.4 million excess deaths in 2021 alone.²⁴ The undercount ratio varied enormously by state. Each episode compounds a credibility problem that outlasts the episode itself – once official numbers are shown to diverge sharply from lived experience, even accurate data become harder to use because nobody believes them anymore.

Similarly, departmental reporting creates incentives for distortion. When the same institution that collects data is also judged by them, the pressure to report success can be stronger than the incentive to report reality. This does not always mean deliberate manipulation. It can mean optimistic classification, incomplete reporting, inconsistent definitions, or weak data-quality checks. But the effect is the same. Leaders receive a managed version of reality.

Take, for instance, the 3,000 to 8,000 data elements that health facilities report every month into India's Health Management Information System (HMIS). Only about 10 per cent are ever converted into usable indicators and roughly half of all fields are routinely left blank.²⁵ The gap between what is reported and what is real can be dramatic. For instance, one state's HMIS recorded 728 infant deaths in April to June 2015 against 3,307 captured by its own parallel surveillance system.²⁶ Education data underwent a similar correction in 2022–23, when school reporting shifted from aggregate counts to a student-by-student basis: 14 million duplicate or “ghost” student entries disappeared from the system overnight.²⁷

For many administrative data sets, collection ultimately happens through a frontline interaction – in a Gram Panchayat (local government body) or urban local body, a school or health facility, or through officials and workers operating close to citizens. Data quality therefore depends not only on central standards but on the conditions of data capture: workload, clarity of definitions, form design, training, connectivity, device usability, validation and whether the information collected is returned in a form that helps frontline staff perform their work.

Three conditions deserve particular attention. Enumerators and frontline workers are not always equipped with adequate training or the skills needed to collect data reliably. Compensation structures rarely reward data quality, leaving little incentive to collect information carefully rather than quickly. And consent principles under the Digital Personal Data Protection (DPDP) Act will require the enumerators to be trained on what the act requires. A better operating model should reduce reporting burden, strengthen frontline training and improve the usefulness of data at source, not simply impose another reporting requirement.

Fourth, systems do not speak to each other. Departments build databases for their own mandates, using different identifiers, formats and standards. Even where there is political willingness to share data, the technical and semantic work required to connect systems remains expensive and repetitive. Cross-organisational data use becomes a series of bespoke exercises.

Interoperability has four dimensions. Technical interoperability enables systems to exchange data. Semantic interoperability ensures that the data have the same meaning. Organisational interoperability aligns roles, processes and responsibilities. Legal interoperability establishes the authority, purpose and safeguards for reuse. A state can have APIs and still fail if any of the other three are absent.

On technical interoperability, most administrative data in India are built to support each department's own programme, rather than cross-sector analysis, and even where digital systems exist, joining them across departments requires manual, one-off reconciliation rather than a routine, repeatable process.

NITI Aayog's NDAP is illustrative of this gap: by its own design, it hosts, standardises and merges data sets that departments have already published for external, largely research-based and public-facing use – and functions as a neatly organised and indexed database repository. However, a data layer requires live administrative systems that can exchange and act on each other's data. Digitisation and the availability of digital tools are an important precondition for that, but they are not the same as data usability.

MoSPI's consultations with states, conducted through structured internal workshops in early 2026, describe a remarkably consistent picture: in most states, data still move between departments as files and emails, dependent on personal relationships between officials rather than institutional processes.²⁸

On organisational interoperability, the problem is both horizontal and vertical. Horizontal interoperability concerns whether departments at the same level of government can exchange and interpret data. Vertical interoperability concerns whether data can move effectively between union, state, district and frontline systems, including back to the institutions responsible for acting on it. Reporting upwards is not sufficient if states and local officials cannot retrieve and use the resulting information for the services and decisions for which they remain accountable.

Social protection and public-service delivery span Central Sector schemes administered by Union Ministries, Centrally Sponsored Schemes implemented through states under shared financing and policy arrangements, and programmes designed and operated by states themselves. Each layer may have its own eligibility rules, registries, reporting requirements, identifiers, update cycles and technology providers.

As a result, the same household, worker, business or beneficiary can appear in multiple administrative systems that were created for different statutory or programme purposes and were never designed to operate together. Fragmentation is therefore institutional before it is technical. Better APIs cannot resolve unclear ownership, conflicting definitions, asymmetric incentives to share data or different accountability relationships between union, state, district and frontline institutions.

DigiLocker illustrates this dynamic. It has done more than most platforms to make issued documents portable, but it remains primarily a storage and retrieval layer: it does not itself adjudicate data quality, and each issuing department stays responsible for the accuracy of what it publishes there. Fields that sit within a single department's remit – ration entitlements within Civil Supplies, for example – are relatively straightforward to keep synchronised. Fields that cut across departments, such as pension or

employment records held elsewhere, are harder: they require agreement on which system is the source of truth for a given individual, and on how to bring departments that are not designated as the source of truth to accept that their version will not be the one relied upon.

This need not mean centralising the underlying data. Several departments are understandably wary that participating in data-sharing means losing control of their own records. Exchange mechanisms – where each department keeps custody of its systems and shares only defined fields under agreed rules – are one way to address that concern: NITI Aayog's Data Empowerment and Protection Architecture and the National Data Governance Framework's state data-exchange platforms are both built on this principle, granting access through consent-based, auditable channels rather than pooling data into a single store. Framed this way, participation can look like a benefit – faster verification, fewer duplicate submissions for a department's own beneficiaries – rather than a loss of control.

Finally, security should be treated as an end-to-end operating requirement rather than a feature of a platform. Centralised repositories can concentrate risk, while fragmented systems can produce inconsistent controls and unmanaged access. States therefore need risk-based architecture, data minimisation, strong identity and access management, encryption, segmentation, audit logging, continuous monitoring, incident response and tested recovery across the full data lifecycle.

India's record of government data-security incidents shows how exposed these systems already are. In 2022, 50 government websites – central and state – were hacked, and eight separate data-breach incidents were traced to government organisations. India has national cyber-security institutions, government-entity security guidelines and sectoral requirements, but the framework for state-government cyber-security remains uneven.²⁹ None of these were sophisticated attacks on hardened systems; they exploited exactly the gaps – unpatched legacy infrastructure, weak access controls, no independent security-audit requirement – that a shared data layer would inherit and amplify unless security is designed in from the start.

The Cost of Inaction

The cost of not having actionable quality data is manifold. NITI Aayog's 2025 assessment estimates that poor-quality data causes fiscal leakage of 4–7 per cent of annual welfare spending – misdirected to the wrong people, the wrong places, or nowhere at all. Database clean-up exercises that simply linked existing siloed records have already demonstrated the scale of the problem: removing 17.1 million ineligible names from the PM-KISAN farmer support scheme saved an estimated 90 billion Indian rupees; eliminating 35 million bogus liquefied-petroleum-gas (LPG) connections saved 210 billion rupees over two years; dropping 16 million fake ration cards is saving roughly 100 billion rupees annually.^{30,31,32}

These figures describe gross fiscal savings, not net public value. Deduplication can legitimately remove duplicate, deceased or otherwise ineligible records, but the same processes can also generate false positives or exclude eligible people when source data, identity-matching or correction mechanisms are weak. Savings should be assessed alongside exclusion rates, reinstated claims, correction volumes, grievance outcomes and the administrative and social cost of errors. The objective is not maximum removal from a register; it is greater accuracy in distinguishing genuinely ineligible records from people the programme is intended to serve.

The social cost falls hardest on people the system is meant to protect. *The Lancet* documented in 2023 how delays to India's census affected access to food support. Several states could not add beneficiaries to the public distribution system, which provides subsidised food grains to families in poverty, because ration-card numbers remained capped using 2011 census figures. This left many people without the food support they were entitled to, including during the pandemic.

There is also a clear geographic pattern. The Human Development Index (HDI) for four of India's states is comparable to that of the Republic of the Congo, while at least two other states' HDI ratings approach that of China. These differences cannot be attributed to data systems alone. However, they illustrate a reinforcing relationship: stronger institutions are generally

better at maintaining reliable data and using them to target resources, while weak data make institutional improvement and accountability harder. States with weaker governance underinvest in data, produce numbers nobody trusts and consequently struggle to target anything well – a cycle that reinforces itself in both directions.^{33,34}

This problem has become more urgent as governments move from dashboards and digital transactions towards AI-enabled analysis, case management and decision support. AI does not remove the need for reliable administrative data; it increases it. Inconsistent definitions, incomplete records, undocumented provenance and weak access controls that a human analyst may sometimes identify or work around can be reproduced automatically and at much greater scale. Used well, the same analytical capability could also help distinguish records that are genuinely ineligible from people the system was meant to protect – a distinction that current deduplication exercises do not always get right. The same reforms required for better delivery are therefore also the foundation for responsible AI adoption.

03

Building a State Data Operating Model

A state data operating model is the practical system through which government turns fragmented administrative data into reliable inputs for services, policy and operational decisions. It combines institutional responsibilities, common standards, reusable data capabilities, assurance and the routines through which evidence is acted on.

Within this model sits a shared data operating layer: the reusable capabilities that allow authorised users and systems to discover what data exist, understand what they mean, assess whether they are fit for purpose, access or exchange them securely, trace how they have been used and correct errors at source. States need these capabilities; they do not necessarily need the same technology platform.

