



WORKING PAPER ON THE ASSESSMENT OF INSTITUTIONAL READINESS FOR THE ADOPTION OF EMERGING TECHNOLOGIES SUCH AS AI, ML, AND BIG DATA IN OFFICIAL STATISTICS



Data Informatics & Innovation Division

**Ministry of Statistics & Programme Implementation,
Government of India
December, 2024**

This page has been left blank intentionally

**WORKING PAPER ON THE ASSESSMENT
OF INSTITUTIONAL READINESS
FOR THE ADOPTION OF EMERGING
TECHNOLOGIES SUCH AS AI, ML, AND
BIG DATA IN OFFICIAL STATISTICS**

Data Informatics & Innovation Division

**Ministry of Statistics & Programme Implementation
Government of India
December, 2024**

FOREWORD

राव इन्द्रजीत सिंह
RAO INDERJIT SINGH



राज्य मंत्री (स्वतंत्र प्रभार)
सांख्यिकी और कार्यक्रम कार्यान्वयन मंत्रालय;
राज्य मंत्री (स्वतंत्र प्रभार) योजना मंत्रालय तथा
राज्य मंत्री संस्कृति मंत्रालय
भारत सरकार
Minister of State (Independent Charge) of the
Ministry of Statistics and Programme Implementation;
MOS (I/C) of the Ministry of Planning and
MOS in the Ministry of Culture
Government of India

FOREWORD

In today's rapidly evolving technological landscape, the role of emerging technologies such as Artificial Intelligence (AI), Machine Learning (ML), and Big Data has become paramount. These advancements have transformed the global data ecosystem, providing governments and institutions new tools to handle, process, and analyse vast amounts of data more effectively.

The Ministry of Statistics and Programme Implementation (MoSPI), as the nodal agency for Official Statistics in India, recognizes the need to mainstream these technologies into our Statistical processes. Traditional methods of data collection and analysis are no longer sufficient to meet the growing demand for more granular, timely, and comprehensive data. As such, leveraging emerging technologies like AI, ML, and Big Data is essential to cater to the growing demand of users.

This working paper presents a comprehensive framework designed to assess the readiness of various Ministries and Departments of the Government in adopting these emerging technologies in the field of Official Statistics. The framework addresses several critical aspects, including strategic coordination, data quality, IT infrastructure, policy frameworks, and human resources. It offers a self-assessment tool that enables Government entities to evaluate their current capacity and identify areas for enhancement in line with these advancements.

Our vision is to create a robust, future-ready National Statistical System that is equipped to handle the challenges of the present & future. By embracing these technologies, we aim to foster a culture of data-driven decision-making, improve transparency, and strengthen public trust in the quality of official statistics.

I encourage all Ministries and Departments to actively participate in this initiative and utilise this framework to ensure their preparedness for the future. The success of this endeavour will ultimately contribute to more informed governance, better policy making, and enhanced service delivery to the citizens of India.

(Inderjit Singh)

EXECUTIVE SUMMARY

The Ministry of Statistics and Programme Implementation is the nodal ministry for Statistics in the country. Traditionally, the Ministry handles huge volume of data regularly. In addition to this Ministry, there are other Ministries/Departments of Government of India who deal with voluminous data, leading to production of Official Statistics.

In today's world, the data ecosystem has undergone a paradigm shift. Large and complex sets of data (also referred to as Big Data) are replacing the traditional data, rendering the conventional data processing system inadequate. Big data is characterized by its volume, velocity, and variety, and can include structured, unstructured, and semi-structured data. These pose new challenges across a range of areas, including methodology, quality assurance, technology, security, privacy, legal matters and skills.

This working paper introduces an assessment framework designed to help Ministries and Departments self-evaluate their readiness to adopt emerging technologies like AI, ML, and Big Data for producing official statistics. The framework focuses on six key themes: *Generic Information, Strategic Coordination, Data Quality and Readiness, Policy Framework, IT Infrastructure, and Human Resources*. The aim of the tool is not to rate any Ministry/Department. Rather, the framework offers a self-assessment tool aimed at helping government entities identify areas for improvement and ensure their institutional preparedness for adoption of emerging technology.

