NITI Aayog

DEVELOPMENT MONITORING AND EVALUATION OFFICE

DATA GOVERNANCE QUALITY INDEX (DGQI) 2.0

Methodology Toolkit

April, 2022

This page is left intentionally blank.

Preface

The Central Government of India, through its Ministries and Departments spends an amount to the tune of more than USD 150 billion on various Central Sector (CS) and Centrally Sponsored Schemes (CSS) schemes. With rapidly evolving governance needs and tremendous growth in data capabilities with the advent of the Fourth Industrial Revolution technologies, it is crucial for governments to ride this transformative wave and shift to evidence-based policymaking for efficient utilization of resources to achieve intended development outcomes.

Against this backdrop, DMEO, NITI Aayog along with NIC/NICSI undertook an exercise to assess the data preparedness of Ministries and Departments (M/Ds) of the Government of India and create an index called the Data Governance Quality Index (DGQI). The idea was to enable these M/Ds to periodically and consistently measure the maturity level of their administrative data systems and its use in decision-making, identify reforms to reach the frontier of seamless data exchange and its synergistic use within the M/D, and clearly define pathways to meet these goals. For developing the methodology of the DGQI measurement tool, an in-depth literature review of various global and domestic data maturity 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, quality control, management and its use, and (c) Data-driven Outcomes where siloed data systems are integrated to create an open data ecosystem where non-personal data is widely shared across institutions and used by multi-disciplinary teams to drive policymaking.

Centered around data systems, the first phase of Data Governance Quality Index (DGQI), henceforth called DGQI 1.0 was conceptualized and conducted in 2020-21 in self-assessment mode. Consequently, DGQI became a regular exercise for monitoring data preparedness levels of M/Ds and utilizing this measurement for driving specific reforms. M/Ds were encouraged to prepare a data strategy and establish a Data & Strategy Unit for improving their data maturity levels.

Following this, DGQI 2.0 was launched with enhanced horizontal focus on all three pillars: Data Systems, Data Strategy and Data driven Outcomes. The vertical scope was also expanded to include non-schematic interventions in addition to schemes.

In this context, DMEO, NITI Aayog has prepared this toolkit on Data Governance Quality Index 2.0 to enable all types of government agencies at central, state and local levels as well as any non-profit organizations implementing development programmes to learn from this experience, undertake a detailed self-assessment of their data preparedness levels for their programmatic interventions and be able to accord objective scores to these interventions. Accordingly, this toolkit can empower government agencies to have an exhaustive assessment of their degree and quality of digitization, identify areas for improvement, and design and implement reforms via a collaborative and structural approach to smoothly transition towards synergistic data ecosystems promoting active data use by policymakers.

It is hoped that this toolkit will aid policy makers at the highest level across the Government and non-government organizations to self-assess their data systems and take affirmative actions to improve data governance and use.

Acknowledgements

This toolkit is the culmination of the 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 extend my thanks to the outstanding National Informatics Centre (NIC- NICSI) team, specifically, Sh. Prashant Kumar Mittal, Sh. Varindra Seth, Sh. Rajiv Rathi, Ms.Garima Sogani and Ms. Shehnaz Nayeem under the leadership of Dr. Neeta Verma, Director General, NIC, who have not only been close partners in the execution of the entire DGQI exercise, but also supported the DMEO team in developing this self-assessment toolkit.

I would also like to thank the DMEO team - Sh. Anand Trivedi, Ms. Gunjan Saini, Sh. Krishn Kant Sharma, Ms. Aishwarya Salvi, Ms. Vaishvii Goel, Sh. Ankit Choudhary, Ms. Meera P R, Ms. Aishwarya Ravikumar, Ms. Veenu Singh, Ms. Anmol Narain, Sh. Kuldeep Pal, Sh. Bishun Chaurasia - for their outstanding efforts, diligence and dedication under the overall guidance of Sh. Shailendra Kumar Dwivedi, Joint Secretary and Sh. Ashutosh Jain, former Deputy Director General towards producing this toolkit on Data Governance Quality Index which would immensely help in driving data-driven outcomes in the Government of India.

Director General,
Development Monitoring & Evaluation Office,
NITI Aayog
April, 2022

Contents

List of Figures

List of Tables

1 Background

Administrative data forms the backbone of decentralized evidence-based decision making in the Government of India. With emerging international evidence of the vital role played by data as an enabler in driving public policy across its lifecycle, the Central and State Governments have paid significant attention to their data systems over the past two decades. Management Information Systems (MIS) and dashboards have been developed for most government schemes and programs. To disseminate this information more widely, Open Data initiatives have also been undertaken. Recently, attempts have also been made to foster data exchange across Ministries/Departments via the Prayas Dashboard at Prime Minister's Office and the Output-Outcome Monitoring Dashboard at Development Monitoring & Evaluation Office (DMEO), an attached office of NITI Aayog. Technologically, the shift towards the India Digital Ecosystem Architecture (IndEA 2.0) by the Ministry of Electronics and IT will also foster seamless data exchange and interoperability within different government digital ecosystems while enabling public-private data collaborations.

In this context, a comprehensive review of present data preparedness levels of all Ministries/Departments (M/Ds) was required to chart the way forward and suggest measures for improvement. In tune with this, the Data Governance Quality Index (DGQI) exercise was initiated with the objective of assessing data preparedness of M/Ds on a standardized framework to drive healthy competition among them, steer focused and coherent action across different stakeholder types, and promote cooperative peer learning from best practices.