The model must work across the main categories of data that government uses: inputs such as budgets, projects, contracts, assets and personnel; outputs and outcomes such as services delivered, beneficiaries reached and indicators achieved; and service data such as applications, permits, subsidies, inspections and complaints. Connecting these data allows ministers to see delivery risks earlier, for officials to identify operational bottlenecks, and for citizens and researchers to scrutinise public action where appropriate.

Hardware, cloud infrastructure, analytical platforms and dashboards can support this model, but they do not constitute it. The missing capability is the institutional and operational layer that allows data held across different systems and departments to be governed and reused consistently.

The state data operating model has six building blocks. Together, they separate the questions of authority, accountability, interoperability, technical capability, assurance and actual use. These six blocks are:

1. **Leadership and mandate:** Set cross-government priorities, establish common rules and resolve disputes
2. **Ownership and stewardship:** Those accountable for each priority data set and maintaining its quality, meaning and use throughout its lifecycle
3. **Standards and interoperability:** How government establishes common identifiers, metadata, definitions, classifications, schemas and exchange requirements so systems can understand one another
4. **Shared data capabilities:** How authorised users and systems discover, assess, access, exchange, trace and correct data without rebuilding the same integration for every use case
5. **Assurance, rights and safeguards:** How government ensures that data use is lawful, secure, proportionate, auditable and subject to correction or challenge where appropriate
6. **Decision routines and public value:** How trusted data are incorporated into real ministerial, departmental and frontline decisions, and how government measures whether their use improves outcomes

These building blocks should be common even where the institutional and technical architecture differs between states. Privacy, cyber-security, sustainability, open standards and proportionality apply across all six rather than forming separate layers of the model.

FIGURE 1

Six building blocks of the state data operating model



Source: TBI analysis

1. LEADERSHIP AND MANDATE

Cross-government data use requires clear political sponsorship, strategic governance and operational accountability. A chief secretary- or minister-level sponsor should set priorities, provide sustained political backing and resolve disputes that cannot be settled between departments. A chief data officer or equivalent central function should own the cross-government framework, set common standards, provide or coordinate shared capabilities and monitor implementation.

The central function must have explicit decision rights and a clear escalation route. It should be empowered to request action on agreed cross-government priorities while leaving departments accountable for the data and decisions within their mandates. There is no need for each state to create a new authority: the institutional form could be a chief data officer's office, a strengthened Data Strategy Unit or directorate, or another arrangement with sufficient whole-of-government authority. The appropriate institutional form is considered in a later chapter, Building State-Level Data Governance.

These responsibilities should be institutional rather than dependent on individual officeholders. The central function needs sufficient capability, durable resourcing and access to senior decision-makers to resolve persistent problems in standards, interoperability, quality or cross-departmental cooperation.

2. OWNERSHIP AND STEWARDSHIP

Every priority data set should have a named owner, an operational steward and a clearly identified technical custodian. The owner is accountable for the data set's purpose, lawful use, risk, access and expected public value. The steward manages definitions, metadata, lineage, quality, correction and reuse across the data set's lifecycle. The technical custodian operates the systems, infrastructure and technical controls through which the data are stored, processed and exchanged.

These must be real operating responsibilities rather than additional titles assigned alongside an official's existing duties. Owners and stewards need defined decision rights, sufficient time and capability, measurable expectations and a clear escalation route where quality, access or interoperability problems remain unresolved. Stewardship should therefore be treated as an ongoing operational function, with responsibility for maintaining documentation, monitoring quality, coordinating changes and ensuring that recurring problems are corrected at source.

Responsibility also needs to extend to where data are created. District and local officials should be accountable for the quality of data captured through the processes they operate, supported by clear definitions, validation rules, training and feedback. For mobile populations and other cross-jurisdictional cases, states also need mechanisms for portable records, cross-district updates and clear responsibility for reconciling conflicting or outdated information. Previous whole-of-government initiatives show that formally assigning responsibility is not enough where owners and stewards lack the capacity, incentives or authority to perform the role.

Being the official source for data does not mean holding everything in one central database. Different institutions may be responsible for different information about the same person or organisation, depending on their legal duties and how the data are used. One may hold the official address, another the tax information. Where responsibilities overlap, governance arrangements should specify which institution is authoritative for which data objects, how changes are synchronised, how conflicting records are flagged and how unresolved disputes are escalated. Technical architecture should implement these decisions, not determine them by default.

3. STANDARDS AND INTEROPERABILITY

Common standards allow independently operated systems to function as part of the same government ecosystem. States should establish common identifiers where appropriate, metadata requirements, semantic definitions, classifications, reference data, schemas and exchange standards.

Interoperability must be addressed across four dimensions: technical, semantic, organisational and legal. APIs alone do not create interoperability if institutions disagree about meaning, responsibility or authority.

4. SHARED DATA CAPABILITIES: THE OPERATING LAYER

The operating layer should provide a reusable set of capabilities rather than prescribe a single platform. At minimum, governments should be able to:

- Discover priority data through catalogues and machine-readable metadata
- Understand them through common definitions, reference data and semantic standards
- Assess and improve quality through validation rules, quality measures and documented limitations
- Access and exchange them through governed APIs, queries, data contracts or controlled environments
- Trace access, changes and reuse through lineage, versioning and audit logs
- Correct errors through processes that return corrections to the authoritative source

These capabilities may be delivered centrally, federated across departmental systems, provided through sector-specific platforms or combined in a hybrid architecture. What should be common is the capability and the rules governing it, not necessarily the technology used to provide it.

The relevant lesson from India's existing digital public infrastructure is not to create another monolithic platform, but to establish common rules and reusable capabilities that allow many independently operated systems to work together without requiring each new use case to negotiate and build integration from scratch.

5. ASSURANCE, RIGHTS AND SAFEGUARDS

More connected data increase both the value government can create and the consequences of poor quality, insecure or inappropriate use. The operating model should therefore combine departmental responsibility, central assurance and independent scrutiny with practical mechanisms through which people can understand, identify and correct material errors in data affecting them.

At the point of collection, people should be able to understand what information is being collected and for what purpose. Where data materially affect a service, entitlement or decision, there should be practical ways to view relevant information, identify inaccuracies and request correction. Where data are reused or linked for consequential purposes, governments should provide appropriate transparency about that use and meaningful routes to question the resulting decision.

Security should apply across the full lifecycle through proportionate identity and access controls, data minimisation, encryption, logging, monitoring, incident response and recovery. Citizen feedback, corrections and grievances should themselves be treated as evidence of recurring data-quality or service problems and fed back to owners and stewards. The objective is not to require individual approval for every legitimate government use of data, but to ensure that data can be used safely while people retain practical routes to understand and challenge consequential errors.

6. DECISION ROUTINES AND PUBLIC VALUE

Data become valuable only when they change what government does. For each priority use case, states should define the decision to be improved, who is accountable for making it, how frequently it is made, what data are required, the minimum quality threshold, what action the evidence should trigger and how errors or unexpected outcomes are fed back into the system. The measure of success is not whether data have been integrated or a dashboard exists, but whether evidence changes an allocation, resolves a bottleneck, improves a service or enables a better decision.

Extending the Operating Model Across Jurisdictions

The same operating model should extend beyond state-government boundaries where a defined service or decision depends on union, local-government or other states' data. Federal interoperability should mean common rules at the boundary, not common ownership of the underlying data.

This does not imply a single national data lake or a central copy of every state data set. India already uses different sector-specific approaches to cross-jurisdictional coordination. The Goods and Services Tax (GST) common portal provides a shared statutory digital framework across union and state tax administrations; One Nation One Ration Card enables portability for a specific welfare entitlement; e-Shram is explicitly designed to share information with central and state authorities through APIs, and support portability for migrant and construction workers; and ABDM provides common interoperability standards for health information. These systems differ significantly in maturity and governance, but they illustrate why national interoperability is more credible when built around defined sectors, services and legal mandates rather than one general-purpose government-data platform.

For state operating models, the design requirement is therefore two-sided. Departments need common standards and governed exchange within the state, while state systems must also be capable of interacting with relevant national or cross-state services where a legitimate use case requires it. Authoritative data should remain with the institution responsible for maintaining them wherever possible, while common semantics, documented interfaces, access controls, purpose limitation, audit and correction mechanisms allow data to be queried or exchanged safely across jurisdictions.

The policy test should remain the same as for intra-state integration: begin with a defined service or decision and ask what information needs to cross the boundary, under whose authority, with what quality threshold, and who is responsible when the information is wrong.

Sustainability applies across the operating model. States should avoid building critical data capability around a single officeholder, temporary project budget or supplier-controlled technology stack. Institutional responsibilities should survive administrative transfers; recurring stewardship, assurance, security, support and platform costs should have multi-year funding; and technical architecture should use documented interfaces, portable data and credible vendor-exit arrangements. The same principle applies to AI and other externally supplied capabilities: governments do not need to build every component themselves, but they should retain sufficient control over their data, interfaces, decision rights and exit options to avoid critical dependency on a single provider.