By adopting this framework, Ministries and Departments can improve their capacity to handle complex data sets, enhance data quality, and promote more effective data dissemination. Ultimately, this effort will foster greater transparency, support data-driven decision-making, and increase public trust in official statistics, contributing to better governance and accountability.

TABLE OF CONTENTS

1. INTRODUCTION	7
2. BACKGROUND	7
3. ASSESSMENT	7
4. KEY ELEMENTS FOR ASSESSMENT Of INSTITUTIONAL READINESS	8
5. SCORING METHODOLOGY	9
6. FINAL GRADING	15
7. CONCLUSION	16

INTRODUCTION

The Ministry of Statistics and Programme Implementation is the nodal ministry for Official Statistics in the country. The Ministry intends to undertake modernization programme to improve the official statistical system. Under such initiative, one of the key roles of the Ministry is to undertake Technical Upgradation. In order to augment that, the Ministry is exploring possibilities of mainstreaming emerging technology, like usage of AI/ML and alternative sources of data including big data in the field of official statistics.

The purpose of the current study is to assess the institutional readiness of such Ministries/Departments and the official statistical system as a whole, regarding the adoption of AI/ML/Big Data for providing more granular and higher frequency data.

BACKGROUND

The data ecosystem has undergone a tectonic shift in the last decade or two. With the rising use of digital tools, there has been a data deluge. Data has become available as a byproduct of all our actions in digital environment. With the availability of such huge data, the National Statistical Offices (NSOs) are expected to provide statistics at a more granular level and with higher frequency.

United Nations Statistics Division (UNSD) has established an expert committee on Big Data (UN-CEBD) in 2014 to help NSOs adopt and collaborate in this emerging field. In view of the data demand, time has come to mainstream these technologies and data sources in the National Statistical System. Accordingly, through this paper it is our endeavor to assess the institutional readiness of the constituents of the Official Statistical System for adopting new technology and data sources. The purpose of this working paper is also to create a quantitative assessment tool for doing the same.

It is clarified that it is not a rating mechanism. This is an attempt to find the status of the readiness of various government Departments and Ministries to adopt emerging technologies so that a comprehensive plan for the government as a whole can be prepared so far as the official statistics is concerned.

ASSESSMENT

The assessment tool can be used to determine the institutional readiness of Ministries/ Department for using emerging technologies like AI/ML/Big Data. The tool is attached at [Annexure-I](#).

This has been prepared by taking reference from the document “Global assessment of institutional readiness for the use of big data in official statistics¹” by UN Global Working Group (GWG) on Big Data for Official Statistics.

KEY ELEMENTS FOR ASSESSMENT OF INSTITUTIONAL READINESS

In order to ensure Institutional Readiness for adoption of Emerging Technologies like AI/ML/Big Data in Official Statistics by the Ministry/ Department, the following key elements are looked at:

S.No	Features	Description
1.	Generic Information	Assesses if the organization has foundational strategies and key leadership roles for AI/ML/Big Data in place.
2.	Strategic Coordination	Evaluates the organization’s ability to collaborate with data partners and manage data sharing for AI/ML projects.
3.	Data Quality and Readiness	Checks if data standards, metadata, and quality assurance are established to handle big data.
4.	Policy Framework	Determines if a governance or legal framework is followed in the adoption of emerging technologies.
5.	IT Infrastructure	Examines if the organization has sufficient technology resources like storage, computing power, and secure cloud infrastructure for big data.
6.	Human Resources	Assesses whether skilled staffs are in place for data science roles and required skills are clearly defined.

¹ Global assessment of institutional readiness for the use of big data in official statistics (taken on 23rd Sept, 2024): https://unstats.un.org/bigdata/task-teams/training/UN_BigData_report_v5.0.html#executive_summary

SCORING METHODOLOGY

The assessment is split into 6 themes viz. *Generic Information, Strategic Coordination, Data Quality and Readiness, Policy Framework, IT Infrastructure, and Human Resources.*

1. There are six themes as indicated above.
2. Each theme consists of a set of questions, with each question being scored on a scale from 0 to 3. A score of 0 indicates absence of required features, while a score of 3 reflects full implementation and adherence.
3. The score for each theme is calculated by simple averaging the scores of the questions within that theme.

