Generative AI & Official Statistics



Presentation
On

Key Enablers & Use Cases For AI adoption in NSOs: Experience of NSO, India

Rohit Bhardwaj, DDG
National Statistics Office, India

Background

Challenges Faced By NSO

- Resource constraints in government statistical offices
- Exponential growth in data volume and variety
- Increasing demands for timely and granular statistics
- Declining survey response rates
- Skills gap between traditional statistical expertise and data science
- Risk of technological inequality between organizations

Global Context

- Canada: Digital Service achieved 61% faster compliance reviews
- Estonia: "Government as a Platform" reduced admin burden by 52%
- Finland: "AI Register" enabled 18 Gen AI apps in urban services

Discussion Points

- **Strategic adoption: What are the key strategic steps for integrating AI as a tool in national statistical offices?**

- **Practical applications: In which areas can AI be applied to increase the quality of official statistics, the efficiency of the data lifecycle, and the potential for innovation?**

- **Governance and ethics: How can national statistical offices identify and mitigate risks related to bias, ethical and legal concerns, and public trust in statistics?**

Enablers for Integrating AI in NSO's Workflow

- **Leadership Support- Leadership support is the key to innovation**
- **Assessment of Readiness- A comprehensive assessment of readiness to adopt is the first step**
- **Identification of Use Cases- Suitable use cases for embedding AI in the workflow**
- **Enabling Institutional Mechanism- An Institutional mechanism to support innovation should be available**
- **Collaboration- Collaboration with Institutions , Industry & Non voluntary sector**
- **Capacity- Availability of Minimum thresh hold capacity in terms of domain and technology should be available**

There has to be a method in madness

Framework for Assessing Institutional Readiness For Adoption of Emerging Technologies like AI/ML/Big Data in Official Statistics



Scan to access Tool of Institutional Readiness

Link: https://mospi.gov.in/sites/default/files/announcements/Institute%20Readiness_VF_L.pdf

Scoring Mechanism

The assessment is split into 6 themes

Theme	Description
1. Generic Information	Leadership roles & strategies for AI/ML/Big Data.
2. Strategic Coordination	Collaboration with data partners and sharing formats.
3. Data Quality & Readiness	Standards, metadata, QA staff and interoperability.
4. Policy Framework	Legal and ethical governance framework for emerging tech.
5. IT Infrastructure	Storage, compute power, cloud security and risk mitigation.
6. Human Resources	Availability of skilled staff and defined skill requirements.

- Each theme consists of a set of questions, with each question being scored on a **scale from 0 (absence of required features) to 3 (full implementation and adherence).**
- Equal weightage across questions to ensure balance

Score Calculation

Scoring Methodology to assess Institutional Readiness :

The score for each theme is calculated by **simple averaging the scores of the questions** within that theme.

Scores are calculated **only at the theme level** and not at the overall level.

Aggregate score is matched with the type given below.

Level	Score Range
Pre-Foundation *(Just beginning to explore big data)*	Aggregate Score < 0.5
Foundation *(Developing leadership, pilot projects)*	0.5 ≤ Aggregate Score < 1.5
Practitioner *(Frameworks established, strategic management)*	1.5 ≤ Aggregate Score < 2.5
Expert *(Embedded data science, skilled staff)*	Aggregate Score ≥ 2.5

The Use Cases

1. Develop an AI-powered tool **to streamline retrieval of specific information from manuals, documents, and web** for the MoSPI

2. Develop an AI-based tool to **extract, process, and analyze legacy data from diverse formats**, enabling efficient access, statistical analysis, and visualization for improved decision-making

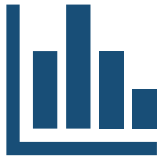
3. Build an **automated algorithm to predict NIC and NCO codes from text descriptions** for efficient industry and occupation classification

4. **AI Powered Chatbot for NSO Website**

The Use Cases contd.

5	Use of AI to undertake semantic search in a data portal
6	Use of AI to classify Scanner Data as Per COICOP to construct CPI
7	ML-based prediction model for MPCE using Household Characteristics
8	AI-based call-based Survey tool
9	Updating UFS Block Maps with Satellite imagery using VLM

The Data Innovation Lab



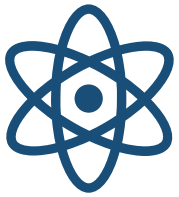
Explore New/big Data Sources
(IoT, e-commerce, GIS, Satellite Imagery, etc.)



Leverage New Innovative Tech.
(Use of Gen AI/ML)



Research on methodological innovations



Cover cross-cutting themes
(utilizing big data for CPI)

Data Innovation Lab was operationalised on 23rd July 2024

- Vision: Revolutionize official statistics through Generative AI
- Alignment: Goal of "Viksit Bharat"

Approaches to Data Innovation

Consolidation of Use Cases	At present, there are nine problem Statements available. Two have been placed in the public domain, seeking solutions
In house POCs	POCs have been undertaken on identified problem statement to understand the bottom line
Research	Internal research is undertaken to examine feasibility and to consolidate global best practices
Collaboration and feedback	Engage actively with academic institutions, multilateral agencies and non-governmental partners
Documentation and Reporting	Prepare necessary documentation/working papers.