Based on a detailed study of four data maturity models widely used in public as well as private domains, it was found that there are three key steps to ensure data preparedness: a) Data Strategy to lay down systemic guidelines, (b) Data Systems to ensure smooth processes of data generation, quality control, data management and its use and (c) Data driven Outcomes where data is utilized and widely shared by institutions to drive decision making. Centered around data systems, DGQI 1.0 was conceptualized and conducted in 2020-21 in self-assessment mode. The M/Ds filled up an online questionnaire to assess their data systems, the responses of which were subsequently used to generate DGQI scores. The exercise showed huge scope for improvement with respect to integrated development of data systems and using data for policymaking on a regular basis.

Consequently, M/Ds were encouraged to prepare an action plan or data strategy to lay down concrete plans to improve their data preparedness levels in general and DGQI scores in particular. They were also advised to set up a Data & Strategy Unit as a cross functional unit with four sub-units: Monitoring, Statistics, Analytics and Technology Unit to have an intersectional lens and work in close coordination with all other divisions of the M/D to foster a data driven culture in the M/D.

Following this, DGQI 2.0 was launched in 2021 with enhanced horizontal focus on all three pillars: Data Systems, Data Strategy and Data driven Outcomes. Additionally, the scope

was vertically expanded for the questionnaire to be able to assess data systems of not just schemes but also non-schematic interventions such as sector dashboards, citizen service delivery portals, etc. A self-assessment questionnaire spanning across twelve themes within the three pillars was circulated with M/Ds on an online portal. Several training webinars were held with M/Ds and resources were shared to assist them along with regular telephonic follow-up calls and one-to-one meetings. Scoring methodology to generate DGQI scores for M/Ds was also finalized in close consultation with M/Ds and transparently shared with them in advance.

These DGQI scores are expected to keep guiding M/Ds while steering necessary reform actions to improve their data governance. Regular tracking of progress against the action plans prepared by the M/Ds has also been ingrained in the self-assessment questionnaire and accorded appropriate scoring to push M/Ds to keep up the reform momentum.

This way, via a structural measure-and-reform approach supported with continued collaboration with M/Ds, DGQI exercise aims to bring about a paradigm shift in data-based governance in India. It is hoped that it will foster a culture of evidence-based policymaking with the use of high-quality, seamlessly interconnected administrative data systems in India.

2 Objectives

This DGQI 2.0 toolkit has been developed with the overarching objective to widely disseminate the learnings from implementing the DGQI exercise so that the framework can be used and contextualized by other stakeholders to self-assess and improve their methods of data governance and management throughout its lifecycle, especially in the public sector.

The toolkit aims to enable government agencies at any tier (Centre, States, Districts or City, implementing agencies, and Public Sector Undertakings) across different geographies to self-assess their data or management information systems and identify areas where there is a need to design better systems or integrate existing systems for more effective programme implementation and monitoring.

The toolkit has the following specific objectives:

a) To enable review and assessment of data preparedness of the data/ MIS systems on objective parameters of a standardized framework.
b) To disseminate a self-assessment diagnostic tool that will enable government agencies to internally contemplate on the need for improving data systems while undertaking the assessment.
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.
d) To enable users and commissioning agencies to use the measurement to drive necessary reforms in the direction by adopting a structural measure-and-reform approach

This document hence presents in detail the architecture, approach and methodology which may be pursued by government agency adopting DGQI framework to measure and improve their data governance.

3 Scope and Use Cases

As mentioned above, this DGQI ready-to-use toolkit has a wide scope across public sector and can be easily adopted by government agencies at any tier (Centre, States, Districts or City, implementing agencies, and Public Sector Undertakings) across different geographies.

Some use cases of how can the toolkit be useful for different government stakeholders in improving their administrative data use for policymaking have been outlined below.

Use Case 1: By Planning Department or apex Monitoring & Evaluation (M&E) Department of Central or State Governments

The Planning Department or the M&E Department of Centre/States often need to analyze and use datasets from various other Departments (handling different sectors) together. Such synergistic use of data is quite often required to track performance against common national goals such as the National Development Agenda or international goals such as the Sustainable Development Goals. This data can then be used for making more informed planning decisions. For instance, in order to achieve a multi-modal transportation system, the Planning Department would not only have to monitor specific projects in roads, railways and waterways sectors, but would also need data on interlinkages between these modes of transportation. In pursuit of such goals, there have been several disparate attempts to create cross-sectoral integrated dashboards that allow monitoring and planning decisions at the highest levels. The PRAYAS Dashboard created by the Office of the Prime Minister of the India, State CM Dashboard of several states like Gujarat, Madhya Pradesh, SDG Dashboard of Andhra Pradesh are some of the prominent examples.

However, siloed administrative data systems hinder such cross-sectoral data use. Some of the required data systems may still be paper based or not available in machine readable formats. If digitized, unit-level and/or latest data may not be available. Even if all required data is available, irregularities in data standards/classifications used may not allow them to be analyzed together. It is also often noted that data from one Department is at a different granularity and frequency than that of another Department.

In order to mitigate some of these challenges, the Planning or M&E Department of the Government can use DGQI framework for a comparative assessment of data maturity across its various departments to identify theme-wise areas for improvement. This would also simultaneously help in peer learning as well as in standardization of data systems that can be easily analyzed together. Apart from measurement, the overall DGQI implementation architecture including its suggested reforms can enable Planning departments to steer multiple line departments towards improved data governance quality standards in a time-bound and focused manner. It will also lay down the foundation for intra and inter departmental coordination and allow for easier one-stop-

access to data enabling stronger evidence-based policymaking in the future.