SPOTLIGHT

What the Operating Model Already Looks Like in Practice

India is not starting from zero. Existing state initiatives already demonstrate individual capabilities that form part of the operating model, even where they do not yet constitute a complete cross-government model.³⁵

Karnataka – Kutumba: Authoritative matching and proactive eligibility.

Kutumba links individual records through a hashed Aadhaar identifier – never storing the actual Aadhaar number, just a one-way 64-character hash – to proactively identify eligible beneficiaries across more than 30 state portals while weeding out duplicates, paired with the Farmer Registration and United Beneficiary Information System for remote subsidy delivery.

Rajasthan – Pehchan Portal: Event-driven data updates. Pehchan Portal links civil registration to the Jan Aadhaar family database in real time. When a death is registered, the event can trigger verification, updating workflows across ration and pension systems, reducing reliance on manual reconciliation while retaining the necessary checks and audit trail.

Tamil Nadu – State Family Database (SFDB): Governed API-based reuse.

The SFDB integrates the Public Distribution System database with 336 schemes across 62 departments. For scholarship programmes like Pudhumai Penn, it links school and higher-education databases via API for automated eligibility checks with no physical application required.

Odisha – Social Protection Delivery Platform (SPDP): Cross-scheme integration. The SPDP integrates records across 75 social-sector scheme databases, enabling cross-scheme deduplication and eligibility-management workflows.

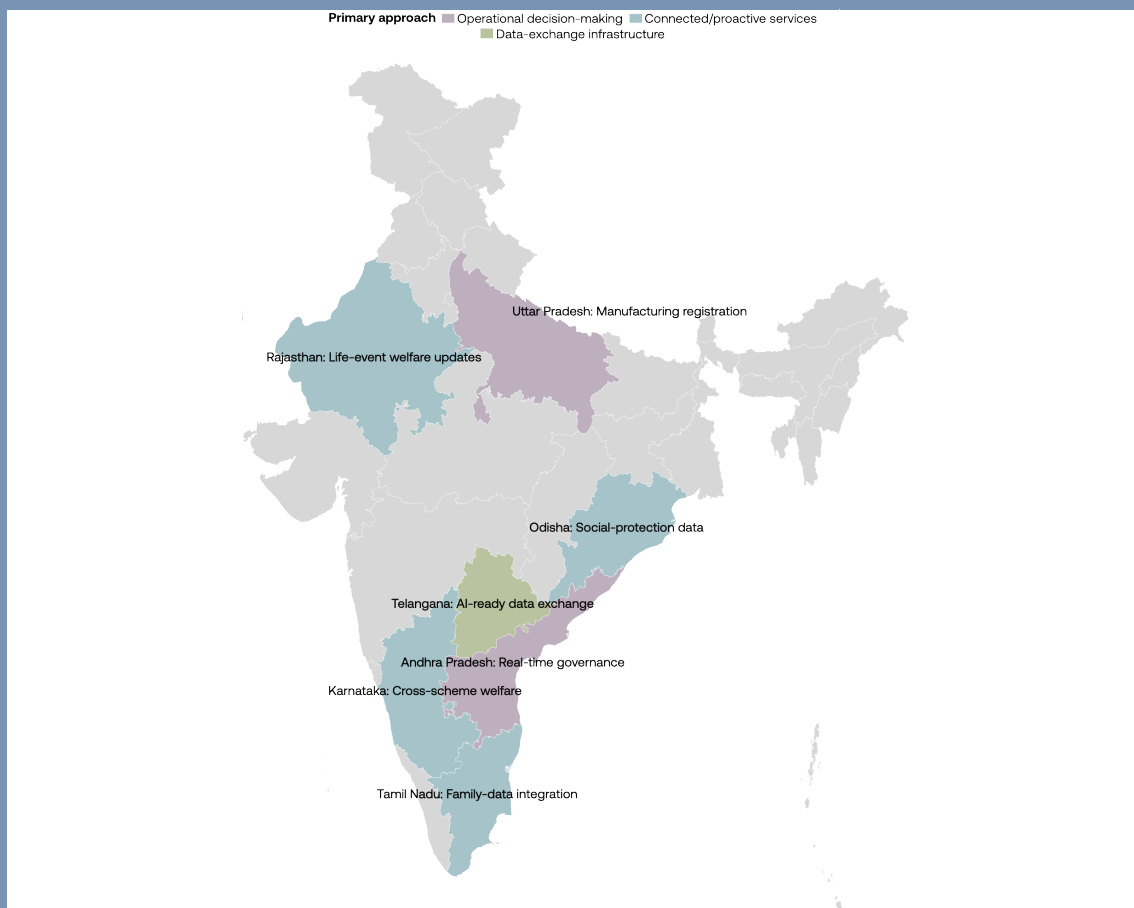
Uttar Pradesh – Factory ID: Improves coverage of foundational data. Factory ID would give each factory a common identifier, allowing information about the same factory to be matched across government records. This would help the state build a more complete picture of its manufacturing sector.

Andhra Pradesh – Real Time Governance Society (RTGS): Embeds data into decision routines. The RTGS, formed in 2017 with a state centre and 13 district centres, pulls live feeds from roughly 47 departments – weather and disaster data, project-site cameras, welfare and grievance systems – on to a single dashboard used for day-to-day decisions rather than periodic reporting. During Cyclone Phethai it pushed evacuation alerts to almost 900,000 people in real time.³⁶

Telangana – Telangana Data Exchange (TGDeX): Controls external reuse. The TGDeX was launched in 2025 with the Japan International Cooperation Agency and the Indian Institute of Science Bengaluru as institutional partners. It packages departmental data as AI-ready data sets – more than 500 of them from more than 20 departments at launch – for reuse by researchers and startups, rather than centralising raw citizen records inside government.^{37,38}

FIGURE 2

Existing state-level data-integration initiatives demonstrate individual capabilities



Source: TBI analysis Note: **Operational decision-making:** Integrated data supporting real-time management or administrative processes. **Connected/proactive services:** Linking data across schemes/departments to improve service delivery. **Data-exchange infrastructure:** Reusable interoperability and data-sharing capabilities enabling wider use.

These are genuine successes, and any state government designing a data operating model should study them directly rather than starting from a blank page. However, these remain largely sector-specific or single-use-case initiatives rather than a routine, system-wide practice, and most states still report that data exchange outside these flagship examples relies on ad hoc manual methods. The question is whether harmonisation can become the default way departments work, rather than an exception that requires a dedicated project team.

These initiatives also illustrate the boundary of a state-only model. They improve integration within individual states, but do not establish interoperability with another state's registry or with national sectoral systems. Cross-state portability therefore depends on additional national standards, shared sectoral infrastructure or other governed exchange arrangements.

Further, it is worth highlighting that a structural limitation runs across these examples: because scheme registry tokenises Aadhaar independently, only the Unique Identification Authority of India (UIDAI) holds the mapping between a given token and the underlying Aadhaar number, so linking beneficiary records generated under different tokenisation schemes back to the same individual is not straightforward. This is by design – it limits the surveillance risk of a single, universal identifier circulating across databases – but it also constrains the cross-database linkage these models otherwise depend on.

Two other comparisons are worth highlighting. National platforms such as NDAP do not replace the need for state-level work; they can only host and merge data that states have already cleaned and validated. This work must happen upstream, within state-government departments and systems. And

digital enrolment is not the same as integration: the Ayushman Bharat Digital Mission has issued more than 960 million health IDs, but a study across six states found a household acceptance rate of only 11.7 per cent in practice, while only about 35 per cent of Indian hospitals have electronic medical-records systems capable of using the linkage meaningfully.^{39,40} Scale of sign-up and depth of actual use are different things, and most current efforts have prioritised the former.

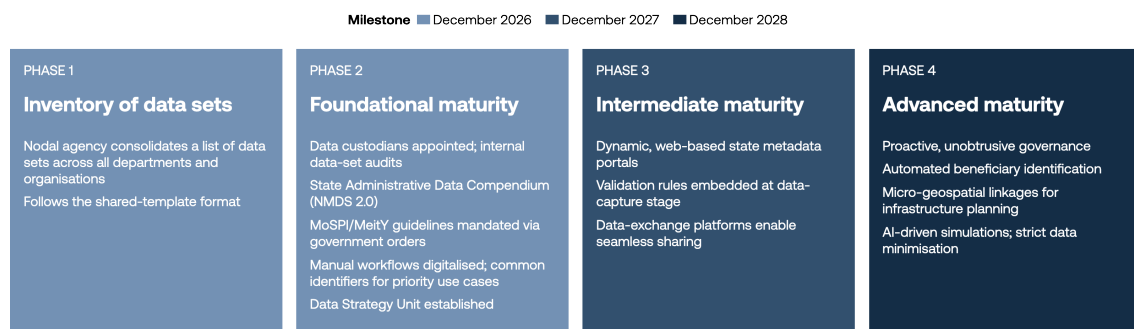
04

Building State-Level Data Governance

India already has a live, time-bound national process addressing exactly this problem. Following the Fifth National Conference of Chief Secretaries in December 2025, MoSPI ran a national consultative workshop in February 2026, state-level internal workshops through March and April 2026, and a National Deliberative Summit in Bhubaneswar on 29–30 April 2026, which produced a four-phase roadmap with milestones running through December 2028.⁴¹

FIGURE 3

The four-phase data-maturity roadmap sets out target milestones for delivery



Source: TBI analysis of MoSPI report

The roadmap is built around a maturity framework with three dimensions – Data Catalogue, Data Consistency and Standards, and Integration Readiness – each assessed at Foundational, Intermediate or Advanced level.