The detailed question-wise scoring mechanism is illustrated as per the following table (M: Ministry; D: Department).

Q.No	Question	Scoring Mechanism
THEME 1: Generic Information		
1a.	Does the M/D have any AI/ML/Big Data (Data Science) strategy in place?	<p>If the Ministry/ Department (M/D) has a fully codified strategy, it would receive a score of 3.</p> <p>If draft strategy is in place, the M/D would receive a score of 2.</p> <p>If the efficacy of AI/ML/Big Data (Data Science) within the M/D is being studied, the M/D would receive a score of 1.</p> <p>If there is no AI/ML/Big Data (Data Science) strategy in place, the M/D would receive a score of 0.</p>

Q.No	Question	Scoring Mechanism
1b.	Is there any designated Chief Data Officer in the M/D?	<p>If the M/D has a full-time designated Chief Data Officer, it would receive a score of 3.</p> <p>If the Chief Data Officer role is present but assigned as an additional duty, the M/D would receive a score of 2.</p> <p>If there is no Chief Data Officer but a division for data is present, the M/D would receive a score of 1.</p> <p>If there is no separate data department or lead Chief Data Officer, the M/D would receive a score of 0.</p>
1c.	Is the Ministry/ Department (M/D) undertaking any project involving AI/ML/Big Data currently?	<p>If the M/D has undertaken AI/ML/Big Data project(s) and it is in the production stage, it would receive a score of 3.</p> <p>If the M/D has started with a project, and it is in PoC stage, it would receive a score of 2.</p> <p>If the M/D has not yet undertaken a project but PoC is expected to start off within the next three months, it would receive a score of 1.</p> <p>If there is no plan to undertake any AI/ML/Big Data projects, the M/D would receive a score of 0.</p>
THEME 2: Strategic Coordination		
2a.	What are the different data owners with whom the M/D is collaborating for its AI/ML/Big Data (Data Science) Project(s)? (Data owners are indicated in options) (The exact number of "Others' data owners would also be used for calculating the scores; it shall be indicated as a note in the questionnaire).	<p>It is scored 3 if five or more options are selected.</p> <p>It is scored 2 if three or more but less than five options are selected.</p> <p>It is scored 1 if one or more but less than three option selected.</p> <p>It is scored as 0 if none of the options are selected.</p>

Q.No	Question	Scoring Mechanism
2b.	<p>Has the M/D discussed/ negotiated a format for provision of data by data-partners?</p>	<p>If the Ministry/Department (M/D) has discussed and agreed on a standard format and it has been circulated to the stakeholders, it would receive a score of 3.</p> <p>If the standard format for data provision has been formulated but is yet to be circulated to the data-partners, the M/D would receive a score of 2.</p> <p>If the M/D is in the process of finalizing standard format, it would receive a score of 1.</p> <p>If there is no plan to negotiate or discuss standard format, the M/D would receive a score of 0.</p>
2c.	<p>Hindrances that the M/D may be facing in implementing Data Science techniques (does not apply to M/D who has no plan for AI/ ML/Big Data projects and accordingly they will not score any marks for this question)</p>	<p>It is scored 3 if one to two options are selected.</p> <p>It is scored 2 if three to four options are selected.</p> <p>It is scored 1 if five to six options are selected.</p> <p>It is scored 0 if more than six options or no option is selected.</p>

Q.No	Question	Scoring Mechanism
THEME 3: Data Quality and Readiness		
3a.	<p>Does the M/D follow any Data and Meta data standards for the implementation process?</p>	<p>If the Ministry/Department (M/D) has fully adopted data and metadata standards, it would receive a score of 3.</p> <p>If the M/D has a big data and meta data standard but is yet to be implemented, the M/D would receive a score of 2.</p> <p>If the M/D is in the process of establishing data and metadata standards, it would receive a score of 1.</p> <p>If no standards are planned for data and metadata, the M/D would receive a score of 0.</p>
3b.	<p>Does the M/D have a framework/guideline for interoperability? (Interoperability means the ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways.)</p>	<p>If the M/D has an implemented framework for interoperability, it would receive a score of 3.</p> <p>If the M/D has a framework for interoperability but is yet to be implemented, the M/D would receive a score of 2.</p> <p>If the M/D is trying to create/adopt a framework for interoperability, it would receive a score of 1.</p> <p>If no framework for interoperability is planned, the M/D would receive a score of 0.</p>
3c.	<p>Does the M/D have a Data Quality Assurance staff in place?</p>	<p>If the Ministry/Department (M/D) has a dedicated Quality Assurance (QA) staff, it would receive a score of 3.</p> <p>If efforts are underway to establish a dedicated QA staff, it would receive a score of 2.</p> <p>If there is no dedicated staff for QA framework for big data but staff doing other quality checks are also engaged in this work, the M/D would receive a score of 1.</p> <p>If there is no plan to have a dedicated QA staff, the M/D would receive a score of 0.</p>

