

Purpose Built: Turning Fragmented Government Data into AI-Ready Infrastructure

A DPI Framework for Governed,
AI-Ready Data Access in India

MARCH 2026



By:
People+ai

Introduction

A handloom business owner in Lucknow asks what export markets are opening up for her products, and an AI bot tells her that "demand for handloom textiles varies with consumer trends." It sounds informed but lacks direction. Ask which platform has rising demand for her price point and weave style, and it recommends a channel suited to an entirely different product category. Ask whether she can sell internationally, and it says yes, without mentioning compliance costs, logistics restrictions, or documentation requirements. These failures are predictable: systems trained on broad, generalized data will produce broad, generalized answers. For merchants, that kind of advice carries a direct financial cost. The models exist. The infrastructure is in place. What's missing is the connection between data held by institutions and the applications trying to use it, and this gap is structural, noted in how government data is organized and maintained. The Ministry of Textiles holds export records with no link to buyer data. Market price databases exist in formats no application can query cleanly. Advisory and policy data is scattered across departments with no common identifiers, no standard definitions, and no defined accountability for keeping it current. The Data Boarding Pass is designed to close this gap.

Why data doesn't travel well

AI systems underperform in public-purpose applications for a predictable set of reasons, and understanding them is the starting point for fixing them. Publicly available data sits in institutional silos where departments have little mandate and lack incentives to share across boundaries. For example, trade tariff data and foreign trade statistics are primarily handled by the "Ministry of Commerce and Industries". Digital data of MSME registration, and benefits like the PMEGP are handled by "Ministry of Micro, Small & Medium Enterprises". Both these ministries work efficiently for the purpose of their departmental oversight. There is no direct incentive for these two ministries to share and combine foreign trade data with MSME benefits data. However, for an entrepreneur who wants to expand internationally and avail export-related benefits, fragmented information spread across separate silos creates a challenge of information asymmetry. Where data is available from multiple sources, it tends to use inconsistent terminology: the same handloom saree described in three different ways across three ministries, the same state spelled differently across

datasets, thus making linkage unreliable. Quality can be difficult to verify, not due to any lack of diligence on the part of institutions, but because they are naturally oriented toward the standards and purposes they were built around, and as new applications of the data continue to emerge, existing structures may not always be positioned to keep pace. An MSME department's job is managing data for its own governance purposes. Making that data useful for an entrepreneur's real-time advisory is adjacent work that falls into a gap between mandates. Add to that the absence of APIs, poor documentation, and no clear pathway for an application developer to cleanly access even publicly owned data, and the picture becomes clear: the problem isn't a shortage of data. It's the absence of a shared framework for moving data from where it lives to where it can do something useful.

What a data boarding pass does



Data Boarding for AI Ready India

Data Boarding Pass unlocks value of data, enables AI impact at population scale; by turning information trapped in fragmented repositories into trusted & connected AI-ready intelligence that can power reliable AI applications, solving real-world problems.

The **Data Boarding Pass** is a conceptual framework that combines technical standards with operational governance to move data from fragmented source systems into AI-ready, connected datasets. These datasets become securely accessible to **Value Creators** within the data ecosystem.

Value Creators include institutional actors such as policymakers and academic researchers who rely on accurate AI insights to address population-scale challenges.

They also include private-sector innovators who build dependable AI applications to solve real-world problems. The framework ensures that all of them can access trustworthy, well-governed data.

At the foundation of the framework is the **Data Passport**.

The Data Passport enables Data Holders to prepare their datasets to a defined, interoperable standard. **A passport-certified dataset is:**

- Discoverable
- Machine-readable
- Published in open formats
- Embedded with explicit permissions
- Backed by verifiable provenance credentials
- Accountable for data quality