FIGURE 4

The MoSPI data-maturity framework: foundational, intermediate and advanced levels

Foundational	Intermediate	Advanced
No state-wide data-set list exists. Departments keep their own disconnected inventories.	A notified state compendium exists, updated periodically, following the NMDS 2.0.	The compendium is a live digital platform; departments update entries directly; details are exposed via APIs.
Standard definitions exist on paper but are applied inconsistently across departments.	Most departments follow MoSPI/MeitY harmonisation guidelines as an ongoing practice, not a one-off exercise.	Common definitions and validation rules are built into system design itself, applied automatically at the point of data entry.
Systems are fully separate. Data moves manually – files, emails and phone calls.	Major systems share standard classifications; data move through defined technical interfaces, though matching still needs manual checking.	Systems share common identifiers by design; data exchange is fully automated through APIs with no manual reconciliation.

Source: TBI analysis of MoSPI report

The roadmap's milestones are concrete: by December 2026, states are expected to have established a governance body (a Data Strategy Unit or equivalent), created a data-set inventory and started releasing machine-readable data with priority APIs; by December 2027, a dynamic web-based catalogue and broader harmonisation across all data sets; by December 2028, fully automated, source-generated data systems, with each state government operating its own data-exchange platform or using the national platform.

MoSPI's roadmap is a strong national foundation for state action. States should use this framework as a common national baseline but not as a complete operating model. It should be supplemented with dimensions covering governance and accountability, data quality and citizen redress, privacy and security, workforce capability and demonstrated public value. Further, it should be implemented as a governance and delivery reform, not treated primarily as a technology programme.

The milestones set by MoSPI establish a baseline that a well-equipped state government should be able to meet ahead of schedule. But meeting it will require an institution with more authority and more durable funding than the Data Strategy Units the roadmap currently proposes as the default mechanism. A unit embedded inside a single department, however well-

intentioned, cannot compel other departments to share data – and the roadmap itself acknowledges this, noting that the central barriers are institutional and political, not technical. Officials resist sharing data because information is power, and interoperability exposes comparative performance and reduces operational autonomy.

Officials who currently control data access lose something tangible when data become transparent and open to query by others – this is not an oversight to be fixed with better change-management slides; it is the central political obstacle. MoSPI's consultations confirm that the states that made real progress did so because a senior political figure treated data reform as a sustained delivery priority across more than one budget cycle, not because the technical design was correct.

Integration without fixing underlying enrolment gaps can exclude people faster than the manual system it replaces. These are the predictable result of connecting databases before correction and appeal mechanisms are in place. Every integration phase needs an explicit exclusion check: who has become newly invisible and why?

In many states, the availability of skilled human capital will be as significant a constraint as technology. India's data crisis is as much a shortage of trained data stewards, district statisticians and field investigators as it is a shortage of infrastructure. Building platform capability without building a skilled workforce in parallel will only produce a more sophisticated way to collect unreliable data.

The legal environment is also evolving. The DPDP Act amends the personal information exemption in Section 8(1)(j) of the Right to Information (RTI) Act, while the RTI Act's broader public-interest override under Section 8(2) remains. How the resulting balance between privacy, transparency and administrative data reuse will operate in practice will continue to be important for state data governance.

National Lessons: Common Rules, Not Data Centralisation

The lesson from India's digital public infrastructure is not that all data should be centralised. It is that common rules, identifiers and exchange mechanisms can allow many institutions and providers to operate as one system.

THE AADHAAR LESSON: INSTITUTIONAL CAPABILITY WITHOUT INSTITUTIONAL COPYING

Aadhaar is instructive not because UIDAI provides an institutional model that states should replicate, but because it demonstrates both the value and the risks of concentrating mandate, capability and accountability around a major data-driven public function.⁴² UIDAI itself followed an unusual institutional path: it began as an attached office of the Planning Commission before becoming a statutory authority under the Aadhaar Act in 2016. Its experience should therefore be read as a source of functional lessons rather than a blueprint for state data governance.

This lesson cuts both ways. Aadhaar's own record includes the documented exclusion failures cited earlier in this paper. These cases show that any state-level data-governance institution – whether a data office, a strengthened directorate or a statutory authority – must build independent oversight, effective correction and accessible appeal mechanisms into its design from the outset. They are not arguments against stronger institutional capacity; they demonstrate why safeguards cannot be added only after integration has scaled. As data become more connected, both the benefits of strong governance and the costs of weak governance increase.

The institutional form can vary. In some states, the right model may be a chief data officer and cross-government data office under the chief secretary. In others, it may be a strengthened Directorate of Economics and Statistics, Planning or Finance function with a formal whole-of-government mandate. A State Data Authority may be justified where the scale, sensitivity

and complexity of cross-government data use require it. What matters is not the label, but authority, durable funding, specialist capability, cross-government reach and clear accountability.

Continuity should be built into the institution rather than dependent on one officeholder. States should use the strongest mechanism available in their administrative context: a formally notified mandate, multi-year funding, a permanent secretariat and specialist staff, documented decision rights and standards, succession arrangements and, where feasible within cadre rules, a minimum expected tenure for the chief data officer-equivalent leadership role. The objective is to ensure that routine transfers do not reset the programme or remove the institutional memory needed to sustain it.

Internally, its functions mirror that of a multi-party consortium, but under one accountable roof: officers who work directly with district administrations to build data dictionaries and map how each department actually uses its data; a technical team that owns the integration platform; and a standards function with real enforcement power, not persuasion alone. Karnataka's three-tier model – a steering committee for cross-government policy disputes, an executive committee for ownership and structural decisions, and an operational working group that reviews and authorises individual data requests – is a workable template other states could adapt rather than design from scratch.

States should identify the authoritative source for core attributes needed to describe people, businesses, places, assets, schemes and government organisations. Authority may sit with different institutions for different attributes and purposes. Other systems should verify or retrieve these data from the authoritative source rather than create competing copies. Where union, state or local responsibilities overlap, precedence, synchronisation and dispute-resolution rules should be agreed explicitly.

One essential function UIDAI lacks but which this authority must have is an independent data-quality audit function, which publishes its findings and provides a structured channel for researchers, journalists and civil society to access appropriate data and flag what government systems miss. A 2023 review by Carnegie Endowment for International Peace found that India's

National Statistical Commission, despite being formally tasked with exactly this audit role, has lacked the statutory backing to perform it for more than 18 years.⁴³ A new authority will repeat that failure if it is not built with independent oversight from the outset.

The operating model should distinguish three levels of assurance. First-line responsibility sits with departmental data owners and stewards. Second-line assurance sits with the central data function, which sets standards and monitors compliance. Independent third-line review should sit outside the delivery function with the ability to publish findings on high-impact data sets and uses. Citizen participation is not a fourth line of assurance but a cross-cutting requirement across all three. Departments need channels through which people can identify incorrect records and service problems; the central data function should monitor patterns in corrections, grievances and exclusions as evidence of systemic data-quality problems; and independent assurance should consider citizen experience and external evidence alongside internal technical assessments.

THE NDSAP LESSON: INSTITUTIONAL MANDATE WITHOUT INSTITUTIONAL FOLLOW-THROUGH

The National Data Sharing and Accessibility Policy (NDSAP) demonstrates what happens when a whole-of-government data mandate is issued without any of the capability, capacity or accountability structures needed to make agencies comply. NDSAP followed an unusual institutional path of its own: it was notified by the Department of Science and Technology in 2012, with the National Informatics Centre (NIC), an executing body sitting inside the Ministry of Electronics and Information Technology (MeitY) building and running the actual platform. NDSAP had not established clear protocols for information gathering, processing and sharing, and unlike Aadhaar, the policy never crystallised into a statutory authority with independent powers of its own. It has remained, for more than a decade, an executive circular that line ministries and states can choose to observe or ignore.⁴⁴

This lesson, too, cuts both ways. NDSAP's own record includes the failures documented earlier: data sets uploaded to the platform were found to be outdated, duplicated, incomplete, lacking in semantic interoperability and

inadequately referenced, with an absence of good quality or any metadata associated with them. A formal review concluded that implementation has “lagged behind stated objectives due to multiple reasons including ambiguity around data-set classification, low data quality, lack of capacity and skills to create, manage and use high-quality data sets, and no effective government-to-government data-sharing mechanism”. These are not incidental technical glitches; they show that any data-sharing mandate (whether a policy circular, a directorate or a statutory authority) must build enforcement power, quality assurance and inter-agency mechanisms into its design from the outset, not treat them as a later software feature.^{45,46}

What NDSAP lacked was authority. Suppliers of data across various ministries and departments struggled to grasp exactly what the Open Government Data (OGD) Platform entailed and the economic value it could generate; they also had limited resources and capacity for implementation. No entity held the standards-setting or enforcement role that is essential for a data office; compliance was voluntary, self-reported and unaudited. This is similar to the pattern described in most country-wide literature on the failures of open-government-data failure: the main root causes include politico-administrative, social, technological, legal and organisational dimensions, including the tendency to copy the OGD initiative without institutionalising it. NDSAP was adopted because open data was a global governance trend circa 2012 not because the underlying institutional capability to sustain it had been built first.⁴⁷

