NITI Aayog

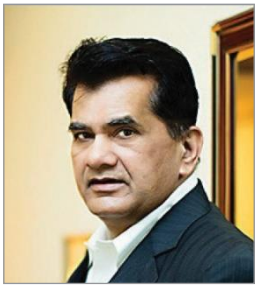
DEVELOPMENT MONITORING AND EVALUATION OFFICE

DATA GOVERNANCE QUALITY INDEX (DGQI)

Methodology Toolkit

March 2021

Amitabh Kant
Chief Executive Officer
National Institution for Transforming India
Government of India
New Delhi, India



MESSAGE

Clearly defined data strategies and strong data systems are fundamental for evidence generation and data-driven governance. The Data Governance Quality Index (DGQI) toolkit provides a unique framework for self-assessment of data preparedness levels across the Government of India. DGQI is based on internationally accepted data preparedness assessment models from private and public sectors but appropriately contextualized for India. The self-assessment approach of DGQI allows for internal review of data preparedness by a government agency on one hand and provides a framework to undertake a comparative assessment of data systems across multiple departments on the other hand.

The toolkit can be used to draw findings to drive reform and future policy initiatives to bolster data-driven governance. To enable data-driven outcomes and public policy decisions, a culture of evidence-based decision making needs to be promoted. An integrated data policy covering the three pillars of data preparedness (data strategy, data systems and data driven outcomes) needs to be adopted by all government agencies. Effective data strategies must be framed by the Ministries and Departments and States and implemented with adequate infrastructural, human and financial resources

This toolkit is an important step in the direction to institutionalize the culture of evidence-based policymaking in India by facilitating continuous and comprehensive ongoing assessment of the Government's data preparedness levels and driving relevant course corrections. For this culture to be truly fostered, it requires continuous emphasis and focus on better data management and analytics practices across the all tiers of the government for India to emerge as a leader in data-driven governance. Findings from DGQI on the key areas of improvements should be diligently followed-up leading to progress towards frontiers in all dimensions in a time bound manner.

MARCH, 2021
NEW DELHI
INDIA

Amitabh Kant

Preface

The Central Government of India, through its Ministries and Departments spends an amount to the tune of more than Rs. 10 lakh Crores on different Central Sector (CS) and Centrally Sponsored Schemes (CSS). These schemes vary in size, scope, objectives and overall implementation. In order to monitor the performance of a scheme, data on its progress in terms of deliverables and outcomes is collected in digitized or non-digitized way at different points of the scheme implementation. Increasingly, as public policy becomes evidence-informed, and the world adopts the Fourth Industrial Revolution technologies like Artificial Intelligence and Big Data in governance, the Ministries and Departments are adopting several measures to make the scheme monitoring more robust by improving their internal data systems.

Against this backdrop, the Development Monitoring and Evaluation Office (DMEO), NITI Aayog, has prepared a toolkit called Data Governance Quality Index (DGQI) to enable the government agencies, at central and state levels to undertake a detailed self-assessment of their data preparedness levels for their programmatic interventions and accord objective scores to them. This toolkit is designed to be executed in self-administered format, which offers two key advantages of increased internal deliberations within and improved ownership of the participating agencies.

For developing this toolkit, an in-depth literature review of various global and domestic data preparedness models was undertaken. Subsequently, three pillars of data preparedness were identified, viz. (a)Data Strategy to lay down systemic guidelines, (b) Data Systems to ensure smooth processes of data generation, management and its use, and (c) Data-driven Outcomes where data is utilized and widely shared across institutions by multi-disciplinary teams to drive policymaking. This edition of the Data Governance Quality Index (DGQI) toolkit covers the second pillar, data systems, with an objective of facilitating an ‘as-is’ assessment of the schemes’ or programmatic management information systems (MIS) across government agencies. Once conducted, the findings thereof may be used by the concerned agencies to prepare a Data Strategy leading to better data-driven outcomes. The subsequent editions of the DGQI toolkit aim to capture the additional two pillars as well.

This toolkit and its findings can lay the foundation for an integrated government-wide Data Policy while enabling the DGQI commissioning agency and the participating agencies to have a structured dialogue for identifying specific areas of improvements and designing customized pathways for expeditiously achieving advanced data preparedness levels.

It is hoped that this DGQI approach and methodology toolkit will aid the policy makers at the highest level across the Government to self-assess their data systems and take affirmative actions to improve data preparedness.

Acknowledgement

This DGQI toolkit is a first-of-its-kind and has been developed as a result of arduous efforts of DMEO, NITI Aayog along with a wide range of stakeholders across the Government of India. First of all, I would like to express my deep gratitude to Sh. Prashant Mittal, MD, NICSI and Sh. Rajiv Rathi, DG, NIC who have provided valuable inputs to develop this self-assessment tool. I would also like to thank NIC HoDs/HOGs and officers from scheme divisions of various Ministries/Departments of Government of India for their guidance during multiple expert consultations held with them. The toolkit has also remarkably benefitted from the inputs received from Mr. Ashutosh Jain, Deputy Director General, Development Monitoring & Evaluation Office, NITI Aayog and all other colleagues at DMEO.

Last but not the least, I would like to thank the team - Mr. Anand Trivedi, Mr. Krishn Kant Sharma, Ms. Gunjan Saini, Ms. Vatsala Aggarwal, Mr. Ankit Choudhary, and Mr. Kapil Saini for their outstanding efforts, diligence and dedication towards producing this toolkit on Data Governance Quality Index which would immensely help the cause of enhanced data preparedness across the Government of India.

Director General,
Development Monitoring & Evaluation Office,
NITI Aayog
March 2021

Contents

1. Introduction

1.1. Data Preparedness in India-Historical Perspective

Data collection and warehousing started as early as 1881 when the first Census was conducted in India. After Independence, National Sample Survey Organization was established in 1950 and Central Statistical Organization in 1951. Data collected through large scale surveys by these organisations, and the administrative data collected by Ministries and the state Governments led to data-driven decision-making in the Central and the State Governments. Scheme-level information generated and collated at various levels i.e., village, block, district and state levels, assisted programme implementation. However, the whole exercise was done manually on formats individually developed under each scheme and overall scheme progress was mostly tracked inputs (fund releases and budget utilization). MIS systems and digital data storage facilities became all pervasive in the last two decades. Gradually, activities and outputs started to get monitored.

With digitization of data, advent of new techniques and ever-increasing importance of data in public policy, the need for even better management of data was recognized. In order to further India's vision towards Open Government and Open Data initiative, National Data Sharing & Accessibility Policy was adopted and data.gov.in was launched to provide all relevant data from Government at single place for wider public use. Many schemes also migrated to dashboard based and basic analytics-driven systems which make complex information available to decision makers in simple charts and figures. Intra-government exchange and integration of data is now being facilitated using ICT platforms such as DISHA, Prayas and Output- Outcome Monitoring Framework (OOMF).