Q.No	Question	Scoring Mechanism
THEME 4: Policy Framework		
4a.	<p>Does the M/D have any governance/ legal/ data ethics & privacy framework available?</p>	<p>If the M/D has a governance/legal/data ethics & privacy framework available by law, it would receive a score of 3.</p> <p>If the M/D has a governance/ legal/data ethics & privacy framework available by policy, it would receive a score of 2.</p> <p>If there are plans to establish a governance/legal/ data ethics & privacy framework but it is not yet implemented, it would receive a score of 1.</p> <p>If no governance/legal/data ethics & privacy framework.</p>
THEME 5: IT Infrastructure		
5a.	<p>Is required infrastructure available (such as storage, compute, etc.) for on-going projects?</p>	<p>If the required infrastructure is available, it would receive a score of 3.</p> <p>If the process of setting up the required infrastructure is underway, it would receive a score of 2.</p> <p>If there is a plan to establish the required infrastructure but it is not yet implemented, it would receive a score of 1.</p> <p>If there is no plan to establish the required infrastructure, it would receive a score of 0.</p>

Q.No	Question	Scoring Mechanism
5b.	Is there a secured cloud infrastructure?	<p>If a secured cloud infrastructure is fully implemented, it would receive a score of 3.</p> <p>If the process of setting up secured cloud infrastructure is underway, it would receive a score of 2.</p> <p>If there are plans to establish secured cloud infrastructure but it is not yet in place, it would receive a score of 1.</p> <p>If there is no plan to implement secured cloud infrastructure, it would receive a score of 0.</p>
5c.	Does the M/D have a risk mitigation plan for addressing data breaches, system failures, and compliance risks?	<p>If there is a risk mitigation plan for addressing data breaches, system failures and compliance risks, it would receive a score of 3.</p> <p>If there is a process of setting up of a risk mitigation plan for addressing data breaches, system failures and compliance risks, it would receive a score of 2.</p> <p>If there is a plan to device a risk mitigation plan for addressing data breaches, system failures and compliance risks, it would receive a score of 1.</p> <p>If there is no plan to have a risk mitigation plan for addressing data breaches, system failures and compliance risks, it would receive a score of 0.</p>

Q.No	Question	Scoring Mechanism
THEME 6: Human Resources		
6a.	Are there employees at M/D actively applying data science/big data techniques in ongoing projects?	<p>If employees are actively applying data science or big data techniques, the organization would receive a score of 3.</p> <p>If no employees are currently applying these techniques but the hiring process is in progress, the organization would receive a score of 2.</p> <p>If the hiring process is being set up but not yet active, it would receive a score of 1.</p> <p>If there is no hiring process set up for data science or big data roles, the organization would receive a score of 0.</p>
6b.	If a project requires data processing, are the specific skill requirements clearly stated in the context of relevant projects or roles where data processing is needed?	<p>If processing the data requires specific skills and those requirements are clearly stated, the M/D would receive a score of 3.</p> <p>If the process of setting up skill requirements for processing data is underway, the M/D would receive a score of 2.</p> <p>If there are plans to establish skill requirements but M/D has not yet started the process, it would receive a score of 1.</p> <p>If no skill requirements for processing data are planned, the M/D would receive a score of 0.</p>

FINAL GRADING

1. Once the assessment is completed, the Ministry/Department receives a level of Pre-Foundation, Foundation, Practitioner, or Expert for each individual theme.
2. To arrive at the level, 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.