Continuity was not designed into the process either. There was no formally notified enforcement mandate, no durable multi-year funding line dedicated to data quality and no permanent specialist secretariat – only the general administrative machinery of NIC and rotating departmental “NDSAP cells”. The 2021 review found the fix had to be manufactured after the fact: it proposed the adoption of specific and uniform data standards and metadata schema, mandatory annual data audits through an approved list of companies, and mandatory two-day training programmes every six months for officials from all NDSAP cells to build both data-management and data-analysis skills – precisely the standing capability that should have existed from year one. Fragmentation compounded this: adoption of data-

sharing at the state level was slow, with only four out of 29 Indian states contributing data to the national portal, meaning the “whole-of-government” claim never extended much beyond the union government’s own departments.⁴⁸

The one function NDSAP never had, which any successor authority must, is an independent, published data-quality audit combined with a real citizen/researcher feedback channel. NDSAP’s implementation guidelines relied on internal departmental self-certification rather than external review. This is the same structural gap documented elsewhere in India’s data ecosystem: an audit function that exists on paper but is never given the statutory teeth to operate independently. A new state or sectoral data authority will repeat NDSAP’s failure if oversight is bolted on only after the first credibility crisis, rather than built in from the start.

Tellingly, the institutional response to NDSAP’s failure was not to reform NDSAP itself. MeitY’s 2022 draft India Data Accessibility & Use Policy proposed setting up an entirely new India Data Office to streamline and consolidate data access and sharing of public-data repositories across government, while NITI Aayog separately built the NDAP, a data lake developed by a private contractor, as a parallel single-platform collection of government data sets with better visualisation tools, APIs and alerts. Both responses imply that the original body lacked the authority to fix itself – but rather than addressing NDSAP’s foundational weaknesses, they built around them. Applying the three-line assurance model, NDSAP effectively had no first line (departmental data owners bore no real accountability for the data sets they published), no functioning second line (no central body had genuine standard-setting or enforcement power over line ministries until the 2022 draft proposal), and no third line at all (no independent, publishing audit function ever existed). A state-level data authority that wants to avoid NDSAP’s fate needs all three lines functioning from day one, not layered on after the platform has already lost public and departmental trust.⁴⁹

GST'S DIGITAL ARCHITECTURE: A CLOSER PRECEDENT FOR CROSS-JURISDICTIONAL DATA EXCHANGE

An even closer precedent for government-data interoperability is India's Goods and Services Tax digital architecture. GST uses a common electronic portal and standardised registration and reporting processes across the country, while union and state tax administrations continue to perform their respective statutory functions through connected systems. The lesson is not that welfare, health or education should replicate the Goods and Services Tax Network (GSTN). It is that cross-jurisdictional data exchange becomes routine when common structures, legal mandates, operating rules and accountability are built into the system rather than negotiated as one-off arrangements.

Unlike NDSAP, GST's interoperability rests on an explicit constitutional mandate rather than an executive circular. The 101st Constitutional Amendment Act 2016 inserted Article 279A, requiring the president to constitute the GST Council within 60 days of the amendment coming into force. The Council is a 33-member constitutional body chaired by the Union Finance Minister that brings the central government and every state into a single decision-making forum, making recommendations on tax rates, exemptions, rules and the phased inclusion of items such as petroleum products still outside GST's scope. Because the Council's authority derives from the constitution, its recommendations carry a force and legitimacy that a policy circular like NDSAP never had – disputes over classification or rate-setting are resolved inside a standing federal body with a permanent secretariat, not negotiated bilaterally between individual departments each time a disagreement arises.

The technical-execution layer is deliberately separated from this policy layer. GSTN is the IT backbone that implements the Council's decisions – handling registration, return filing, e-invoicing, e-way bills and payment reconciliation through a single national portal rather than requiring taxpayers or state administrations to interact with dozens of separate state-level systems. This split is structurally significant: the GST Council decides on the rules, and GSTN is accountable for how those rules are executed technically, with each function answering to a different, clearly defined form of accountability.

GSTN's ownership itself evolved to reinforce this: it began in 2013 as a hybrid entity with 51 per cent private ownership by financial institutions and 49 per cent government ownership, designed to give it agility in hiring and technology procurement, before the GST Council later approved its conversion into a fully government-owned entity split evenly between the central government and the states collectively. That shift is a governance lesson in itself – a core sovereign function transitioned to joint centre-state ownership once the technology was established, rather than remaining dependent on private control.

GSTN's interoperability with the wider financial and administrative ecosystem is also rule-bound rather than improvised. The system integrates external entities such as banks, the Reserve Bank of India (RBI) and GST Suvidha Providers through an open architecture and well-defined APIs, which is what allows real-time tax-payment updates, bank account validation and third-party filing software to reconcile against the same authoritative data set without separate one-off agreements for each integrating party. GSTN does not hold tax money itself; funds flow to the RBI and onward to the respective government accounts, while GSTN is paid a user charge by government for maintaining the infrastructure. This is a clear separation of custody from execution, which limits the operational and security risk concentrated in any single node.

Accountability for data security and confidentiality is similarly built into the governance structure rather than bolted on. The central government retains strategic control over the composition of GSTN's board, special-resolution mechanisms and shareholder agreements, precisely because the system is authorised to collect, retain, store, process and transfer the trade-transaction data of the entire country and is therefore required to meet the cyber-security obligations set out in the IT Act, 2000. This gives GST's architecture something NDSAP conspicuously lacked: a designated, empowered body with legal responsibility for the integrity and security of the shared data, answerable to a constitutional decision-making forum, rather than diffuse and voluntary compliance by individual departments.

The result is that cross-jurisdictional data exchange under GST is genuinely routine rather than negotiated case-by-case. A taxpayer registering or filing in any state interacts with the same national portal, under the same standardised process, with Input Tax Credit and Integrated GST settlement handled automatically between jurisdictions rather than through manual state-to-state reconciliation. This is precisely the “whole-of-government” outcome that NDSAP aspired to but never achieved for open data – not because GST’s underlying technology is unusually sophisticated, but because it paired a constitutionally mandated federal decision body, a legally accountable execution arm, defined data-custody rules and an explicit integration architecture with external systems; the four elements NDSAP left informal, voluntary or entirely absent.

International Lessons: Functions, Not Institutional Templates

Global experience does not point to a single institutional model. It points to a set of functions that need to exist, regardless of whether they sit in a central data office, a strengthened statistical or planning function, or a statutory authority. These examples should therefore be treated as functional references rather than direct institutional templates. Their administrative scale, constitutional arrangements, institutional capacity and trust environments differ significantly from those of Indian states. The relevant question is which governance functions transfer, not whether the institutional form or technology can be copied.

Estonia: Distributed ownership and governed data exchange. Estonia does not rely on a single central government database. Authoritative registers remain with the institutions responsible for maintaining them and exchange data through X-Road,⁵⁰ a secure data-exchange layer. The once-only principle means that citizens and businesses should not repeatedly be asked for information that government already holds through an authoritative source. The relevant lesson is not to copy X-Road as a product.

It is to identify authoritative sources, standardise how data and services are described, establish lawful access and ensure that exchanges are traceable, rather than creating an uncontrolled central copy of government data.

United Kingdom: Ownership attached to critical data assets. The UK's model requires critical data assets to have named owners and stewards, accurate metadata, central registration and active quality management.⁵¹ The practical lesson is that a catalogue creates limited value unless an accountable person is required to maintain quality, approve reuse, resolve problems and explain the risks associated with the data.

Australia: A dedicated regulator to prescribe data standards. Australia does not leave technical data standards to be negotiated separately by each sector or agency. The Consumer Data Right, a legislative framework for data portability rolled out sector by sector starting with banking and energy, is supported by a dedicated Data Standards Body (DSB),⁵² which sets the technical rules (API specifications, data schemas, security profiles) and consumer-experience standards that every participating data holder and accredited recipient must adhere to. Authority is deliberately separated by function rather than concentrated in one body: the DSB, sitting within the treasury, develops the standards; a ministerially appointed Data Standards Chair formally makes them; the Australian Competition and Consumer Commission accredits and enforces compliance among participants; and the Office of the Australian Information Commissioner regulates privacy and investigates data breaches independently of all three. The relevant lesson is that technical standardisation, enforcement and independent privacy oversight are distinct functions that work because they sit with different, clearly accountable bodies rather than being combined in one and left to informal or voluntary coordination between departments.

Organisation for Economic Co-operation and Development (OECD): Data governance judged by public value. OECD guidance treats data as a strategic asset used across policymaking, service delivery, organisational management and evaluation.⁵³ This shifts the test from whether a platform or catalogue exists to whether data improve decisions, services, accountability and trust.

Across these examples, six design tests emerge for Indian states:

1. An authoritative source is identified for each core entity and priority data set.
2. A named owner is accountable for quality, lawful use, access and correction.
3. Systems can exchange data without creating unnecessary or uncontrolled copies.
4. Access, queries, changes and reuse are logged and auditable.
5. Citizens and independent reviewers can identify and challenge errors.
6. Each integration is linked to a defined decision, service or public-value outcome.