1.2. Data Preparedness in India – Current Scenario

As of now, an internal Management Information Systems (MIS) is developed for most government programmes, which provides required information regarding coverage and outputs of the programme, e.g., HMIS for National Health Mission which tracks information uploaded by the States/UTs which enables planning, management, and decision-making based on grading of facilities and various health indicators at block, district, state as well as national level. Such programme MIS typically have capabilities to generate standardized analytical reports on the basis of data collected. Further, Ministry of Statistics and Programme Implementation (MoSPI), through Twenty Point Programme (TPP-2006) and Infrastructure and Project Monitoring Division (IPMD) monitors key infrastructure projects within the Government. TPP monitors 65 items on 162 parameters related to different programmes and schemes spread across different nodal Ministries and Departments (M/Ds[1]). The Government also launched Digital India programme in 2015 to ensure digital availability of government services to citizens. This Programme is being managed by National e-Governance Division (NeGD). NeGD provides project development and programme management support to e-governance related measures taken by Ministries. Some of the State Governments also present the work done by their various departments through dashboard based analytical systems (e.g. Pratibimba by Govt. of Karnataka). These measures have ushered in a new era of accountability. Overall, it is clear from the

[1]Throughout this document, M/D has been referred as an abbreviation for Ministries/Departments

background above that governments in India have been quite proactive in ensuring adoption of newer technologies in data management and thereby improving programme outputs and outcomes.

However, there still remains lot more to be done with reference to data preparedness level in the Government of India, especially with respect to programme monitoring and management. Given the above, it is imperative that a comprehensive review of data preparedness is conducted for government data systems for scheme management and decision support information systems. Development Monitoring and Evaluation Office (DMEO), an attached office of NITI Aayog, has developed this toolkit to enable a comprehensive self-assessment of data preparedness levels to come up with a scorecard of Data Governance Quality Index (DGQI) for the government agencies at the central and state level.

1.3. Context to DGQI

DMEO's rich experience with various Ministries/Departments of the Government of India on account of exercises such as the output-outcome monitoring of CS/CSS schemes, evaluation of schemes and monitoring of global indices has highlighted that there was a glaring disparity in data preparedness levels of scheme monitoring systems within the Ministries/ Departments (M/Ds) across various CS/CSS schemes.

Subsequently, the need for developing a toolkit to facilitate a comprehensive assessment of data preparedness of all the M/Ds emerged. DMEO developed this toolkit by referencing suitable documents in national and international context applicable to government agencies in specific and broader organizational data maturity in general with an aim of scoring and comparatively ranking data systems' preparedness of government agencies. .

1.4. Intent of the document

This toolkit has been developed to enable the Central and the State governments assess data preparedness levels of their own schemes and identify areas where better IT systems are needed for more effective programme implementation and monitoring. It can also help a DGQI commissioning agency such as state planning departments to build a comparative assessment of the data systems' maturity across its various departments or agencies while identifying the theme-wise areas for improvement within the government. In this case, the commissioning agency may also use the tool to source and encapsulate best practices from the ministries/departments and disseminate them within the government for enabling cross-learning. However, the subject of the toolkit is not very specific to Central and State M/Ds, and in fact, can be adopted by other government stakeholders such as statutory bodies, PSUs etc. There may be 2 direct benefits of adoption of this toolkit by government stakeholders. First, it would highlight the adopting body's data preparedness levels and the areas for improvement to help in better programme implementation and overall outcomes of the projects. Secondly, by improving data preparedness at different levels of government (central govt., state govt. etc.) it would have multiplier effects in improving the overall efficiency and accountability of governance. The overall intent of this document is to disseminate this framework and the questionnaire to all the relevant stakeholders in the form of a ready-to-use toolkit.

2. Objectives & Scope

2.1. Objectives

The overarching objective of this document is to disseminate this toolkit to all the relevant government stakeholders. As mentioned previously, the intent of the DGQI toolkit is to enable Ministries/ Departments and state departments to assess themselves at various levels of data maturity on the basis of a standardized framework, which in turn would facilitate deepening of digitization in the Government of India. .

It is hoped that in the long run, DGQI will help in laying the foundation of more integrated monitoring systems, for e.g., a single, online, API-integrable 'Overarching Dashboard' kind of monitoring system of all the CS/ CSS schemes of all M/Ds, ultimately leading to a state-of-the-art data-driven decision making.

The DGQI toolkit has been developed with the following objectives:

a. To enable review and assessment of data preparedness of the data/ MIS systems of the Ministries/Departments on objective parameters of a standardized framework.
b. To prepare a self-assessment diagnostic tool that will enable the M/Ds to internally contemplate the need for improving data systems.
c. To enable the commissioning agencies to conduct a comparative assessment of data preparedness and source best practices in IT systems which can enable improved cross-learning between the participating agencies.

This document presents in detail the approach and methodology which may be pursued by an adopting government agency. Further, a suggestive operational approach has also been discussed.

2.2. Scope:

The document has primarily been developed for Central Government Ministries and Departments. The DGQI exercise can be undertaken by commissioning agencies to ensure participation either from all the ministries/departments or a selected group of ministries/departments based on their scale and scope of work. Even within the participating ministries/departments, the commissioning agency may decide to either include all the schemes of each of the M/Ds or a selected set of schemes based on the budget and nature of the schemes. Typically, it is prescribed that a pareto analysis should be conducted for the identifying the most critical schemes.

However, the scope and applicability of the Index is much wider and deeper across the government machinery. The state governments may adopt the document with few tweaks to assess their Ministries and Departments on current levels of data preparedness. Further, even at district level, a customized shortened tool adapted from the DGQI questionnaire may be used to assess the data preparedness of different departments in the district. Improvement in data systems could be very useful in making departments and district administrations more responsive in-service delivery and in providing useful data to collate at the state-and district-level to draw insights. The Index could also be used to assess data preparedness levels in same Department across states (for eg. Department of Health and Family Welfare across states). This may be useful in planning schemes at central level and allocating resources for project monitoring in the states. Moving ahead, other government set-ups such as PSUs, autonomous bodies etc. may also adopt the exercise in its spirit.

Reforms undertaken to improve the standardization and maturity of data systems as a consequence of the internal DGQI assessment can be used by the Central and State governments to develop integrated government-wide shared data systems, e.g. the CM's dashboard, sector-level dashboards and departmental outcomes-monitoring systems which involve inter-agency data exchange and coordination. The tools may have to be marginally edited to take into account the different structure, and data requirements of these organizations while keeping the broader principles intact.