Use Case 2: By IT Department of Central or State Governments

The IT Ministry or Departments of governments regularly roll out e-governance and data management related frameworks, policies and standard guidelines, in line with continuous technological changes. However, the adoption of these policies and standards across different line ministries and departments across Central and State governments are found to be varying, depending upon the top-down push within the government. In such situation, it becomes difficult to ensure harmonious adoption across all Departments which is a precursor to inter-departmental data exchange required to realize the goal of government-as-a-platform.

Hence, IT Ministry or Departments can use the DGQI approach to undertake a comparative assessment of compliance across Departments and ensure quicker adoption of existing policies.

Use Case 3: By Finance Department of Central or State Governments

The Finance Departments across governments are rapidly moving to performance-based budgeting to bring in accountability in government expenditure. However, to be able to undertake this exercise in an effective manner, Finance Departments need reliable data on performance of government programmes from several Departments. For instance, several Ministries/Departments of the Government of India have begun to annually report data on the performance of their Central Sector/Centrally Sponsored Schemes as part of the Output Outcome Monitoring Framework (OOMF) published as a part of the Union Budget each year and utilized for periodic reviews at the highest level within the Government of India. Though performance data has become easier to gather with rapid digitization of government services and programmes, the Finance Department needs to be sure of the reliability of reported data before using this for budgeting decisions. However, at present, lack of data quality, reliability and timeliness in reported data plays a major roadblock in the path of institutionalizing performance budgeting in India.

To overcome these challenges, Finance Department of Centre/State governments can use DGQI to conduct a comparative assessment of the processes governing data quality of different performance data streams across various program divisions and departments to identify areas for improvement and trigger necessary reforms in the direction. This can go a long way in improving reliability of administrative data reported for government schemes/programmes, give financial advisers greater confidence to use the performance data for budgeting decisions, and thereby, demonstrate accountability and transparency in government budgeting and expenditure in the long run.

Use Case 4: Any other Department of Central or State Governments / Public Sector Undertakings (PSUs) with multiple divisions or agencies under the Department /

PSU

A single department often has multiple divisions operating under it. For instance, a State Education department may have different scheme divisions responsible for implementation of Samagra Shiksha Scheme, Mid-Day Meal Scheme and Scholarship Schemes. Similarly, Ministry of Power has several different implementing agencies responsible for different sectoral activities such as Central Electricity Authority, Rural Electrification Company, Power Finance Company etc.

In such cases, monitoring the department or sector-wide goals requires coordination with multiple agencies or divisions. However, these divisions often have their own administrative data systems which operate in siloes with each other. For instance, the Mid-Day Meal data system is not yet integrated with the UDISE data system of the Ministry of Education. In such cases, synergistic data use to monitor sector wide performance may become difficult. For example, it may be quite helpful to estimate the correlation between availability of mid-day meals on school enrolment rates. However, with one variable available in Mid-Day Meal MIS and the other available in UDISE MIS with no common school identifier, it may not be possible to use administrative data for this purpose.

In addition, at times, data reported by divisions is not available at the right granularity or frequency in digitized formats at which it may be required by the department heads for quick decision making. For example, project level data is required for monitoring infrastructure-oriented schemes such as Sagarmala, Bharatmala Pariyojana etc. However, it is often noted that such scheme data systems/MIS have highly aggregated data. Project level data is not digitized and often stored in separate files. Hence, project level monitoring reviews and decision making by the department Heads takes significantly more time.

It is hence useful for all departments to adopt the DGQI framework to assess the data maturity levels of their divisions to ensure their scheme divisions collect high quality, granular, and near-real-time data. Further, it will also ensure that all MIS systems within the department can communicate with each other by following similar classifications/standards. This way, DGQI can help in creating better MIS right at the design stage of schemes and provide easier, one-stop-access to all sector related data to department heads. This can go a long way in regularly monitoring scheme performance, making suitable corrections in implementation, inform design of new schemes/policies and hence eventually improve performance of the sector.

Use Case 5: By local government bodies/ district administrations

Local government bodies and district administrations are responsible for a lot of data collection during the day-to-day implementation of government services and schemes for upward reporting to above authorities. District administrations and local bodies need to

hence ensure they are reporting good quality data. However, owing to paper based data collection which is later on fed on digital systems, data is prone to human errors. Further, it often does not follow the data quality protocols or the classifications/standards set by above levels.

While this data, if rightly collected, can be used by these bodies to improve the efficiency of public service delivery on a regular basis, it is often witnessed that such forms of data use remain quite restricted. For instance, real time transactional data on purchases of foodgrains from PDS shops can be used to monitor the total stock of available foodgrains and escalate requirement for procurement/transportation of additional stock as and when need arises. Grievance redressal of citizens regarding PDS can also be digitized via a mobile app. However, such practices are rare to witness and data is more often than not collected merely for reporting purposes at these levels. Thus, the design of administrative data systems to meet the decentralized decision-making needs of field functionaries needs a closer look.

There is huge merit for such local bodies to use DGQI toolkit to measure the data maturity levels of their current programmes and public services while understanding the degree of its use for decentralized decision-making. They can use this tool as a way to identify what all data they are collecting, how can these data collection processes be streamlined to reduce duplication of efforts and optimize reporting burden. Further, they can also use it to identify gaps and develop capabilities to foster decentralized data use to improve the efficacy of public administration.

Use Case 6: By any other government division/agency/unit implementing or monitoring a single scheme/initiative/programme at any tier

Finally, even smaller government divisions or agencies that are responsible for implementing and/or monitoring any government initiative can make use of the DGQI framework. Such divisions need to oversee day-to-day implementation of a government programme to make mid-course corrections such as checking for leakages in a subsidy scheme, exclusion/inclusion errors in government programmes, fund supply in line with physical progress etc.