These are the functions Indian states should adapt. The institutional label and technology stack should be selected according to each state's administrative structure, legal environment, capability and delivery priorities.

Preparing State Data for AI

AI raises the stakes. Governments are under pressure to show progress on AI adoption. Officials are being asked to identify use cases, procure tools and demonstrate efficiency gains. Technology providers are offering systems that promise better analytics, automated casework, fraud detection, predictive targeting and more seamless conversational access to government information and services.

Some of these tools will be useful. But AI will not compensate for weak data foundations – it will expose them faster and at greater scale. If records are incomplete, AI systems trained on them inherit the gaps. If categories carry embedded bias, automated systems can reproduce and amplify that bias in eligibility decisions at machine speed, in ways that are difficult to detect without a deliberate audit.

If data ownership is unclear, accountability for AI-enabled decisions will be unclear. If access rights are poorly managed, AI can widen access to sensitive information. If metadata and lineage are missing, officials may not know whether an output is based on reliable evidence.

AI systems require precision that human analysts can work around: explicit metadata, stable data models and machine-readable provenance. A data set that is perfectly usable for a human reading a report can be entirely unusable for an AI system trying to answer the same question because an AI system may infer missing context incorrectly or rely on undocumented assumptions, making explicit metadata, definitions and provenance essential. MoSPI's technical framework defines the requirements for being "AI-ready" using five layers, providing a more precise specification than most state AI strategies currently work with.

FIGURE 5

The five layers of MoSPI's AI-readiness framework

Layer	What It does
Discovery	A machine-readable catalogue for AI systems to find available data sets and indicators without a human pointing them to it.
Semantics	A shared vocabulary (NMDS 2.0 plus a national statistics dictionary) that gives every term a single, machine-readable definition.
Delivery	Data structured so a machine can tell dimensions (what a record is about) apart from measures (the values) and query both through an API.
Integration	Context files that map field names to their statistical meaning, so two systems can connect without a human translating between them.
Governance	Version control is anchored to a national standards body, so as definitions change, machines always pull the current version, not a stale one.

Source: TBI analysis of MoSPI report

Semantic interoperability also has a multilingual and multi-script dimension that the framework does not make explicit. Common definitions need to work across the languages and scripts in which source records are created, labels and free text are entered, and services are delivered. This is not simply a translation problem: transliteration, spelling variation, local terminology and inconsistent encoding can make records that are equivalent to a human difficult for automated systems to match reliably. States should therefore include multilingual terminology management, language metadata, encoding standards and validation rules within their semantic-governance and data-quality work.

The most concrete expression of this is MoSPI's Model Context Protocol (MCP) Server – a technical component that lets an AI system ask, in effect, “what does this indicator mean, where does it come from and how should I interpret it?” and receive a reliable, machine-readable answer, without a human in the loop translating between systems.⁵⁴ MCP can improve the discoverability and controlled use of existing APIs by AI tools. However, it is an interface pattern, not a data-governance model. It does not establish lawful basis, authoritative sources, semantic consistency, data quality, access rights, evaluation or accountability. States should therefore not treat MCP implementation as evidence that their data are AI-ready.

AI readiness should be assessed against a defined use case. A high-impact application is one where an AI-supported output can materially affect an individual's eligibility for benefits, access to essential services, financial obligations, enforcement outcomes, public safety, rights or another consequential government decision. For these applications, states need a clear purpose and legal basis, an assessment of whether the data are representative and suitable, testing for error and unfair impact, meaningful human review, explanation and appeal routes, security testing, ongoing monitoring, incident response and named accountability for the resulting decision. Departments remain responsible for their decisions made through these processes even where an AI system informs, recommends or coordinates the action.

For example, an AI system used to identify potentially fraudulent welfare claims should not suspend a benefit solely on the basis of a model score. The department should test false positives, document the data and variables used, require an authorised official to review adverse cases, retain an auditable reason for the action and provide the affected person with a route to contest the decision. A low-risk internal tool used to summarise non-sensitive documents would not require the same level of assurance; controls should be proportionate to the consequences of error.

This is the practical difference between an AI pilot that works once on a clean demo data set and a reusable government capability: the first depends on a team manually preparing data for a single use case; the second depends on infrastructure that any new use case can plug into without starting over. States procuring AI tools without this underlying layer should expect every new use case to require its own bespoke data-cleaning project – which is the pattern most government AI pilots have followed so far and is precisely what makes them pilots rather than capabilities.⁵⁵

06

Recommendations

India's states vary widely in capability, infrastructure, resources and institutional maturity. A single implementation model will not work everywhere, even if the core principles are common.

States with stronger digital infrastructure can move quickly towards integration, advanced analytics, AI-enabled decision support and research access. Their challenge is often coordination, standards, trust and political ownership.

States with weaker foundations may need to begin with priority data sets, basic stewardship, structured reporting, cloud readiness and a small number of high-value use cases. Their challenge is often capability, financing and execution discipline.

Sector-led states with strong systems in health, agriculture, education or welfare but weak cross-government coordination should focus on interoperability, shared identifiers, common catalogues and data-sharing rules.

States with low trust in official data should prioritise independent quality assurance, transparency, civil-society engagement and correction mechanisms before moving into high-stakes AI or automated eligibility use cases.

The principle is the same across contexts. Start from the decisions government needs to make, identify the data required to make them, fix the institutional and technical foundations, and build use cases that prove value.

The approach should be modular. It should allow states to begin where there is political demand and delivery value, while moving gradually towards a common architecture. Resource-constrained states should be able to use open standards, shared components and pooled procurement rather than bespoke vendor-led platforms. That said, seven broad recommendations hold for most Indian states.

Recommendation 1: *Establish an empowered cross-government data function with a clear delivery mandate.*

Fragmented data cannot be solved by departmental goodwill alone. A central function is needed to set rules, convene institutions, support implementation and maintain accountability.

States should establish or strengthen a cross-government data function with the mandate to govern priority public-sector data, coordinate standards and support delivery use cases. Depending on the state context, this may take the form of a chief data officer's office, a strengthened Data Strategy Unit or directorate, or a statutory State Data Authority. This would build on MoSPI's proposed Data Strategy Unit by giving the central function an explicit whole-of-government mandate, a clear escalation route and durable funding.

States should begin by assigning political sponsorship, appointing a chief data officer or equivalent lead, identifying departmental data stewards and approving a limited first-year mandate focused on priority data sets and use cases. The initial institutional design should also include continuity measures: permanent secretariat capacity, documented decision rights and standards, succession and transition arrangements and, where feasible, a minimum expected tenure for the lead role. This reduces the risk that routine administrative transfers reset the programme or leave it dependent on one individual.

The required capability is not only technical. The authority needs policy, legal, data-engineering, privacy, service-design and delivery expertise. It must be close enough to the centre of government to influence priorities, but practical enough to work with departments and districts.

The risk it mitigates is institutional fragmentation: the tendency for every department to build its own system, define its own rules and resist shared accountability.

Recommendation 2: *Build governed data capabilities around priority decisions.*

States should develop the capabilities to make high-value data usable for defined decisions, including discovery and metadata, authoritative source management, semantic interoperability, quality management, secure access and exchange, lineage and audit, and correction at source. These capabilities may be implemented through shared, federated, sector-specific or hybrid technical arrangements depending on the use case and existing environment. This delivers on the Integration Readiness dimension of MoSPI's maturity framework without assuming every state should build the same platform. Governments do not need more data for its own sake. They need reliable, governed data for targeting, delivery, planning, monitoring, service improvement, accountability and safe AI adoption.

Where a priority use case crosses jurisdictional boundaries, the operating layer should also support governed exchange with relevant national and other state systems through agreed sectoral standards, interfaces and access rules. Cross-state interoperability should be designed around defined services and decisions rather than through wholesale replication or pooling of another jurisdiction's data. Sectoral governance arrangements should also define authoritative attributes, conflict-resolution rules and an escalation route where union and state records or responsibilities diverge.

Durable funding should be based on a multi-year operating model rather than an implementation-project budget. States should estimate the recurring cost of the central data function, departmental stewardship, platform and integration services, security operations, independent assurance, support, capability building, upgrades and vendor-exit arrangements. The costing should differentiate between permanent institutional capability and specific use-case expenditure, identifying which responsibilities are centralised and which remain with departments.

States should begin with a small number of priority domains where the fiscal, social or political value is clear: welfare targeting, health-system performance, education records, infrastructure projects, agriculture support, disaster risk or grievance redress.

For each priority use case, states should define the decision to be improved, the official responsible for that decision, how frequently it is made, the data required, the minimum quality threshold, the action that the evidence should trigger and the feedback or correction route. This prevents data programmes becoming inventories without users or dashboards without decisions.

Data use should be embedded into existing ministerial, departmental and district delivery routines. The measure of success is not whether a dashboard exists, but whether evidence changes an allocation, resolves a bottleneck, improves a service or triggers corrective action.

The necessary institutional capability includes data architecture, metadata management, data-quality management, access control, privacy and data-protection procedures, management of lawful basis, purpose and citizen rights – including consent where consent is the appropriate basis – and operational routines for using evidence in decisions.