3. Approach & Methodology

3.1. Key Reference Data Maturity Assessment Models

To begin with, DMEO reviewed existing frameworks for assessing data preparedness of organizations. Nine data maturity models were identified spanning private and public sector organizations in both Indian and international contexts.

Based on an initial evaluation, following four models were shortlisted for a detailed study to understand their key focus areas and methodology. They were chosen based on their relevance, exhaustiveness and representativeness (coverage across public and private sectors and in India as well as internationally).

1. **US Federal Government Data Maturity Model**: This model was developed and integrated as a part of the US Federal Data Strategy, the 2020 action plan which was laid down in March 2018, by the US President's Management Agenda. This Agenda included a new Cross-Agency Priority (CAP) Goal: Leveraging Data as a Strategic Asset. The model provides a common language and framework to help promulgate common solutions and best practices across federal agencies towards advancing data-driven decision making.

2. **Data Governance Maturity Model (IBM):** The Data Governance Maturity Model is developed by IBM Data Governance Solutions. It provides an informed, objective, documented assessment of "current state" of data governance maturity. It also works on defining the strengths and weakness in existing data governance organization, technologies, processes and activities.

3. **Data Maturity Assessment Framework (SCM):** The Data Maturity Assessment Framework is published by Ministry of Housing and Urban Affairs. The framework is prepared to promote a spirit of competitive benchmarking amongst 100 Smart Cities that will enable them to assess themselves at varying degrees of data maturity through its twin pillars of 'Systemic' and 'Sectoral' maturity with respect to a standardized framework covering aspects of enabling policies, governance structures, data processes and capacities.

4. **Data Maturity Management Model (CMMI):** The Data Management Maturity (DMM) Model is developed by CMMI Institute. It provides the best practices in six different categories to help organizations build, improve, and measure their enterprise data management capability allowing for timely, accurate and accessible data across the entire organization.

   The key thematic areas within these four key reference models are indicated below:

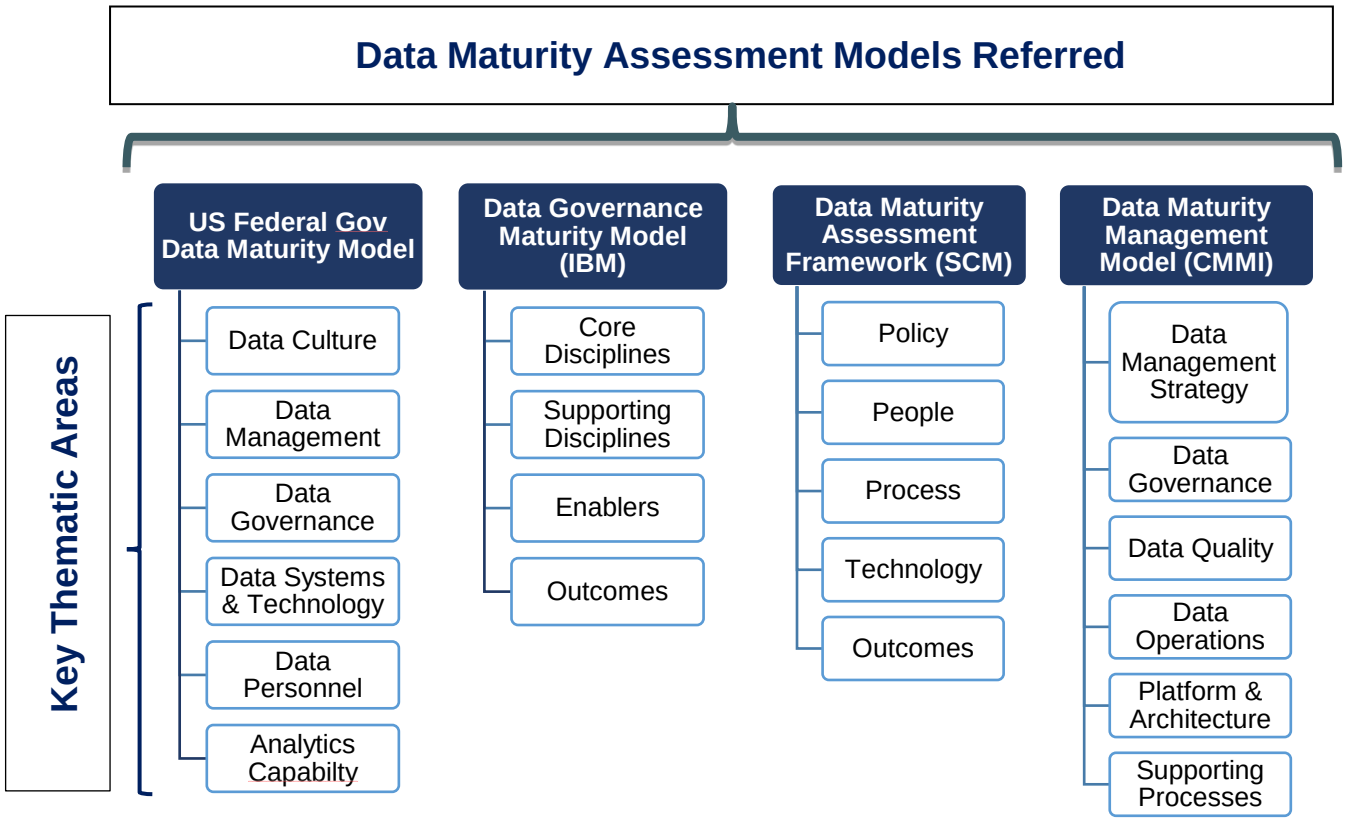


Figure 1: Key Reference Models with thematic areas

3.2. Overall Approach:

Based on a detailed analysis of the above-mentioned data maturity assessment models, a theory of change for data preparedness was developed. Three key pillars of data preparedness were identified viz., Data Strategy, Data Systems and Data Outcomes. This theory of change formed the basis for design of DGQI as discussed in detail subsequently.