A simple average is prescribed, as each question is considered of equal importance to ensure a comprehensive assessment of the Ministry/Department's readiness in that specific theme. Further, the variation across the scoring range can be maximum 3. Hence, the effect of extreme observation will not be substantial.

Aggregate scores are matched with the type given in table below.

Level	Score Range	Significance of Level
Pre-Foundation	Aggregate Score < 0.5	The organization is just beginning to explore big data strategies and projects.
Foundation	$0.5 \leq$ Aggregate Score < 1.5	Big data leadership and strategies are being developed, with a few pilot projects in progress.
Practitioner	$1.5 \leq$ Aggregate Score < 2.5	Frameworks are established, data scientists are in place, and big data projects are strategically managed.
Expert	Aggregate Score \geq 2.5	Data science is deeply embedded, with skilled staff leading projects, offering internal and external training.

An illustration on how to fill the self-assessment tool is provided in the [Annexure - II](#) titled "An Illustrative Example to Assess Institutional Readiness in Big Data".

CONCLUSION

The growing importance of emerging technologies like AI/ML/Big Data in official statistics necessitates a structured approach to assess institutional readiness. The assessment framework given in this working paper provides a comprehensive tool for evaluating the institutional readiness of Ministries and Departments in adopting emerging technologies such as AI/ML/Big Data and identifying areas for improvement. The adoption of these technologies is crucial for modernizing official statistics, enhancing data quality, and meeting the growing demand for more granular and timely information. Ultimately, this effort will help the government create a more responsive and effective data ecosystem that supports evidence-based policymaking.

Disclaimer

Assessment and recommendations made in this working paper do not overlap with the work conducted by the Ministry of Electronics and Information Technology (MeitY). While MeitY focuses on IT infrastructure and technological guidelines for facilitating Data Science, Mo-SPI's job is to use this infrastructure and guidelines to provide better data to policy makers. Our intent is not to duplicate the work of MeitY and the ambit of this paper is only to assess institutional readiness for adopting emerging technologies in the field of Official statistics. This distinction is important to promote a holistic approach in government to adoption of Emerging Technologies like AI/ML/Big Data in Official Statistics.



Data Informatics & Innovation Division

**Ministry of Statistics & Programme Implementation
Government of India**

December, 2024

**Self-Assessment
of Institutional
Readiness for
Adoption of
Emerging
Technologies
like AI/ML/Big
Data in Official
Statistics**



A. Organisational Information

S. No.	Particulars	Details
1	Ministry/Department (M/D) Name	
2	Data products of the M/D	1. 2. 3. 4.
3.	Details of the nodal officer	
	a. Name	
	b. Designation	
	c. E-Mail ID	

B. Questionnaire

(M: Ministry, D: Department)

Q.No.	Question	Scoring Mechanism
1. Generic Information		
1 a.	Does the M/D have any AI/ML/Big Data (Data Science) strategy in place?	A. Fully codified AI/ML/Big Data (Data Science) strategy in place B. Draft strategy is in place C. The efficacy of AI/ML/Big Data strategy (Data Science) is being studied D. No AI/ML/Big Data strategy (Data Science) in place
1 b.	Is there any designated Chief Data Officer in the M/D?	A. Full-time designated Chief Data Office B. Chief Data Officer role is present but assigned as an additional duty C. No Chief Data Officer but a division for data is present D. No separate data department or lead Chief Data Officer
1 c.	Is the M/D undertaking any project involving AI/ML/Big Data currently?	A. AI/ML/Big Data project(s) in the production stage B. AI/ML/Big Data project(s) in PoC stage C. PoC is expected to start off within the next three months D. No plan for AI/ML/Big Data projects
2. Strategic Coordination		
2 a.	How many of the different data owners with whom the M/D is collaborating for its AI/ML/Big Data (Data Science) Project(s)? (Data owners are indicated in options) (The exact number of "Others" data owners would also be used for calculating the scores; it shall be indicated as a note in the questionnaire).	A. Other Govt. M/D, (including departments of same Ministry) B. Satellite image provider, C. Research Institute, D. Mobile phone operator, E. Social Media provider, F. Any other source (Please specify): _____