However, they often struggle with reliable data available at the right granularity/frequency to be able to make such decisions. For example, in absence of a unique farmers' database, there is no way for the scheme division to ensure that fertilizer subsidy is being routed to eligible farmers. Further, due to limited digitization of land records and soil health cards, it is not possible to ensure fertilizer sales can be made as per the nutrient requirements of the soil.

DGQI toolkit can be helpful for such divisions to self-assess themselves on various parameters of data management. This will help them identify areas for improvement, capacity gap and digitize their systems in a coherent manner that collects high quality

data and aids day-to-day implementation. They can also use this tool as a way to identify what all data they are collecting, how can they collect and use it better while optimizing their reporting burden.

The above six use cases are only a glimpse of the way the DGQI toolkit can be beneficial for driving development of synergistic data systems and promoting data driven decision making at various levels. Apart from the stakeholder-specific benefits listed above, there can be a wide-ranging impact of the deepening of digital systems triggered by the DGQI exercise.

Reforms undertaken to improve the standardization and maturity of data systems as a consequence of the DGQI assessment can be used by the Central and State governments to develop integrated government-wide shared data systems. These systems can foster a culture of data use among the departments. Self-assessment mode can help in driving ownership from various stakeholders for identification and diagnosis of reform actions required to deepen digitization in India at various levels. Comparative assessment allows for peer learning and harmonization of efforts across departments. It can also be used as a needs assessment for identifying components that may be required to implement an integrated data governance framework across the Government of India.

These efforts will have multiplier effects in improving efficiency as well as effectiveness of public administration while bringing in more accountability and transparency. It has the potential to make public service delivery more responsive to citizen needs and strengthening the government to citizen engagement in the country. This way, adoption of DGQI toolkit can play a key role in embedding evidence-based policymaking in the DNA of the Government of India. This can go a long mile in preparing India to be able to ride the rapidly evolving technological and data wave across the world.

The next sections present the overall architecture, operational approach and methodology adopted as a part of DGQI framework. Further, it is extremely crucial for users to suitably customize the toolkit based on their tier, role and type of intervention to be able to make effective use of the framework. Detailed guidance on how can the above-mentioned user types customize and adopt the DGQI toolkit has been provided in Section 7.

4 Architecture

To meet its intended objectives, the architecture of the Data Governance Quality Index (DGQI) exercise was based on two main arms: first, measurement and second, action.

A standard yet comprehensive self-assessment questionnaire with the ability to assess diverse types of data systems used for different kinds of interventions was used for self-measurement of data maturity levels by M/Ds. There was a special emphasis on self-assessment instead of external evaluation to allow for internal contemplation among M/Ds on the need to improve their data maturity levels. This way, the measurement is more self-reflective instead of being a competitive tool, reducing the incentive to provide biased responses.

Next, this measurement was used to drive action by M/Ds to improve their data preparedness level. To drive these reform actions, three main levers of change were identified: first, developing M/D-wise action plans for time-bound improvement in their respective DGQI scores; second, setting up necessary institutional architecture at M/Ds in the form of Data & Strategy Units (DSUs) to lead the implementation of these action plans; and finally, capacity-building of both the existing as well as new human resources within the M/Ds to implement the action plan and sustain these improvements in data ecosystem at M/Ds in the long run.

The M/Ds were encouraged to actively focus on the above three levers of action. DMEO, NITI Aayog provided necessary support to the M/Ds to be able to undertake these massive steps. As it was found that an action plan or data strategy is pertinent to be framed by all M/Ds to lay down a concrete plan with clear actionables and create provisions for infrastructural, human and financial resources, an indicative outline of the action plan (see Annexure 1) was prepared by DMEO, NITI Aayog to support M/Ds to come up with exhaustive and harmonious yet customized elements of their plans. Next, as an institutional mechanism, a 'Data and Strategy Unit (DSU)' was required to be set up within each M/D to steer the development and implementation of action points embodied in these plans. A detailed Terms of Reference (see Annexure 2) for setting up these DSUs was prepared by DMEO, NITI Aayog to support M/Ds with indicative structure, size, and placement of the unit. For capacity development, regular capacity-building efforts were undertaken by DMEO, NITI Aayog in the form of broadcast webinars and one-to-one sessions. M/Ds were also encouraged to partner with other stakeholders and embed DGQI related components in their capacity development plans.

DGQI is action oriented and hence to ensure continued focus on these reforms, two themes named 'Data& Strategy Unit (DSU)' and 'Action Plan' were added to the self-assessment questionnaire. These themes check and score M/Ds for well-designed and implemented plans supported by well-staffed units. Within action plan, it was decided that the M/Ds would not only be scored for preparing a plan, but also on timely compliance on the action points that they have laid down for themselves. This tracking and incentivization of strategy implementation was identified to be an essential way of driving reform actions without which the paradigm shift could not be achieved.