Figure 2: Theory of Change for Data Maturity

The thematic areas in identified key reference models were mapped with the theory of change. This formed the basis of identifying key thematic areas to be covered in the DGQI toolkit. The mapping was as under:

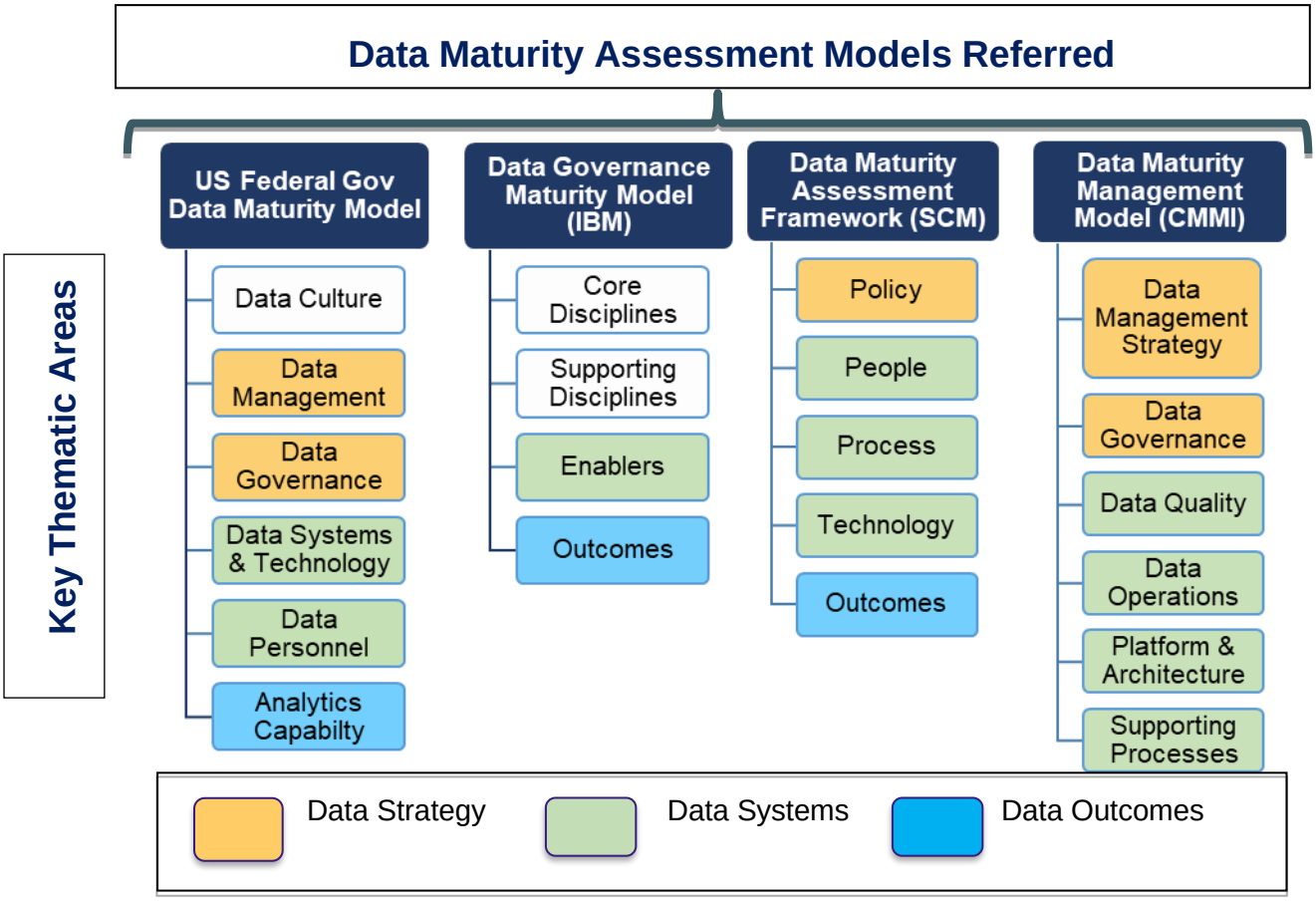


Figure 3: Key Reference Models mapped with Theory of Change for Data Maturity

First of all, data strategy is required to lay down systemic guidelines for data governance by organisations.

Next, there is a role for well-defined and organised data systems encompassing various data processes such as data generation, ensuring data quality, use of technology, data analysis to create evidence, dissemination of evidence in user-friendly manner and existence of capable data management teams. Data systems are to be supported by enablers such as adequate financial allocation, correct placement of data management teams to ensure coordination with decision makers and configuration management to take care of other technical support.

The first and the second pillar work in conjunction with each other to enable the third pillar of data-driven outcomes. However, the existence of data strategies and systems alone cannot ensure that data is converted to information and is actually utilised as evidence to guide decisions. The same has to be fostered within institutions through a step-by-step approach. This would involve integrated data use facilitated by exchange of data among various agencies, development of strong data analytical capabilities within Ministries/Departments and finally well-articulated data use plans. These aspects hence get covered under the third pillar – data-driven outcomes.

The focus of this DGQI toolkit is on the second pillar – data systems. While data strategy is an extremely useful precedent for well-defined data systems and data driven outcomes are the ultimate goals to be reached, the same have not been covered in this first edition of the

DGQI toolkit. The key reason was to formulate a toolkit to first assess data systems, and identify challenges and best practices therein. The learning by using this toolkit then may be used to focus on rectifying this pillar. The inputs from this assessment then may be used by the Central and the State Ministries/Departments in building a strategy going forward. It is planned that subsequently the scope of the toolkit may be broadened. The upcoming versions of toolkit would assist Ministries/Departments and other government agencies undertake a phased approach for strengthening their data processes leading to better data outcomes. However, for States where articulate steps to institutionalize an integrated data strategy have already been undertaken, aspects within the rest of the two pillars, i.e., data strategy and data outcomes, may also be included in the first round of self-assessment by appropriately modifying this toolkit.

3.3. Index Methodology

Under the realm of the overall approach, six key themes have been identified under data systems pillar covered by the Data Governance Quality Index:

Data Generation: Data generation measures the ability of the respective ministries/departments to efficiently generate useful data in the course of their programme implementation. It covers areas related to the level of digitization, frequency and granularity of data generation. It also assesses if mobile phones, location tracking and GIS mapping is used to authenticate the generated data.

Data Quality: Data Quality covers processes of scientifically and statistically evaluating data in order to determine whether they meet quality benchmarks. The key areas covered under this theme relate to profiling of data, data quality assessment processes (for e.g. data pipeline design, well defined data schema etc.), data cleaning, use of latest technologies and mobile phones in the process.

Use of Technology: This theme assesses if emerging technologies are being utilized to improve data robustness. It assesses if MIS of ministries/departments have linkages with PFMS for ensuring transparency and Jan-Dhan, Aadhar and Mobile [JAM-trinity (if applicable)] for delivering last mile services. It also explored if other data sources such as remote sensing or social media data is utilized in addition to data collected by ministries/departments to get a nuanced understanding. Finally, it also measures if emerging technologies like block chain, big data analytics, machine learning, artificial intelligence, IoT are being used to collect data or to draw analytical insights from it.