Q.No.	Question	Scoring Mechanism
2 b.	Has the M/D discussed/negotiated a format for provision of data by data-partners?	<ul style="list-style-type: none"> A. Standard format circulated to stakeholders Question not answered B. Standard format formulated but not yet circulated C. Standard format in the process of finalization D. No plan to negotiate a standard format
2 c.	How many hindrances the M/D may be facing in implementing Data Science techniques (Does not apply to M/D who has no plan for AI/ML/Big Data projects and accordingly they will not score any marks for this question)	<ul style="list-style-type: none"> A. Collaboration with data source owners outside Government B. Collaboration with data source owners within Government C. Legislative Issues D. Human Resources E. Financial Issues including but not limited to cost of sourcing/storing the data F. Privacy issues related to public trust G. Methodological aspects H. Information technology issue I. Ready Data Availability J. Any other challenge (Specify):
3. Data Quality and Readiness		
3 a.	Does the M/D follow any Data and Metadata standards for the implementation process?	<ul style="list-style-type: none"> A. Data and Metadata Standards are fully implemented B. M/D has data and metadata standard but is yet to be implemented C. M/D is in the process of establishing data and metadata standards D. No standards are planned for data and metadata
3 b.	Does the M/D have a framework/guideline for interoperability? (Interoperability means the ways in which the data is formatted that allows diverse datasets to be merged or aggregated in meaningful ways)	<ul style="list-style-type: none"> A. M/D has an implemented framework for interoperability B. M/D has a framework for interoperability but is yet to be implemented C. M/D is trying to create/adopt a framework for interoperability D. No framework for interoperability is planned
3 c.	Does the M/D have a Quality Assurance staff in place?	<ul style="list-style-type: none"> A. A dedicated Quality Assurance (QA) staff is available B. Efforts are underway to deploy a dedicated QA staff C. No dedicated staff for QA framework but staff doing other quality checks are also engaged in this work D. No plan to have a dedicated QA staff
4. Policy Framework		
4 a.	Does the M/D have any governance/legal/data ethics & Privacy framework available?	<ul style="list-style-type: none"> A. Governance/legal/ data ethics & Privacy framework available by law B. Governance/legal/ data ethics & Privacy framework available by policy

Q.No.	Question	Scoring Mechanism
		C. Plans to establish a Governance/legal/ data ethics & Privacy framework, but not yet implemented D. No Governance/legal/ data ethics & Privacy framework planned
5. IT Infrastructure		
5 a.	Is required infrastructure available (such as storage, compute, etc.)?	A. Yes, required infrastructure is available B. Process of setting up required infrastructure is underway C. Plan to set up required infrastructure exists but not yet implemented D. No infrastructure planned
5 b.	Is there a secured cloud infrastructure?	A. A secured cloud infrastructure is fully implemented B. The process of setting up secured cloud infrastructure is underway C. Plan to establish secured cloud infrastructure exists but has not yet implemented D. No cloud infrastructure planned
5 c.	Does the M/D have a risk mitigation plan for addressing data breaches, system failures, and compliance risks?	A. There is a well-defined risk mitigation plan in place. B. A process for setting up a risk mitigation plan is underway. C. There is a plan to devise a risk mitigation plan in the future. D. No, there is no plan to have a risk mitigation plan.
6. Human Resources		
6 a.	Are there employees at M/D actively applying data science/big data technique in ongoing projects?	A. Employees are actively applying techniques B. Hiring process for relevant employees is in progress C. Hiring process being set up but not yet active D. No hiring process for these roles
6 b.	If a project requires data processing, are the skill requirements clearly stated in the context of relevant projects of roles where data processing is needed?	A. Yes, skill requirements clearly stated B. Process of setting up skill requirements is underway C. Plan to establish skill requirements exists but M/D has not yet started the process D. No skill requirements planned

Assessment Level

Level (Score)	Definition
Pre-Foundation (0-0.5)	The organization is just beginning to explore big data strategies and projects.
Foundation ($\geq 0.5-1.5$)	Big data leadership and strategies are being developed, with a few pilot projects in progress.
Practitioner ($\geq 1.5-2.5$)	Frameworks are established, data scientists are in place, and big data projects are strategically managed.
Expert (≥ 2.5 to 3)	Data science is deeply embedded, with skilled staff leading projects, offering internal and external training.