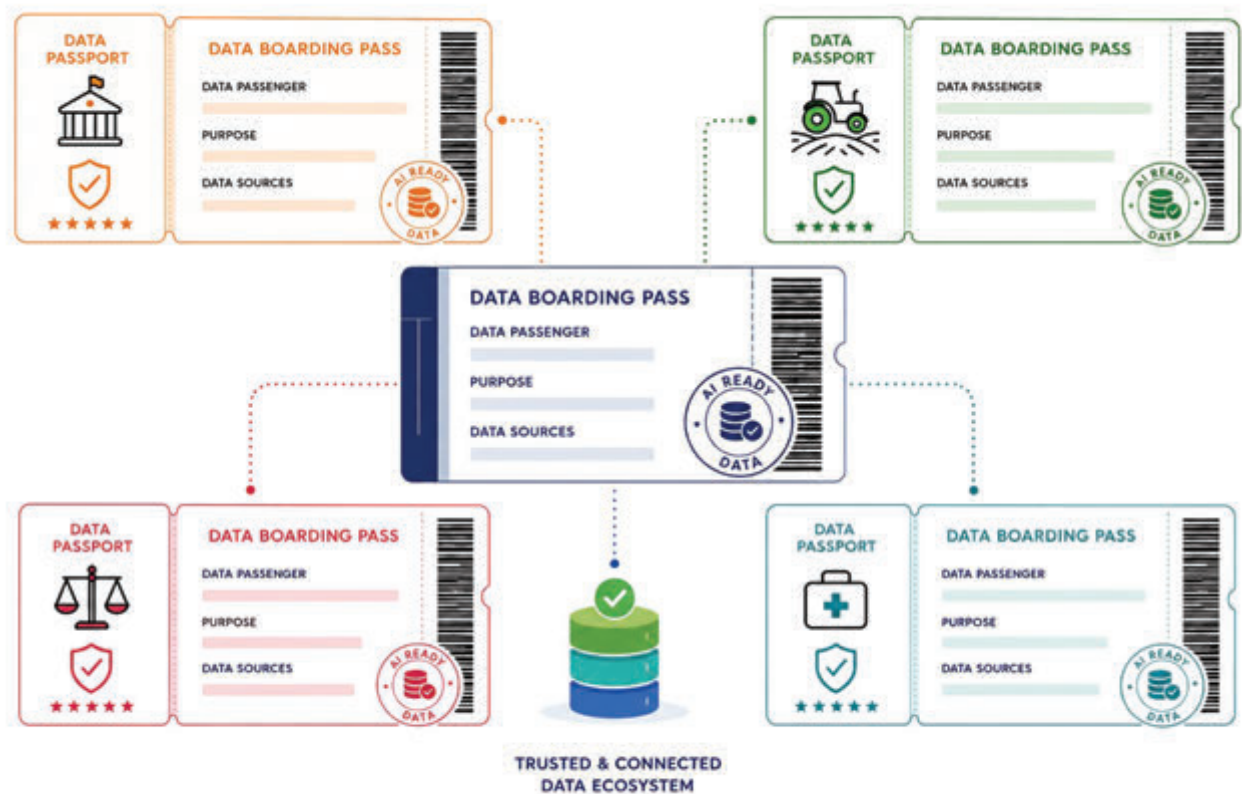
The Data Passport certifies that the dataset is fit for responsible reuse. It is the proof that the data is ready to “board” the ecosystem.



Building on this foundation is the **Data Boarding Pass**.

The Data Boarding Pass is issued to a Value Creator for a specific, declared use case. It grants governed access to multiple datasets (passports) that have been linked, enriched with semantic metadata, and prepared for AI applications.

Importantly, access is always purpose-specific. This is not a constraint but a deliberate design feature. The same datasets do not automatically transfer to new or unrelated use cases. Governance is embedded by design and that assurance is precisely what makes Data Holders willing to share their data in the first place.



The technical aspect beneath these two artefacts implements four steps of the data value chain in the journey of AI readiness