Data Analysis, Use and Dissemination: One of the core themes, it covers if the collected data is being analyzed and used for evidence creation and decision making. Given the present context, it gauges whether ministries/departments are undertaking basic cross-sectional analyses only or regression and predictive analysis as well. The use of dashboards for visualization of data is also checked to ensure that information is disseminated in a user-friendly manner. It also assesses if other social media platforms are also being increasingly used for information dissemination and whether websites have features to support multi-lingual interfaces and are GIGW compliant.

Data Security and HR Capacity: While data security requires an in-depth analysis in itself, the same is briefly captured in the index also to reflect its importance. It assesses if antivirus updates and internal audit systems are in place to ensure data is not corrupted or prone to threats. These were identified to be the minimum requirements expected to be met and are not meant to be exhaustive in nature. To look at HR capacity, the existence of dedicated

data quality teams has been considered. Again, this is by no means an exhaustive measure of capacity development but was adopted as the starting point.

Case Studies: The present questionnaire for this theme focuses on scheme-level MIS. Any intervention done at the Ministry/ Department level or any innovative approach that may not be captured in the structured questions of the tool can be highlighted through best practices. These best practices can be provided as case studies. This theme is expected to help unlock the hidden potential not only in terms of enhanced decision making through inter-ministerial collaboration but also by opening doors for learning from challenges faced and the solutions devised by peer ministries.

A snapshot of weightages allocated to various themes under DGQI has been provided below. The weights have been assigned on the basis of their relative importance in driving data-driven outcomes while building in the learning from the key reference models and their methodologies. The local context of India was also kept in mind while assigning these weightages. Hence, maximum weight has been given to data analysis, use and dissemination (30%) followed by data generation (20%), data quality (15%) and best practices (15%). While digitized collection is indispensable for next steps, data analytics is of utmost importance to convert data to evidence. Sufficient focus on data quality control and utilization of best practices come next to ensure good quality.

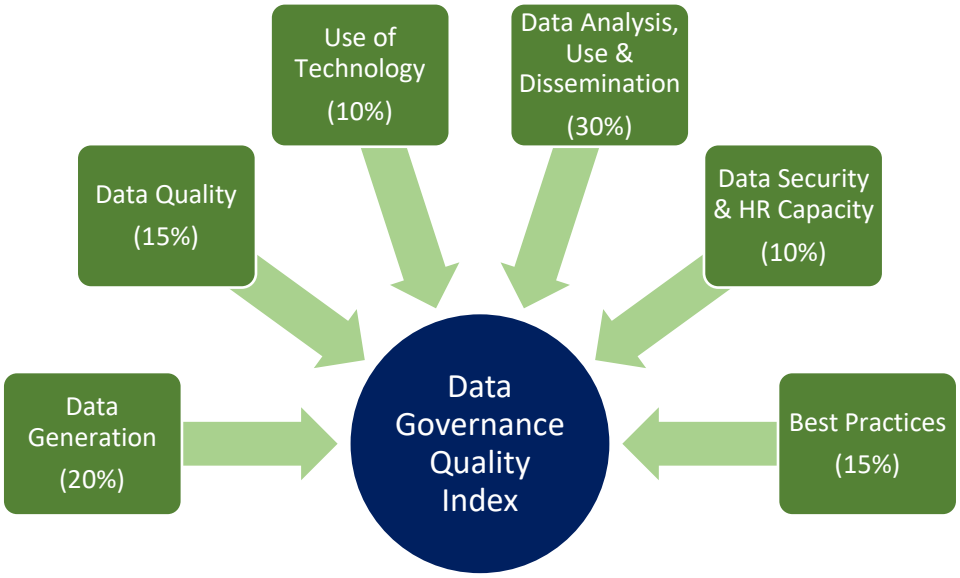


Figure 4 DGQI: Themes & Weightages

Within each theme, key dimensions and questions to assess the preparedness of M/Ds on these dimensions were developed. The survey questionnaire designed as a self-assessment tool can be found at **Annexure 1**. Subsequently, weightages were assigned to each question/dimension within every theme as shown below in Table 1.

Theme	Theme Weightage	Question No. (In Survey)	Dimension/Question	Question Sub-weightage within theme
Data Generation	**20%**	Part B, Q A	Data Generation: Granularity and Digitization	40%
		Part B, Q B	Data Generation: Frequency in the scheme	40%
		Part B, Q C 1	Data Generation: Use of mobile surveys, geocoding or geofencing	10%
		Part B, Q C2	Data Generation: Use of location tracking devices	10%
Data Quality	**15%**	Part B, Q D (a)	Data Profiling	20%
		Part B, Q D (c,e)	Data Quality Assessment	25%
		Part B, Q D (b,d)	Data Cleansing	25%
		Part B, Q E	Use of Mobile Phones in QC	30%
Use of Technology	**10%**	Part B, Q J	Linkages with other Platforms	60%
		Part B, Q K	Use of Alternative Data Sources	20%
		Part B, Q L	Use of ML/AI, Blockchain, IoT, Big Data	20%
Data Analysis, Use & Dissemination	**30%**	Part B, Q F	Data Analysis	30%
		Part B, Q G	Dashboards	20%
		Part B, Q H	Data Visualization: Types of visualization	15%
		Part B, Q H	Data Visualization: Visualization on maps	15%
		Part B, Q I	Dissemination/Communication	10%
		Part A, Q B1	Portal features for differently abled	5%
		Part A, Q B2	Multi-lingual interface	5%
Data Security & HR Capacity	**10%**	Part A, Q B3	Antivirus updates	35%
		Part A, Q B6	Transaction systems with internal audit systems	30%
		Part B, Q D (f)	Existence of Data QC teams	35%
Case Studies	**15%**	Part A Q C	Best Practices	100%

Table 1 DGQI: Question wise weightages

After defining weightages and sub-weightages, a scoring mechanism for each question was defined to attribute a score between 0 (lowest) to 5 (highest). The same is available at **Annexure 2**. Based on these weights and scoring mechanism, DGQI score ranging between 0 to 5 may be calculated for every CS/CSS scheme. To arrive at the scores for each Ministry/Department, a simple average of scheme-level DGQI scores can be undertaken. Hence, by using the toolkit, every participating ministry/department will be able to see its performance in the form of a DGQI score between 0 to 5, 0 being the lowest and 5 being the highest.

Owing to different functions and scope of M/Ds, a straightforward cross-comparison of M/Ds may be unsuitable and may not yield relevant findings. Participating M/Ds, therefore, may

be divided into the following six categories: Administrative, Strategic, Infrastructure, Social, Economic and Scientific. These categories are suggestive only and appropriate modifications may be made as per the context of the commissioning and participating agencies.