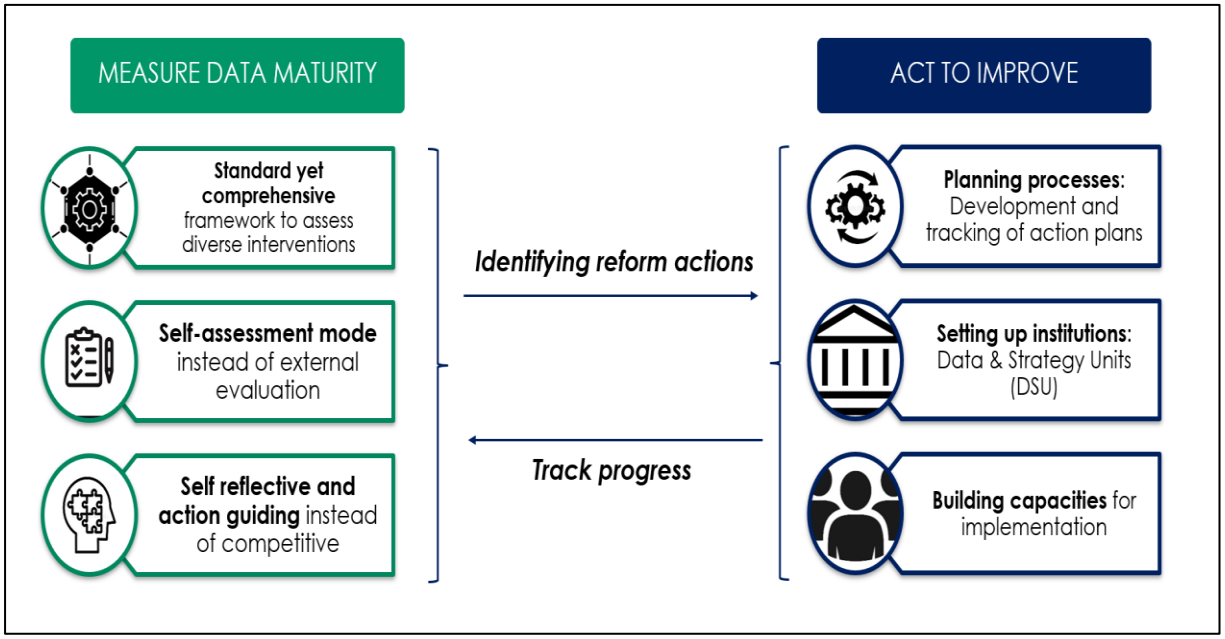

Figure 1 DGQI Architecture

This way, DGQI tries to utilize baseline measurement for identification of reform actions by participating agencies. Next, it embeds regular tracking of reform actions within future rounds of measurement. This way, by regularly tracking progress on reforms (activities) as well as on systems (outputs) and data driven outcomes (outcomes), it adopts an agile approach to steer through the entire results chain towards transformed administrative data in India.

Other government agencies aspiring to adopt DGQI for their self-assessment can learn from this structural measure-and-reform based architecture so that they not just measure but also use it to drive reform actions. All three levers of change must be focused upon (action plans, institutional setup to implement plans and capacity development) by the commissioning agencies while devising their reform processes. However, they may need to suitably modify the outline of the action plan and the structure of the institution based on their present status and organizational needs. For instance, if a State government wants to develop their own action plan, while the overall structure of the action plan as presented in Annexure I may continue to be used, instead of just a scheme wise strategy, the action plan may also need a department wise strategy. Similarly, if a State department wishes to set up a DSU for the implementation of their digitization plans, while they may continue to have integrated monitoring, statistics, technology and analytics units, their roles and responsibilities may need to be suitably modified to be applicable at State level. Similarly, for a district or city government, the structure of their action plans should nudge their departments to think more from the perspective of decentralized data generation and use for improved citizen services delivery. In summary, while the overall principles and architecture of DGQI may be adhered to for greater effectiveness, the finer

elements of its implementation must be customized to meet specific stakeholder needs.

5 Operational Approach

There were two main enablers that supported a smooth execution of the DGQI exercise.

First, the entire exercise has been conducted in close partnership with participating Ministries/Departments in the spirit of cooperative action. The self-assessment questionnaire and methodology were transparently shared with all Ministries/Departments in advance and finalized in close consultation with them. This helped ensuring that the tool is exhaustive and responsive to requirements of different participating agencies. It also helped in ensuring that the Ministries/Departments take active interest in understanding the questionnaire, which in turn helped them contemplate on their areas for improvement. Further, M/Ds were provided tailormade recommendations to improve their data preparedness after every round of the exercise. Such one-to-one feedback and regular contact with nodal officers at Ministries/Departments has aided in inviting proactive interest from officials in the domain of data governance, which is expected to help sustain the objectives of the exercise in the long run.

Second, conducted in the middle of the Covid-19 pandemic, the entire exercise has been technology driven. The self-assessment questionnaire was hosted on an online dashboard with separate credentials provided to all M/Ds. It offers a macroscopic picture of overall progress on the DGQI initiative as well as a microscopic view of different M/Ds and schemes on each theme of the index. With easy-to-update questionnaire and action points modules, automated scores generated in real-time as soon as the M/Ds fill in information, user-friendly visualizations offering comparative analysis, capabilities to undertake trend analysis and downloadable compliance and progress reports, this dashboard is a one-stop-shop for all M/Ds to regularly review and take this exercise forward in an effective manner. Regular training webinars and preparation of training resources on the website for public dissemination have also facilitated in capacity building and ensuring that the questionnaire is correctly responded.

This way, other government agencies planning to adopt DGQI framework must focus on integrating the above two aspects in their operational approach for undertaking the exercise in an effective manner.

6 Methodology

Based on a detailed assessment of several data maturity models (see Annexure 3 for details), 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. All pillars are explained below.

6.1. Data Strategy

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

Under the data strategy pillar, two themes are covered within DGQI 2.0: a) Data and Strategy Unit and b) Action Plan.

These two themes have been explained in detail below.

6.1.1. Data & Strategy Unit

Ministries/Departments were advised to set up a Data & Strategy Unit (DSU) as a central unit to steer the development and implementation of an action plan or data strategy to improve their data preparedness levels as part of the DGQI exercise.