These routines need corresponding incentives and accountability. Departmental performance frameworks and senior officials' delivery objectives should include responsibility for maintaining priority data, resolving identified quality issues, responding to agreed cross-government data requirements and demonstrating how evidence has informed priority decisions. Ministerial or chief secretary reviews should distinguish compliance activity – for example publishing a data set or producing a dashboard – from actual outcomes, such as whether a bottleneck was resolved, a service improved or an identified data-quality problem corrected.

This helps mitigate the risk of technology-led reform, of procuring a platform before clarifying the decisions, users, rules and responsibilities that make the platform valuable.

Recommendation 3: *Make data ownership and stewardship operational.*

Governments should require every high-value data set to have a named owner and a steward, a defined public or operational purpose, metadata, sensitivity classification, quality rules and measures, access and reuse

conditions, retention and correction procedures, and lineage. Ownership establishes accountability; stewardship turns that accountability into continuous operational practice. This extends MoSPI's Data Catalogue to make clear who is accountable for each data set.

Data without ownership cannot be trusted. When nobody is responsible for a data set, quality problems persist, access decisions become ad hoc and errors are difficult to correct.

States should begin by identifying the top 20 to 50 data sets most important for service delivery, fiscal management and political priorities. These should be catalogued, assessed and assigned to named owners.

Stewardship should be treated as a lifecycle function inside departments, supported by common central standards, tooling and communities of practice. Stewards should monitor quality and usage, coordinate changes in definitions and schemas, resolve interoperability and correction issues, maintain documentation and feed recurring problems from frontline users, citizens and external reviewers back into the data set's management. It should be a substantive operating responsibility, not an optional digital role.

The risk it mitigates is unmanaged reuse – data being shared, linked or used without clarity on meaning, quality, permission or accountability.

Recommendation 4: *Create independent data-quality assurance and safe external access.*

A data system that cannot be questioned, corrected or audited will not be trusted. The government should establish an independent authority to assess priority data sets, identify inclusion and exclusion risks, and support appropriate access for researchers, civil society and innovators. This is particularly important where data are used for welfare eligibility, targeting, enforcement, health planning, education, disaster response or AI-enabled decision support. In these areas, errors do not remain technical. They affect rights, access and lives.

MoSPI's maturity framework has no dimension for this: its three axes – Data Catalogue, Data Consistency and Standards, and Integration Readiness – assess whether data exist and connect, not whether they can be trusted or independently scrutinised.

Independent scrutiny is essential to complement internal reporting and demonstrate that priority data sets can be trusted. Data that shape public decisions should be subject to scrutiny, especially where they affect eligibility, rights, resource allocation or public claims about performance. Citizens affected by high-impact administrative data should have practical routes to identify inaccurate records, understand where an error has materially affected a service or entitlement and seek correction without having to navigate the internal ownership structure of government. Aggregated patterns in complaints and corrections should be treated as data-quality signals and fed back to data owners and stewards.

Civil society organisations, researchers, journalists and universities often identify inconsistencies that government systems miss. A mature data architecture should create safe channels for open data, restricted research access and controlled sandboxes. Not all data should be open. But where data can be opened safely, this strengthens accountability. Where data cannot be opened, structured research access can still improve scrutiny and learning.

This also supports the wider data economy. Public-data sets, geospatial information, research data and standardised APIs can help firms, researchers and innovators build services and analysis that government alone cannot produce. The economic value of data depends on trust, usability and access rules. Data that are locked away, undocumented or unreliable have limited public value.

States should begin with quality reviews of data sets used for high-stakes schemes. They should publish quality summaries where possible and create controlled research-access routes where full openness is not appropriate.

However, not all reuse requires the same access model. States should distinguish between open data, cross-government shared data, restricted research data, highly sensitive operational data and data that should not be reused. Each tier should have a defined legal basis, purpose, approval route, security requirement and audit mechanism.

The required capability includes statistical review, audit, privacy protection, secure research environments, implementation of privacy-enhancing technologies and clear legal gateways for data access. The risk it mitigates is credibility failure – the erosion of public trust when official data are inaccurate, unverifiable or visibly disconnected from lived experience.

Recommendation 5: *Build legal, privacy and procurement safeguards for reuse.*

Governments should establish clear legal gateways, privacy controls, retention rules, audit logs, vendor obligations and procurement standards for data reuse across departments and with external partners. This is the safeguard MoSPI's roadmap assumes will exist but does not build itself, at a moment when the legal environment – the DPDP Act and Data Governance Framework Policy still in draft since 2022 – is still being settled.

Data-sharing should be supported by clear legal gateways, standard approval processes and architectures that remain under government control. States need lawful, auditable and reusable routes for data access that protect citizens while enabling delivery.

Procurement should require open standards, documented APIs and schemas, exportable metadata and lineage, data portability, security testing, knowledge transfer, exit assistance and state control over core data models, interfaces and architecture. Contracts should measure working interoperability, improvements in data quality and delivery outcomes, not only whether a platform has been deployed.

Procurement should prioritise iterative delivery rather than locking states into a fully specified multi-year platform before testing user needs, data quality and operating processes. Where procurement and budgeting rules allow,

states should use staged discovery and implementation, minimum viable capabilities with clear criteria for scale-up, and modular contracts with documented interfaces between components. Acceptance criteria should test working data exchange, quality and user outcomes rather than simply completion of predetermined features. Contracts should also allow for the replacement of components or suppliers without requiring the state to rebuild the underlying data architecture.

The risk this mitigates is unlawful or unsafe reuse, cyber-exposure and vendor lock-in systems that may technically connect data while leaving the state unable to explain, audit, secure, move or govern them.

Recommendation 6: *Use a State Data Balance Sheet to link data to major reforms.*

States should publish or internally maintain a Data Balance Sheet that identifies priority data assets, liabilities, debt, risks and dividends. No equivalent instrument exists in MoSPI's framework, which gives states a maturity score but no ongoing way to link that score to budget and delivery decisions.

A State Data Balance Sheet would provide a practical way to make data governance visible to leaders. It would show which data sets exist, where a state government would face data liabilities and risks, and where better use of data creates opportunities. It would allow ministers to govern data as they govern money: by identifying assets, liabilities, risks and return on investment. It should include five categories:

1. Data assets: Key registries and databases, service logs, geospatial information, project records, administrative data sets and open data resources
2. Data liabilities: Missing baselines, outdated records, duplicate beneficiaries, unverifiable fields, undocumented data sets and low-quality departmental data

3. Data debt: Legacy systems, manual reporting, spreadsheet-based processes, non-standard identifiers, fragmented platforms and repeated reconciliation work
4. Data risks: Privacy exposure, exclusion risks, cyber vulnerabilities, vendor lock-in, weak access control and poor auditability
5. Data dividend: Savings, better targeting, faster services, improved investment confidence, stronger research access and AI readiness, reported alongside material exclusion, correction and error costs rather than as gross savings alone

Governments need a practical management instrument, not just a data strategy. A balance shows clearly which data sets can be trusted, which create fiscal or social risk, which legacy systems generate repeated reconciliation work and where better data have produced value. This is not a financial-valuation exercise and does not imply the sale or monetisation of personal or administrative data. It is a management instrument for prioritising investment, risk reduction and public value.

States should begin with the top 20 to 50 data sets linked to budget, delivery and political priorities. The balance sheet should be updated continuously and reviewed through budget, digital-investment and delivery processes. It should become a management instrument, not a reporting annex.

For practical purposes, the Data Balance Sheet should be reviewed alongside budget and delivery priorities to help leaders decide which data sets to fix first, which reporting burdens to remove, which legacy systems to retire and which AI use cases are safe to scale.

The risk it mitigates is invisible data debt: the tendency for governments to fund new platforms while the underlying data assets remain outdated, duplicated, undocumented or unreliable.

Recommendation 7: *Sequence reform around priority decisions and delivery use cases.*

The MoSPI roadmap should remain the high-level framework because it is nationally agreed and uses language that state governments already recognise. Our recommendations build on that framework by setting out the practical actions, institutional arrangements and safeguards needed to move from data inventory towards reliable, governed and increasingly automated use of data.

However, successful implementation of the recommendations also requires them to be sequenced. For instance, governments should resist the temptation to start with dashboards or AI tools. Reform should proceed from mandate, priorities and stewardship to inventory and quality improvement, to governed integration and service use, and only then to high-level indicators and advanced AI applications. This matters because indicators and AI systems built on weak data can produce false confidence, unfair outcomes and poor decisions.

SPOTLIGHT

Sequencing Implementation Through the MoSPI Roadmap

Phase 1 – Inventory: Establish mandate, priorities and baselines

Confirm chief secretary- or minister-level sponsorship and identify the central coordinating function. Select three to five priority decisions, services or schemes; name the relevant departmental owners and stewards; inventory the data sets and systems on which they depend; and establish initial baselines for data quality, legal authority, security, exclusion risk and manual-reporting burden. Existing national and state systems should be reused where appropriate, but broad migration should not begin before these foundations are understood.

Phase 2 – Foundational: Fix quality and standards at source

Agree common definitions, identifiers, classifications, reference data and minimum metadata for the priority domains. Replace the highest-risk reporting practices for spreadsheet, email and memo; introduce validation at the point of capture; and establish correction, appeal and feedback routes. Publish the first machine-readable data sets and APIs for priority uses and use early delivery or fiscal results to demonstrate the value of the programme.