Another point of consideration while developing the toolkit was to validate the applicability of each question for each category of M/D. For instance, visualization on maps may not be applicable for some M/Ds with no spatially spread out schemes, etc. In order to take care of this, it is suggested that responses received on each question are carefully studied. If any question is not answered or response to the question is negative for all M/Ds in a particular category, the question may be considered to be “Not Applicable” for that category by the commissioning agency. In such cases, average score for that question can be awarded to all schemes of M/Ds in such category. For e.g. All M/Ds in Economic category where funds are disbursed to some other government entity, the Ministry may not be having data with themselves regarding usage of funds. In such case, data granularity and digitization related questions for these M/Ds may be awarded average score of other schemes. The rationale here is that M/Ds within a category can be considered peers and learn from each other if any other M/D in the same category is using a better technique. In case a government agency is using this toolkit without any peers filling up the same, a separate approach may be taken. In such cases, it is suggested that if any question/ section is not applicable to the adopting agency, the question/ section may be removed from the overall scoring process. While doing so, appropriate redistribution of weightage, preferably proportionate weightage distribution, across questions should be done to maintain the basic structure of the scorecard.

Additionally, applicability of linkage with JAM trinity needs to be verified for every scheme of each M/D. Only if the scheme is beneficiary-oriented or if the M/D has provided a positive response to any one of the questions on JAM trinity (Part B Q J (2), (3), (4)), the question should be considered applicable to the scheme. Operational Approach

The following operational approach may be adopted to carry out the self-assessment of data preparedness exercise. The toolkit (available at Annexure 1) may be designed and launched in the form of an online survey with assistance from NIC. Login credentials, for filling up online survey form, should be created for each participating agency and may also be shared with the NIC divisions of the those respective agencies.. Further, JS/ Director level nodal officers ought to be nominated from each participating agency or Ministry/Department who can assist in coordinating across scheme divisions and in driving the entire exercise at the participating agency level. To facilitate the Ministries/ Departments in understanding the task at hand, many rounds of workshops/ webinars may need to be conducted by the commissioning agency till there is enough clarity to fill out the tool. The commissioning agency may also need to conduct regular follow-ups with the participating agencies, both telephonically and through emails, to prompt them for timely filling up of the survey and also extend support at all stages. Finally, after receiving all the requisite data from all the participating agencies, a comprehensive data analysis may be done and DGQI scores may be finalized.

3.4. How to use this methodology

Different government bodies/ agencies may find the exercise to be useful to them. As mentioned earlier, although this toolkit to self-assess data preparedness has been

developed primarily for Central and State M/Ds, its subject is all pervasive and could be equally useful in the context of other government agencies. Following steps may be used while adopting this methodology.

a) Understanding the methodology of DGQI and assessing the present data preparedness systems against the index
b) Deciding on the number of pillars to be included in the self-assessment exercise though it is advised that initial assessment be based on the data systems pillar
c) Adapting the questionnaire in the context of the participating government agency
    i. Modify questions' language/ options as required
    ii. Removing non-applicable questions/ sections
    iii. Some other question(s) may be added which seem relevant for the agency and fall within the overall framework
d) Getting the questionnaire developed in a software version which could be canvassed to all relevant stakeholders
e) Ensuring participation from all the relevant stakeholders (may involve top level officials to ensure participation from all)
f) Modifying scoring method wherever it seems necessary
g) Adjusting weights as per the context and roles of the agency
h) Generating theme-wise scores and aggregate agency-wise scores
i) Discussing actionables against the shortcomings identified based on the agency-wise and theme-wise scores.

3.5. Points to be taken care of while using the tool and methodology

Several documents, including both national and international frameworks have been used to come up with the methodology in this toolkit. The research tool, given in **Annexure 1** has been developed based on the learning from these frameworks. Though the subject matter is non-specific, the tool has been indeed customized and fine-tuned to assess central government ministries and departments. Further, since this is a first of its toolkit in Indian government context, there may be certain aspects which may have been left out in the process of contextualization and can be further improved by the user agency as per their context. The following Do's and Don'ts list may be used by any other government body which wants to use the tool and methodology-

Do's	Don'ts
• The questionnaire should be edited on the basis of the context of the government body. Themes, question therein and weights may be fine-tuned to better represent the data related policy framework of the governments undertaking the exercise.	• The participating bodies (divisions, departments, ministries etc.) may not be explained the details of the methodology before seeking their responses. This will help reduce any effort to engineer the responses to gain better score.
• The weights and scoring criteria of different questions should be pre-defined and frozen before taking in the responses.	• Participating bodies should not be allowed to get themselves excluded from the exercise on the basis of flimsy grounds where they might say

	that the exercise / part of exercise is not applicable to them. Any such request from any participating body should be thoroughly examined before exempting them from the exercise.
• Several rounds of orientation and awareness generation sessions regarding the intent of the exercise should be conducted. This will help participating bodies see the merit in the exercise and provide quicker responses.	• The participating bodies should be made clear in advance that the responses would not be allowed to be changed once the index calculation is done. They should be encouraged to seek help from the organizing body in case there is any gap in understanding any part/ section of questionnaire so as to help them provide correct response the first time itself.
• Dedicated persons should be there at the organizing agency to respond to any query from participating bodies. Similarly, the participating bodies should be asked to nominate nodal officer/ SPOC to establish an effective communication. Further, a dynamic spreadsheet should be maintained by the organizing agency to document all the communication between them and the participating bodies.	• In case any participating body wishes to change its responses after index score calculation, a detailed response should be taken from them which cites the reasons for the same. The updated questions should be filled in the data-set by the organizing team and at no-point the response filling window should be reopened for blanket changes in responses.
• The questionnaire should be canvassed online by providing a login-id and password to nodal officers of participating bodies. It will save precious time in data entry and cleaning.	
• The agency which steers the exercise should deploy sufficient number of people to help the participating bodies understand the online portal/ questions. Since the nature of the exercise is fairly technical, this step may help in getting correct and unbiased responses from the participating bodies.	
• As per the need of the government agency, it should be decided whether a scoring on the index is sufficient or a ranking of participating bodies is also required. Many a times, the ranking on the basis of scores may trigger unnecessary comparisons between the	

participating bodies and may jeopardize the larger objective.	
• The participating bodies should be categorized on the basis of their functional/ domain similarities. This will help reduce any unfair comparison among dissimilar bodies.	

3.6. Limitations of the toolkit:

The current version of the toolkit only focuses on the assessment of the MIS/ Dashboards of the CS/ CSS schemes of the Ministries/ Departments. Whereas, the Ministries/ Departments may also have several other dashboards encompassing other areas of data monitoring like some M/Ds may have dashboards put in place for project monitoring or take into account a sectoral overview, which do not fall under current scope of the toolkit. However, flexibility has been offered to M/Ds to highlight these in best practices section of the questionnaire.