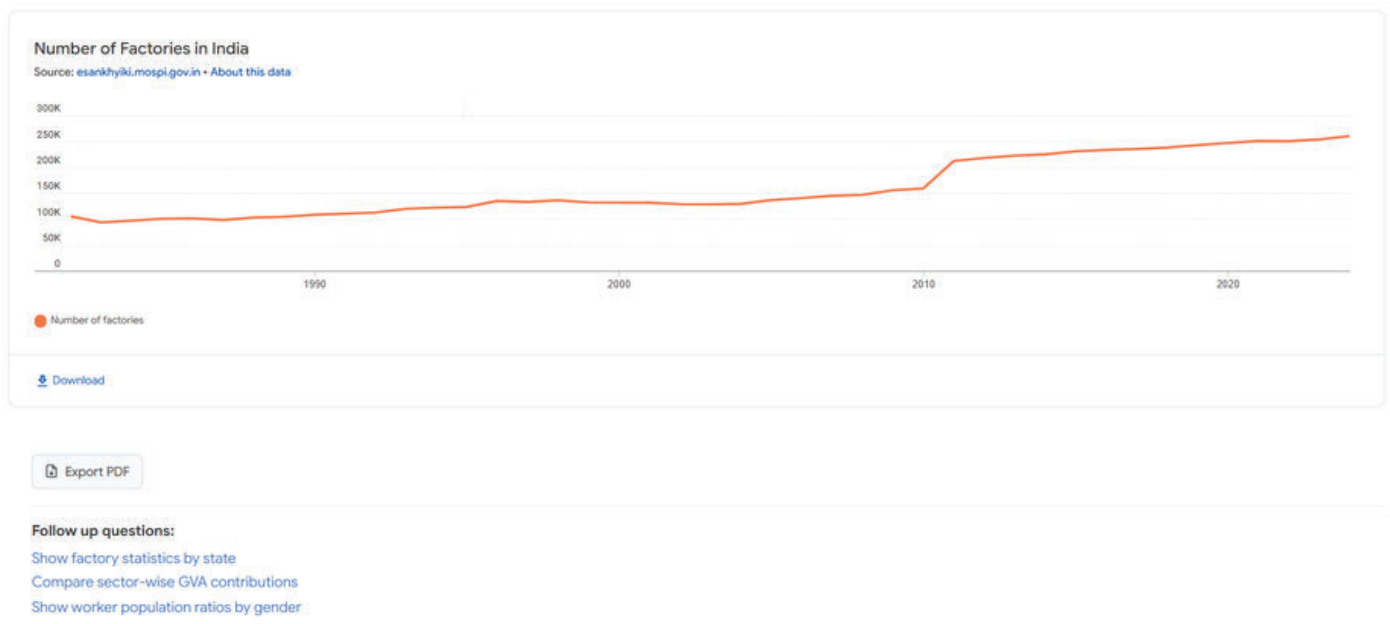
Prepare	Connect	Access	Apply
Discoverability Interoperability Authorization Provenance Accountability	Linkage Entity Resolution Enrichment Semantic MetaData Training Ready	Open access protocols Governed Interfaces API/ MCP/ A2A/ RAG Toolkits	Insight Reports Dashboards Chat Agents Recommendation Engines

What changes when the data is ready

The difference between a generic AI response and a purpose driven one is almost never the model. It's whether the model has access to the right data, at the right granularity, verified and linked correctly. A handloom business owner's question about which export market to enter next gets a meaningfully different answer when the system can draw on geographically-linked trade data, current buyer demand patterns, and verified scheme eligibility rather than a generalised national dataset.

The question is the same. The underlying data readiness is what changes the answer. This applies equally to a policy analyst asking why factory employment in one state diverged from GDP growth, or a farmer trying to understand which mandis are offering the best price for their crop this week, or a district health officer correlating vaccination coverage with facility capacity. In each case, the AI is only as useful as the data infrastructure supporting it.

The India Data Commons prototype demonstrated this directly at the AI Impact Summit: analysts could ask complex, comparative questions across economic and demographic datasets and receive specific, sourced answers in seconds, because the underlying data had been prepared, connected, and made accessible through a governed interface.



Access and explore the prototype here - [Our Prototype](#)

Why this matters beyond any single sector

The ambition of the Data Boarding Pass scales precisely because it doesn't try to solve everything at once. Each boarding pass is scoped to a specific use case, a specific value creator, a specific set of data sources. But the passports and the standards beneath them are reusable. Once a dataset meets passport standards, it can board multiple use cases. Once a value chain for entrepreneurs is established, the pattern applies to health, to finance, to urban infrastructure.

What this creates, over time, is not a single data platform but a shared protocol layer across India's public and private data driven by a common use case. The sectors being worked on today like agriculture, dairy, MSME, governance, investment, represent hundreds of millions of people whose interactions with government and markets are shaped by data quality they can't see and rarely benefit from directly.

Getting those datasets to the Data Passport standard changes the ceiling for every application built on top of them.

The real leverage in AI adoption for public good is not the AI model. It has always been the mechanism that gets relevant and reliable data to the right use at the right time. The Data Boarding Pass is, finally, a systematic way to build that mechanism for achieving a connected public data ecosystem for AI readiness.

ACKNOWLEDGEMENTS

This article was written with inputs from:

1. Aparajita Choudhury K, Solutions, Data & AI, **People+ai**
2. Pooja Manjooran, Program, Data & AI, **People+ai**
3. Anuj Gupta, Product, Data & AI, **People+ai**
4. Sonia Menezes, Research Writer, **People+ai**
5. Shalini Kapoor, Chief Strategist, Data & AI, **EkStep Foundation**
6. Tanvi Lall, Director, Strategy, **People+ai**