The Four-Pillar Collaborative Framework

- Government (NSO India , Ministry of Education & MeitY)
- Academic Institutions (IITs, IIMs)
- Students (Hackathons & Projects)
- Startups (Tech innovators & incubators)

The Collaborations

Key Collaboration	Key Arrangement Themes	Outcomes/Benefits
IIT Madras	Collaboration for undertaking projects	IIT Madras has expressed interest in execution of four projects under DI Lab in collaboration with Sarvam AI.
IIIT Delhi		The Incubator has submitted ten projects for collaboration.
IIT Gandhinagar		A hackathon has been organised.
IIM Ahmedabad		A workshop has been held to identify mutual areas of Interest.
Sarvam AI	Demonstration of Use cases for the Official Statistics system	A use case for AI-based data collection has been planned.
IIM Ahmadabad	Provides overarching framework of collaboration	A workshop has been organised to finalise the area of collaboration
M/s Edubuild Pvt Ltd(Start up)	To develop use cases related to AI-based tools for facilitating field works and rejuvenation of legacy data	Work has started on the use cases with financial support from NSO, India.
Ministry of Education	Collaboration to conduct a Hackathon	Hackathon for five use cases planned in June 2025

The Outreach

1. MoUs signed with 15 Academic institutions
2. Two Hackathons held
3. 17 mentors empanelled from IIT, NPCI, NASSCOM, RBI etc.
4. 8 Pilots/POCs underway
5. Participation in workshops to sensitise stakeholders;
   IIM Ahmedabad: Joint workshop

Knowledge Repository

1. Working paper and self assessment tools for readiness for adoption of AI/ML
2. Working Papers on best practices for Data Dissemination & Portals
3. Compilation of research paper of the last five years on use of AI/ML and alternative data sources in Official Statistics
4. Membership of UNCEBD & TT on AI , PARIS21, India is Offered seat in Bureau of UNCEBD

Lessons & Way Forward

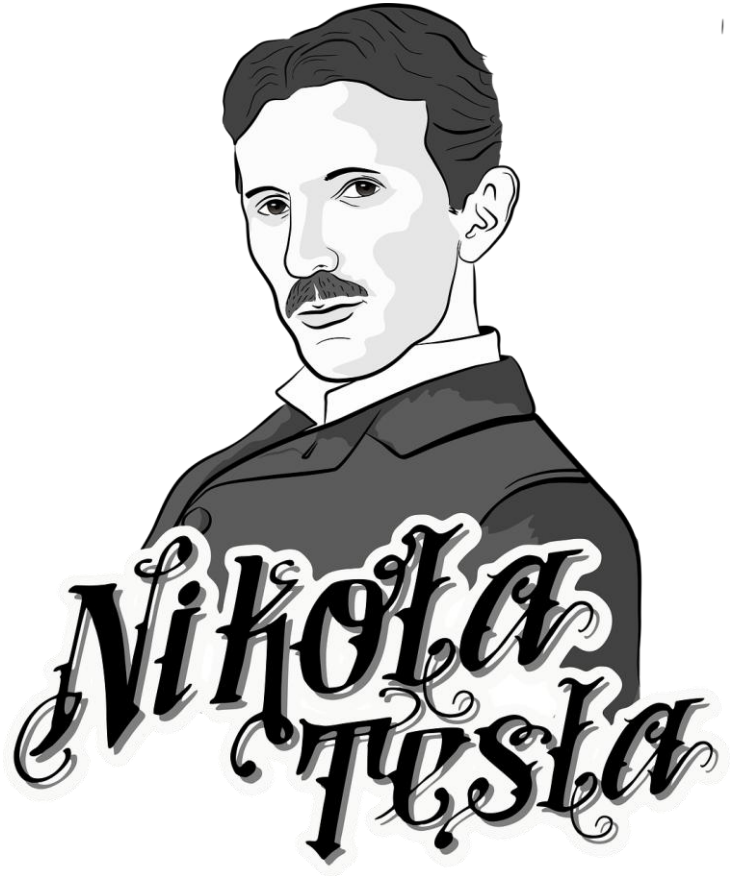
LESSONS LEARNED

- Be proactive in partnerships
- Define clear use cases
- Multi-stakeholder models work
- Innovation vs accountability needs balance
- Knowledge sharing accelerates adoption

WAY FORWARD

- Secure dedicated funding
- Attract AI/data science talent
- Formalize ethical Gen AI guidelines
- Build evaluation frameworks
- Enable knowledge sharing across govt.

It is Just a Beginning!!



Inventor Of Alternative Current

"The scientific man does not aim at an immediate result. He does not expect that his advanced ideas will be readily taken up. His work is like that of the planter - for the future. His duty is to lay the foundation for those who are to come and point the way."

-Nikola Tesla

Questions??

Contact:
- Rohit Bhardwaj, DDG, NSO India(rohitb.rgi@nic.in)