The present focus is only on one pillar of data preparedness, i.e. data systems; the reason for which have been discussed above.

3.7. How to reach us:

DMEO, NITI Aayog would be happy to help any government or non-government entities interested to adopt and implement the DGQI toolkit. For detailed understanding of the toolkit and further support, the concerned DMEO team which designed the DGQI toolkit can be reached at *dgqi-dmeo-niti@gov.in*.

Annexure-1: DGQI Questionnaire

Part -A (To be fed at Ministry/ Department Level) **Date : / /2021**

A. Background Information

1.	Ministry:		(Short name)
2.	Department:		
3.	Name of the Central Sector Schemes:	a.	
		b.	
		c.	
4.	Name of Centrally Sponsored Schemes:	a.	
		b.	
		c.	

B. General

1.	Does the website of M/Dhave features to support differently abled as per GIGW/NIC norms	☐ Yes
2.	Do all the M/D portals support Multi-lingual interfacesas per GIGW/NIC norms	☐ Yes
3.	Does the M/D follow Antivirus update	☐ Yes
4.	Does the M/D follow norms of electronic waste disposal	☐ Yes
5.	Does the M/D use de-gauging before electronic machine disposal	☐ Yes
6.	Are transaction systems put in place are conducive to internal audit	☐ Yes

C. Best Practices – Kindly describe up to 3 best practices of using Information Technology & Data Analytics in driving smart, near real-time and granular decisions in your ministry/department

Best Practice 1
1a. Describe the best practice in terms of the objectives of the initiative, the technological solution developed and the implementation of the initiative (in about 200 words)
1b. Explain the positive impact generated due to the technological solution implemented (in about 50 words)

Best Practice 2
2a. Describe the best practice in terms of the objectives of the initiative, the technological solution developed and the implementation of the initiative (in about 200 words)
2b. Explain the positive impact generated due to the technological solution implemented (in about 50 words)

Best Practice 3
3a. Describe the best practice in terms of the objectives of the initiative, the technological solution developed and the implementation of the initiative (in about 200 words)
3b. Explain the positive impact generated due to the technological solution implemented (in about 50 words)

Part -B (To be fed at Program/ Scheme level) – As many Part-B to be filled as number of schemes mentioned in Point no. 3&4 of A. Background information of Part – A of self assessment form

Scheme Name : **Short Name(If any):**

A .Data Generation: Granularity and Digitization

1.	**At what granularity is data generated and at what level the data is digitized for this scheme**	**Paper**	**Digital**	

	a. At the M/D (National)	☐ Yes	☐ Yes	
	b. State	☐ Yes	☐ Yes	☐N/A
	c. District	☐ Yes	☐ Yes	☐N/A
	d. Panchayat	☐ Yes	☐ Yes	☐N/A
	e. Village	☐ Yes	☐ Yes	☐N/A
	f. Individual	☐ Yes	☐ Yes	☐N/A
	g. Project	☐ Yes	☐ Yes	☐N/A

N/A= not applicable; M/D = Ministry/Department If the attributes are applicable for Digital (as Yes)

B. Data Generation: Frequency in the scheme

1.	At what frequency is data generated	Paper	Digital
	a. Realtime (transaction data)	☐ Yes	☐ Yes
	b. Daily	☐ Yes	☐ Yes
	c. Weekly/ Fortnightly	☐ Yes	☐ Yes
	d. Monthly	☐ Yes	☐ Yes
	e. Quarterly	☐ Yes	☐ Yes
	f. Half-yearly	☐ Yes	☐ Yes
	g. Yearly	☐ Yes	☐ Yes

C. Data Generation: Others

1.	Does scheme use any of the following for data generation	
	a. Mobile surveys - CAPI	☐ Yes
	b. Geo coded photos	☐ Yes
	c. Geo-fenced information	☐ Yes
2.	Use location tracking devices	☐ Yes

D. Data Quality

1.	Does scheme use protocols to check data quality	☐ Yes
	a. Rigorous data profiling and control of incoming data	☐ Yes
	b. Data pipeline design to avoid duplicate data	☐ Yes
	c. Accurate gathering of data requirements (well-defined data schema)	☐ Yes
	d. Enforcement of data integrity	☐ Yes
	e. Integration of data lineage traceability into the data pipelines- use of metadata	☐ Yes
	f. Dedicated data quality control teams	☐ Yes

E. Use of mobile phones

1.	Does the scheme use mobile phones for monitoring	☐ Yes ☐ No
2.	If 'Yes', what are the purposes for which mobile phones are used	
	a. Outreach as a part of social audit or participatory monitoring	☐ Yes
	b. Feedback	☐ Yes

	c. Collect data remotely	☐ Yes
	d. Telephonic survey (manual/IVR)	☐ Yes
	e. Geotagged photograph uploading	☐ Yes
	f. Geo fenced data generation	☐ Yes
	g. Location and GPS data	☐ Yes
	h. Multimedia data – voice, video, images as evidence	☐ Yes

F. Data Analysis

1.	Does the scheme do data analysis	☐ Yes ☐ No
2.	If 'Yes', what are the methods used	
	a. Exploratory data analysis	☐ Yes
	b. Modeling and algorithms	☐ Yes
	c. Correlation	☐ Yes
	d. Causation	☐ Yes
	e. Regression analysis	☐ Yes
	f. Predictive	☐ Yes
	g. Data mining	☐ Yes

G. Dashboards

1.	Does the scheme have dashboards	☐ Yes ☐ No
2.	If 'Yes', what are the purposes for which Dashboards are being used	
	a. Visual presentation of performance measures	☐ Yes
	b. Identifying preempt trends	☐ Yes
	c. Measure efficiencies/inefficiencies	☐ Yes
	d. Generate detailed reports showing new trends	☐ Yes
	e. Make more informed decisions based on collected business intelligence	☐ Yes
	f. Align strategies and organizational goals	☐ Yes
	g. User friendly one stop access to multiple reports	☐ Yes
	h. Gain total visibility of all systems instantly	☐ Yes
	i. Quick identification of data outliers and correlations	☐ Yes

H. Data Visualization

1.	Type of Data Visualization used by the scheme	
	a. Bar chart	☐ Yes
	b. Histogram	☐ Yes
	c. Scatter plot	☐ Yes
	d. Heat maps	☐ Yes
	e. Treemaps	☐ Yes
	f. Gantt chart	☐ Yes

	g. Others- Stripe graphics, streamgraph, etc.	☐ Yes
2.	Does M/D visualize information on maps	☐ Yes

