Making NSO Data AI-Ready: A Case Study from NSO, India

Enabling National Statistics Offices for the AI Era through NMDS

National Statistics Office
Ministry of Statistics and Programme Implementation

Date : 2nd Feb,2026

Why AI Readiness Matters

- Shift to API & AI-based data consumption patterns by stakeholders
- NSOs Provides automated access via APIs for AI systems
- AI systems accessing information from other sources may lead to mistrust for NSO
- Lessen risk of misinterpretation
- Preserving statistical intent

Challenge: Ensuring NSOs remain the trusted source of statistical information in an AI-driven world

What AI Readiness Means

Data and metadata readiness for:

Machine-Readability

Data that can be automatically processed and understood by AI systems without human intervention

Interpretability

Clear context and meaning that enables accurate interpretation by automated systems

Actionable with Minimum Human Intervention

Structured in ways that allow AI systems to utilize data directly for analysis and decision-making

Proposed Solutions

Machine-Readable Components:

- **Catalog :** Indicators, Identifiers, URLs
- **Indicator Metadata :** Machine readable Metadata
- **Data Structure :** Dimensions, measures, attributes, code lists
- **NSO Vocabulary Template :** Concept name, NSO specific definition, Code list
- **Context File :** Semantic meaning

Target Architecture for AI Readiness

Core Components

- → Discovery
- → Meaning/Context
- → Structure
- → Codes
- → Data

Alignment With SDMX Principles

- → Registries
- → Controlled vocabularies
- → DSD-like structures
- → Dimension and attributes
- → Code lists
- →Measure

Components available in Machine readable format

NSO Foundations

- **NMDS 2.0 :** Data set level metadata standards exposed by APIs through metadata portal
- **Datasets Identifiers:** Structure available through National Data Portal
- **eSankhyiki :** Structured data arrangement with API and availability of observation
- **Classifications :** Follows standard codes and classification
- **Governance :** Established data governance

Foundation Strength: These existing systems provide a solid base for AI readiness implementation

Gaps Identified

Machine-readable components needed:

Catalog for Discovery

Enable AI systems to discover available datasets and indicators

NSO Vocabulary for Uniform Interpretation

Standardized terminology ensuring consistent understanding across systems

Semantic Context Files for Understanding Meaning

Rich contextual information that preserves statistical intent

Templates Needed

To institutionalize AI readiness, Six templates are needed for creating machine readable file like JSON exposed by APIs::

- Machine-readable catalog template (Indicators, Identifiers, URLs)
- Indicator Metadata Template (NMDS 2.0)
- Data Structure Template at Indicator level (dimensions, measures, attributes, codelists)

Templates needed

- NSO Vocabulary Template (concept name, NMDS-sourced definition, NMDS reference)
- Context File Template at Indicator leel (field-to-concept semantic mappings)
- Data File for accessing the actual observation at Indicator level

Conclusion & Recommendations

- Generative AI works both as a challenge and an opportunity for data readiness
- Formats and structures required to effectively ground generative AI applications in data are not in place
- AI Readiness is more a Data and Metadata challenge
- AI readiness can be anchored with NMDS 2.0 and eSankhyiki portal
- Wrapped by NSO vocabulary, context, and discovery mechanisms
- It is a resource intensive task as machine understandability is needed at indicator level

Building on existing foundations while addressing identified gaps will position NSO as trusted and primary source of data in the AI era

Thank You

Rohit Bhardwaj, DDG,NSO, India
rohitb.rgi@nic.in