Within this theme, it was assessed if the Ministries/Departments have taken necessary steps in this direction to establish the necessary arrangements that are required for the development and maintenance of a robust data strategy.

6.1.2. Action Plan

Ministries/Departments were also advised to develop an action plan or data strategy with clear actionables, definite timelines and responsibilities to improve their data preparedness levels as part of the DGQI exercise. An indicative outline of the action plan was also shared for reference and guidance.

Within this theme, it was assessed if the Ministries/Departments have developed action plans as per the outline. In addition, the compliance by the Ministries/Departments in completing the action points within the timelines set by themselves were measured.

6.2. Data Systems

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

Under the data systems pillar, six themes are covered within DGQI 2.0: data generation (ability of M/Ds to collect and digitize data at high granularity and frequency); data quality (practices adopted by M/Ds to undertake data quality assessment of incoming

data); data analysis, use & dissemination (use of collected data for analysis and decision making, open data and modes of dissemination); use of technology (use of emerging technologies and alternative data sources); data security & HR capacity (measures to ensure data security and protection of personal data and existence of data QC and analysis teams); and data management (adoption of lifecycle approach to data management).

These six themes have been explained in detail below.

6.2.1. Data generation

This theme measured the ability of Ministries/Departments to collect and report generate data on inputs, outputs and outcomes of their schemes. It covered areas related to the granularity and frequency of digitization and also covered if approaches like CAPI surveys, GIS mapping, transactional data collection etc. are used to improve quality of generated data.

6.2.2. Data quality

This theme measures whether Ministries/Departments undertake data quality assessment procedures to evaluate the quality of incoming data and make suitable corrections. Key areas included under the theme pertain to data quality assessment, automation of data quality assessment and use of latest feedback and backcheck mechanisms to further validate data quality.

6.2.3. Data analysis, use & dissemination

This theme measured the ways in which collected data is analyzed and used by Ministries/Departments for evidence creation and decision making. Use of dashboards and other modes of dissemination were also included within this theme. Key areas also include ensuring accessibility of data, machine readability of data and open data systems for wider dissemination.

6.2.4. Use of technology

This theme covered linkage of Ministries/Departments' portals with other platforms like PFMS, JAM, GSTN, Udyog Aadhaar, LGD etc. wherever applicable. Use of alternative data sources outside the government like remote sensing data, social media data etc. to improve data robustness and use of emerging technologies like AI/ML, Drones, etc. in scheme monitoring are other key areas.

6.2.5. Data security & HR Capacity

This theme measured the capacity of Ministries/Departments to ensure data security and privacy related concerns of their data systems. It also covered questions on human resource capacity of data quality and analysis teams for various schemes of Ministries/Departments.

6.2.6. Data management

This theme covered areas related to data management across its lifecycle, i.e., guidelines for data management, data storage and historical data management.

6.3. Data driven outcomes

The first and the second pillar work in conjunction with each other to enable the third pillar of data-driven outcomes. Mere existence of data strategies and systems alone cannot ensure that data is used seamlessly and optimally 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 across administrative silos, development of strong data analytics capabilities, and well-articulated data use plans. These aspects hence get covered under the third pillar – data-driven outcomes.

Under this pillar, four themes have been identified under DGQI 2.0: Synergistic data use within M/Ds (creation of better exchange systems within M/Ds to drive integrated data use); inter-agency collaboration (data-based collaborations with other agencies to drive better data-based outcomes); prescriptive analytics (creation of data culture by moving to prescriptive analytics); and good practices (good practices in using data in driving smarter, granular and quicker decisions).

These four themes have been explained in detail below.

6.3.1. Synergistic data use

Often, data available within the Ministries/Departments may not be fully used for decision-making. Alternatively, sometimes, data gaps exist for enabling better policymaking. Therefore, this theme covered how Ministries/Departments have identified gaps in their available data from decision-making perspective, what corrective actions have been taken to fill-in those gaps, and steps taken internally to create better exchange systems to drive integrated data use.

6.3.2. Inter-agency data collaboration

This theme covered how Ministries/Departments have undertaken data-based collaborations with other agencies to drive better data-based outcomes and create a rich data culture in the organization.

6.3.3. Prescriptive analytics

This theme covered how Ministries/Departments are trying to create a data culture by moving to prescriptive analytics and developing mechanisms for institutionalizing it in the long run.

6.3.4. Good practices

This theme highlighted good practices adopted by Ministries/Departments in using data

in driving smarter, granular and quicker decisions for informing policy along with its quantified impact. It is expected to help unlock hidden potential by opening doors for cross-learning from challenges faced and solutions devised by peers.

6.4. Scoring

A self-assessment questionnaire was devised around the above twelve themes and DGQI scores were arrived at on the basis of responses filled up by M/Ds to this questionnaire. The questionnaire consisted of two parts: Part A (to be filled at M/D level) and Part B (to be filled for each CS/CSS scheme/non-schematic intervention at CS/CSS/NSI level). The questionnaire can be viewed at Annexure 4.

The response to each question was scored on a scale of 0 to 5, which was then aggregated using weighted averages to arrive at scores at themes, pillar and overall index level (all scores range between 0 to 5). The data systems pillar was appropriated an overall weight of 60% as it is a major pillar where outputs of data strategy are visible which then also play a key role in the ability of M/Ds to achieve desired data driven outcomes. Remaining 40% weight was accorded to the data strategy and data driven outcomes pillar combined. This 40% was distributed equally between data strategy (20%) and data driven outcomes (20%).