I. Dissemination/Communication

1.	Use of IT by scheme for dissemination	
	a. Web site information	☐ Yes
	b. Web site dashboard	☐ Yes
	c. Mobile App	☐ Yes
	d. Social media	☐ Yes

J. Linkages with other Platforms

1.	Does the MIS of the scheme have linkages with PFMS	☐ Yes
2.	Does the MIS of the scheme have linkages with Aadhar	☐ Yes
3.	Does the MIS of the scheme have linkages with Mobile numbers	☐ Yes
4.	Does the MIS of the scheme have linkages with Bank Account	☐ Yes

K. Use of Other Data Sources

1.	Does the MIS of the scheme use remote sensing data	☐ Yes
2.	Does the MIS of the scheme use night light data	☐ Yes
3.	Does the MIS of the scheme use social media data	☐ Yes
4.	Does the MIS of the scheme use private sector generated data	☐ Yes
5.	Is IT system compliant to Local Govt Directory(LGD)	☐ Yes

L. Use of Machine Learning, Artificial Intelligence, Block Chain and Internet of Things

1.	Does the scheme apply Machine Learning	☐ Yes
2.	Does the scheme apply Artificial Intelligence	☐ Yes
3.	Does the scheme apply Blockchain	☐ Yes
4.	Does the scheme use Internet of Things (IoT)	☐ Yes
5.	Does the scheme use Big Data analyticals	☐ Yes

M. Number of users/ stakeholders of MIS/Dashboard of the Scheme

1.	Number of users at Central Level	
2.	Number of users at State Level	
3.	Number of users at District Level	
4.	Number of users at Sub-District/taluka/tehsil/Block Level	
5.	Number of users at GP/Village Level	

N. Application Software

1.	Is it Open Source	☐ Yes
2.	Developed in-House	☐ Yes

3.	Developed by out-sourced agency	☐ Yes
4.	Does scheme owns the partial/full IPR	☐ Yes
5.	Could it be replicated for similar requirement	☐ Yes

Annexure-2: Scoring mechanism

Sr. No.	Question No.	Question	Scoring Logic
1	Part B, Q A	Data Generation: Granularity and Digitization	Scoring is to be done primarily for the Digital Part Only. If the Scheme is collecting Paper-based information at any granularity at State or Sub-state level, give '0'. Regarding Digital scoring, give '1' at National Level (only), '3' at State level (only), '4' at district level and '5' at individual/village/project level.
2	Part B, Q B	Data Generation: Frequency in the scheme	Scoring is to be done primarily for the Digital Part Only. If the Scheme is collecting Paper-based information at any frequency at Quarterly or higher level, give '0'. Regarding Digital scoring, give '1' at Yearly, '2' at half-yearly, '3' at Quarterly, '4' at monthly/fortnightly/weekly and '5' at Sub-state level.
3	Part B, Q C 1	Data Generation: Use of mobile surveys, geocoding or geofencing	If one of the responses is 'Yes', score '3', if two or more are 'Yes, score '5' and no response is 'Yes', score '0'.
4	Part B, Q C 2	Data Generation: Use of location tracking devices	If response is 'Yes', score '5', else '0'.
5	Part B, Q D (a)	Data Profiling	If response is 'Yes', score '5', else '0'.
6	Part B, Q D (c,e)	Data Quality Assessment	If no response is 'Yes', score '0'. If anyone response is Yes, score '3'. If two are yes, score '5'.
7	Part B, Q D (b,d)	Data Cleansing	If no response is 'Yes', score '0'. If any one response is Yes, score '3'. If two are yes, score '5'.
8	Part B, Q E	Use of Mobile Phones in QC	If M/D uses mobile phones in any one application, score '2'; if for 2 applications, score '3'; if for 3 applications, score '4'; if for 4 applications, score '5', if used in no application, score '0'.
9	Part B, Q J	Linkages with other Platforms	If the scheme is individual beneficiary based - If linkage with PFMS and JAM trinity - '5', linkage with either of the two - '3', If no linkage with either - '0' If the scheme is NOT individual beneficiary based - If linkage with PFMS - '5', If no linkage with PFMS - '0'
10	Part B, Q K	Use of Alternative Data Sources	If the scheme is LGD compliant - give a score of '3'. In addition to this, if the scheme is using any one of the other 4 non-conventional data sources - '5'. If the scheme is using non-conventional data but not LGD compliant, score - '2'. If neither LGD compliant nor using non-conventional data, score - '0'. If LGD compliance is not applicable,if the scheme is using any one of 4 non-conventional data sources - '5'. If schemes is not using non-conventional data, score '0'.

11	Part B, Q L	Use of ML/AI, Blockchain, IoT, Big Data	If the scheme is using any one of the 5 technologies - '2', If using 2/3 technologies - '4' and if using 4/5 technologies - '5'.
12	Part B, Q F	Data Analysis	If the scheme uses no method, score '0', any one method, score '1'; if for 2/7 methods, score '2'; if for 3-4 methods, score '3"; if for 5-6 methods, score '4' and if for all 7 methods, score '5'.
13	Part B, Q G	Dashboards	If response to the question is 'no', score '0'. If the response to question is Yes but no detail about purpose is given, give '1'. Within purposes mentioned, if any one purpose is mentioned, score '2'. If 2-3 purposes mentioned, score '3', if 4-6 purposes mentioned, score '4' and if 7-9 purposes mentioned, score '5'.
14	Part B, Q H1	Data Visualization: Types of visualizations	If no response is given, score '0'; if any one visualization technique adopted, score '2', if 2 techniques adopted, score '3', if 3 techniques adopted, score '4', if 4 adopted, score '5'.
15	Part B, Q H2	Data Visualization: Visualization on maps	If Yes, score '5', else '0'.
16	Part B, Q I	Dissemination/Communication	If no response is given, score '0'; if any 1 is used, score '2', 2 are used, score '3', if 3 methods used, score '4' and if all five methods adopted, score '5'
17	Part A, Q B1	Features for differently abled	If Yes, score '5', else '0'
18	Part A, Q B2	Multilingual interfaces	If Yes, score '5', else '0'
19	Part A, Q B3	Antivirus updates	If Yes, score '5', else '0'
20	Part A, Q B6	Transaction systems with internal audits	If Yes, score '5', else '0'
21	Part B, Q D (f)	Existence of Data QC teams	If Yes, score '5', else '0'
22	Part A Q C	Best Practices	If no response, score ‘0’. If case study is given but it is not a best practice, score ‘1’. If only 1 case study qualifies as best practice, score '3'. If 2 or more case studies qualify as best practices, score ‘5’.

Development Monitoring and Evaluation Office

NITI Bhawan, Sansad Marg, New Delhi-110001

contact-dmeo@gov.in | https://dmeo.gov.in/