Phase 3 – Intermediate: Build governed exchange around priority services

Establish a dynamic catalogue and connect only the data required for agreed use cases through documented APIs, data contracts or controlled analytical environments. Test access controls, lineage, quality thresholds, legal gateways and operational use before expanding. Bring high-value existing services on to the shared operating layer and retire duplicate reconciliation and reporting processes where evidence supports doing so.

Phase 4 – Advanced: Scale, automate and enable AI safely

Move towards source-generated and automated data exchange; scale proven patterns across departments; modernise or retire legacy integrations; and embed multi-year funding, workforce capability, cybersecurity, independent assurance and vendor-exit arrangements. Advanced analytics and AI should be introduced only where data quality, representativeness, human oversight, explanation, appeal and accountability are demonstrably adequate.

Headline indicators should be integrated only once the underlying input, service and outcome data meet agreed quality thresholds.

The proposed sequencing therefore reinforces the MoSPI roadmap by linking each phase more directly to delivery priorities, data quality, institutional accountability and the conditions required for responsible analytics and AI.

The institutional capability required is delivery management. This agenda needs milestones, owners, funding, procurement plans, risk management and ministerial review. It should be managed as a core government reform rather than as a back-office IT project.

The risk it mitigates is premature automation: using advanced tools before state institutions can trust the underlying data, explain resulting decisions or correct errors.

Conclusion

India's first digital-government achievement was scale. Aadhaar, UPI and the platforms built on top of them proved that government technology can reach hundreds of millions of people quickly and reliably. The next phase should focus on converting India's digital scale into stronger institutional and delivery capability. The test is whether states can convert the data flowing through these systems into a resource they can govern and use – reliable enough to target resources precisely, timely enough to correct course before failures accumulate and trusted enough that citizens accept the numbers when they are published.

The practical starting point is deliberately small. A handful of important decisions, a limited set of high-value data sets, named owners, measurable quality baselines and one governed integration pathway that proves value. Scale should follow evidence, not precede it.

That conversion will not happen as a byproduct of more infrastructure. It requires institutions with real authority, standards that are enforced rather than recommended and political leadership willing to hold the effort accountable across more than one budget cycle. State governments that build this capability first will not simply have better dashboards. They will be better able to see problems clearly, act on evidence and correct both the data and the decisions that follow from it.

08

Acknowledgments

The authors would like to thank the following for their input and expertise:

Srinath Chakravarthy, National Institute for Smart Government

Balasubramaniam Gauthaman, Centre of Data for Public Good, Indian Institute of Science

Astha Kapoor, Aapti Institute

Aparna Krishnan, Abdul Latif Jameel Poverty Action Lab (J-PAL)

Vikram Sinha, Artha Global

Shreya Thakur, Wadhvani Institute for Artificial Intelligence

Endnotes

- 1 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1825145&Reg=48&lang=2>
- 2 <https://uidai.gov.in/en/aadhaar-dashboard/enrollment-dashboard>
- 3 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>
- 4 <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2301764®=3&lang=1>
- 5 <https://mospi.gov.in/flipbook-viewer?url=%2Fpdf%2Fdata-harmonisation-playbook.pdf&title=A%20Practitioner%27s%20Handbook%20on%20Data%20Harmonisation>
- 6 <https://www.worldbank.org/en/publication/globalindex/report>
- 7 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1955614&lang=2®=48>
- 8 <https://www.niifindia.in/wp-content/uploads/2025/01/NIIF-Research-2501-Monthly-Economic-Report-on-India.pdf>
- 9 <https://dbtbharat.gov.in/static-page-content/spagecont?id=18>
- 10 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1825145&Reg=48&lang=2>
- 11 <https://egov.eletsonline.com/2025/01/niti-aayog-turns-10-transforming-governance-through-national-data-and-analytics-platform/>
- 12 <https://www.pib.gov.in/FactsheetDetails.aspx?ModuleId=16&NotelId=150834&id=150834&lang=2&Reg=48&>
- 13 <https://360analytika.com/statistical-insight-on-digilocker-in-india/>
- 14 <https://rebootdemocracy.ai/blog/voices-in-every-language-bhashadaan-india>
- 15 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1847833&Reg=48&lang=2>
- 16 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2246222&Reg=3&lang=1>
- 17 <https://openknowledge.worldbank.org/server/api/core/bitstreams/c5f1fc39-7a91-41e6-9965-6355465b5f72/content>
- 18 <https://www.pib.gov.in/PressReleaselframePage.aspx?PRID=2200569&lang=2&Reg=48>

- 19 <https://www.impriindia.com/insights/policy-update/transforming-indias-credit-ecosystem-through-the-unified-lending-interface-uli2025/>
- 20 <https://voxdev.org/topic/social-protection/balancing-corruption-and-exclusion-incorporating-indias-aadhaar-public-food>
- 21 <https://privacyinternational.org/long-read/4468/failures-digitisation-indias-food-security-programme-exclusion-married-women-odisha>
- 22 <https://www.deccanherald.com/amp/story/india/centre-announces-census-2027-schedule-3571126>
- 23 <https://mospi.gov.in/uploads/publications%5Freports/publications%5Freports1780040415321%5F0624fb13-fb47-40bc-b470-7c7e9635c3ef%5FPLFS%5F2025%5FF%5FRE%5F29052026.pdf>
- 24 <https://pmc.ncbi.nlm.nih.gov/articles/PMC12615008/>
- 25 <https://pmc.ncbi.nlm.nih.gov/articles/PMC8324195/>
- 26 <https://www.ovid.com/jnls/jfmpc/fulltext/10.4103/jfmpc.jfmpc%5F1967%5F24-study-on-external-consistency-of-health-management>
- 27 <https://educationforallinindia.com/analysis-of-udiseplus-2022-23-and-2023-24-data-understanding-trends-and-national-education-goals/>
- 28 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>
- 29 <https://www.aninews.in/news/national/general-news/50-govt-websites-hacked-8-data-breaches-in-2022-23-alone-it-minister-informs-rajya-sabha20230203213814/>
- 30 <https://www.niti.gov.in/sites/default/files/2025-09/india-data-imperative-the-pivot-towards-quality.pdf>
- 31 <https://cag.gov.in/uploads/download%5Faudit%5Freport/2022/4--Report-no---5--ENGLISH-Chapter---2-0639c52a04277b3.55097063.pdf>, <https://cag.gov.in/en/audit-report/details/55961>, and <https://cag.gov.in/uploads/download%5Faudit%5Freport/2016/Union%5FCivil%5FNational%5FFood%5FSecurity%5FReport%5F54%5Fof%5F2015.pdf>
- 32 <https://www.careratings.com/>
- 33 <https://www.niti.gov.in/sites/default/files/2025-09/india-data-imperative-the-pivot-towards-quality.pdf>
- 34 <https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/06/data-for->

development-profiles%5Fbc0e3218/84baa8f3-en.pdf

35 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>.

36 <https://rtgs.ap.gov.in/> (access may be restricted outside India)

37 <https://tgdex.telangana.gov.in/>

38 <https://www.medianama.com/2025/07/223-telangana-tgdex-india-first-state-led-ai-data-exchange/>

39 <https://www.projectstatecraft.org/post/abha-card>

40 <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2155449&Reg=3&lang=2>

41 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>

42 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>.

43 <https://carnegieendowment.org/research/2023/06/indias-statistical-system-past-present-future>

44 “A decade into India's open government data journey.” Observer Research Foundation, 2022.

45 <https://www.ideasforindia.in/topics/governance/unleashing-the-full-potential-of-indias-open-government-data-initiative>

46 <https://zenodo.org/records/10678058>

47 “Why do Open Government Data initiatives fail in developing countries? A root cause analysis,” *The Electronic Journal of Information Systems in Developing Countries* (2024).

48 “How is open data changing India?” World Economic Forum (2015).

49 <https://www.orfonline.org/research/towards-evidence-based-policymaking-indias-open-data-initiatives>

50 <https://e-estonia.com/solutions/interoperability-services/x-road/>

51 <https://www.gov.uk/government/publications/identify-and-record-critical-data-assets>

52 <https://dsb.gov.au/>

- 53 <https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/10/the-oecd-digital-government-policy-framework%5F1dd6aa8/f64fed2a-en.pdf>
- 54 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>.
- 55 <https://www.mospi.gov.in/uploads/publications%5Freports/publications%5Freports1781181484342%5F6540ca58-b721-4853-87a6-61d912dbefb5%5FSummit%5FReport.pdf>

Follow us

facebook.com/instituteglobal

x.com/instituteGC

instagram.com/institutegc

General enquiries

info@institute.global

Copyright © October 2026 by the Tony Blair Institute for Global Change

All rights reserved. Citation, reproduction and or translation of this publication, in whole or in part, for educational or other non-commercial purposes is authorised provided the source is fully acknowledged Tony Blair Institute, trading as Tony Blair Institute for Global Change, is a company limited by guarantee registered in England and Wales (registered company number: 10505963) whose registered office is One Bartholomew Close, London, EC1A 7BL.