Hence, overall DGQI Score = 60% *(Data systems pillar score) + 20% *(Data strategy pillar score) + 20% *(Data driven outcomes pillar score)

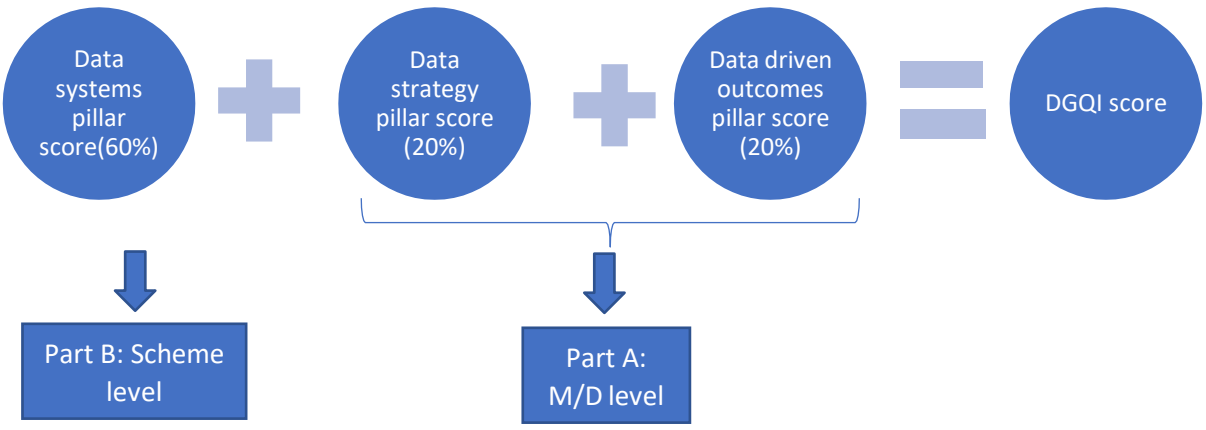

Figure 2 Calculating overall DGQI Score

To arrive at each pillar score, each theme was further assigned a weight, as shown below.

6.4.1. Data systems pillar

Apart from use of technology, all remaining themes within data systems were decided to be allocated equal weightage as all these systems were found to be equally important for

ensuring robust data systems. Use of technology was allocated a weight of 10% to promote the use of emerging technologies across data systems.

Pillar	Theme	Theme weightage within data systems pillar
Data Systems	Data Generation	18%
	Data Quality	18%
	Data analysis, use & dissemination	18%
	Use of technology	10%
	Data security & HR capacity	18%
	Data management	18%
	All themes	100%

Table 1 Theme wise weightages within data systems pillar

Hence,

Data systems score = 18% * (Data generation score) + 18% *(Data quality score) + 18% * (Data analysis, use & dissemination score) + 10% * (Use of technology score) + 18% *(Data security & HR capacity score) + 18% * (Data management score)

As explained above, data systems pillar scores were based on scheme-level information provided in Part B of the self-assessment questionnaire. Hence, for each scheme (filled up in Part B), a data systems score was generated using the above formula. Then, a simple average of these scheme-level scores was calculated to arrive at a combined data systems score.

6.4.2. Data strategy pillar

Both DSU and action plan were appropriated equal weightages as both were found to be equally important components of data strategy pillar.

Pillar	Theme	Theme weightage within data strategy pillar
Data Strategy	Data & Strategy Unit	50%
	Action Plan	50%

Table 2 Theme wise weightages within data strategy pillar

Hence,

Data strategy score = 50% * (Data & Strategy Unit score) + 50% * (Action plan score)

6.4.3. Data driven outcomes pillar

Driving intra-ministerial and inter-ministerial integrated use of data is one of the key outcomes for fostering a data culture. Similarly, good practices offer huge scope for peer learning. Hence, these three themes were given the highest equal weightages.

Pillar	Theme	Theme weightage within data driven outcomes pillar
Data driven outcomes	Synergistic data use within M/D	30%
	Inter-agency data collaboration	30%
	Prescriptive Analytics	10%
	Good practices	30%

Table 3 Theme wise weightages within data driven outcomes pillar

Hence,

Data driven outcomes score = 30% * (Synergistic data use within M/D score) + 30% * (Inter-agency data collaboration score) + 10% * (Prescriptive Analytics score) + 30% * (Good practices score)

6.4.4. Question wise weightages

As each theme had multiple questions within it as a part of the self-assessment questionnaire, each question was also accorded an appropriate weightage within the theme. The same is tabulated below for all questions.

Pillar	Theme	Question No.	Question	Question Weightage within theme
Data Strategy	Data & Strategy Unit	Part A, B1	Constitution	5%
		Part A, B2	Head	5%
		Part A, B3	Verticals	10%
		Part A, B4	Strength	20%
		Part A, B5	ToR	20%
		Part A, B6	Review mechanisms	20%
		Part A, B7	Frequency of review	20%
	Action Plan	Part A, C1	Action plan	5%
		Part A, C2	Sections	5%
		Part A, C3	Schemes	5%
		Part A, C4	Timelines	5%
		Part A, C5	Responsibilities	5%
		Part A, C7	Compliance Scoring	75%
Data Systems	Data Generation	Part B, A1	Requirements gathering	10%
		Part B, A2	Collection	10%
		Part B, A3	Digitization	20%
		Part B, A4	Granularity	20%

Pillar	Theme	Question No.	Question	Question Weightage within theme
		Part B, A5	Frequency	20%
		Part B, A6 and 7	Use of technologies in generation	20%
	Data Quality	Part B, B1	QC mechanisms	20%
		Part B, B2	QC automation	20%
		Part B, B3	Data quality assessment	40%
		Part B, B4	Use of mobile phones in QC	20%
	Data analysis, use & dissemination	Part B, C1	Types of data analysis	15%
		Part B, C2	Cross sectoral analysis	10%
		Part B, C3	Documentation of data analysis	10%
		Part B, C4	Use of data analysis	15%
		Part B, C5	Modes of dissemination	5%
		Part B, C6	Use of dashboards	15%
		Part B, C7	Data visualization	5%
		Part B, C8	Data visualization on maps	5%
		Part B, C9 and 10	Data Accessibility for all	5%
		Part B, C11 and C13	Open data	5%
		Part B, C13	Open data – 2	5%
		Part B, C12	Machine readable data	5%
	Use of technology	Part B, D1	Linkage with PFMS	10%
		Part B, D2	Last mile linkage of PFMS	20%
		Part B, D3	Linkage with other platforms	20%
		Part B, D5	Linkage with LGD Codes	20%
		Part B, D4	Use of alternative data sources	10%

Pillar	Theme	Question No.	Question	Question Weightage within theme
		Part B, D6	Use of emerging technologies	20%
	Data security & HR capacity	Part B, E1	Antivirus updates	5%
		Part B, E2	Security audits	5%
		Part B, E3 and 4	SSL certification	5%
		Part B, E5	Firewalls	5%
		Part B, E6	External communication	10%
		Part B, E7, 8	Personal data protection	10%
		Part B, E9	Personal data protection -2	10%
		Part B, E10	Data QC team	25%
		Part B, E11	Data analysis team	25%
	Data management	Part A, D1,3,4,5	Data management architecture	25%
		Part A, D2	Data management Compliance	10%
		Part B, F1 and 2	Distributed storage	25%
		Part B, F3 and 4	Cloud storage	25%
		Part B, F5	Historical data management	15%
Data driven outcomes	Synergistic data use within M/D	Part A, E1 and 2	Identification of data gaps	40%
		Part A, E3	Data exchange	60%
	Inter-agency data collaboration	Part A, F1	Collaborations	50%
		Part A, F2	Types of collaborations	50%
	Prescriptive Analytics	Part A, G1	Prescriptive analytics	50%
		Part A, G2	Frequency	25%
		Part A, G3	Modes	25%
	Good Practices	Part A, H	Good practices	100%

Table 4 Question wise weightages within each theme

Question wise scoring mechanism for each question can be found at Annexure 5.

6.5. Special Cases

To consider the non-applicability of certain questions or sub-parts of questions, NA option was explicitly included in the DGQI self-assessment questionnaire. For acertain question, if NA option is selected, its weight has been redistributed among other questions within the theme. However, if it is the case that only certain sub-parts (a,b,….) of a question were not applicable, a case-by-case mechanism of how they will be taken care of at the scoring stage was devised. The same can be found at Annexure 6.

Similarly, if any question was disabled based on skipping patterns of the self-assessment questionnaire, it was accordingly given appropriate score. For example: If action plan was not formed, M/Ds would be scored zero on all other questions related to action plan that get automatically skipped.

6.6. Summary

This way, in order to arrive at DGQI scores, a three-tiered weighted average process is used:

(a) First, weighted average of question wise scores within each theme.

(b) Second, weighted average of theme-wise scores within each pillar. Within this step, for data systems pillar, initially, data systems scores are calculated for each scheme separately. To aggregate the same into a single score at M/D level, a simple average of these scheme level scores is calculated to arrive data systems pillar score.

(c) Third, weighted average of pillar-wise scores to arrive at final DGQI score for the M/D.

7 How to use DGQI

As mentioned in section 3, the DGQI framework can be used by several government agencies at various tiers (Central, State, District, Local Bodies, PSUs etc.). to improve their data preparedness.

Previous few sections of this toolkit detail out the architecture, approach and methodology of the Index for other users to be able to utilize the same. However, as mentioned earlier in Section 3, the framework may need to be suitably customized by the users for their interventions and data types before being able to use the same. Indicative steps which may be followed by the users of this toolkit to adopt and use DGQI framework have been provided below:

1. **Understand & Customize**: Begin by understanding the objectives, architecture and methodology of the DGQI exercise as implemented by DMEO, NITI Aayog using this toolkit. The framework and associated self-assessment questionnaire can then be suitably customized to suit the needs of the organization using DGQI. This adaption can include modifications in questions and options as per requirement (for eg: national granularity option may not be relevant at state level assessments), removal of non-applicable questions/sections (some questions about national policy use may not be applicable to districts), addition of other questions which may seem relevant (questions on ease of reporting of data by field level functionaries may be more applicable at district/block level assessments; districts or local bodies are more heavily involved in delivery of services to citizens – hence suitable questions may need to be added to cover data systems of such non-schematic interventions). Similarly, methodology may be suitably modified by user agencies depending upon their data requirements (for eg: more weightage may be appropriated to data driven outcomes by any State government which is already at an advanced level of data systems maturity).

Detailed guidance on how can DGQI toolkit be customized for the six use cases introduced in Section 3 is provided below.

Use Case 1: By Planning Department or apex Monitoring & Evaluation (M&E) Department of Central or State Governments

As mentioned in Section 3, the Planning Department or apex Monitoring & Evaluation Department of governments presently face challenges due to siloed data systems hindering effective data use to use it for regular monitoring and for making strategic planning decisions.

The Planning/M&E Department can customize and use DGQI toolkit to conduct a comparative assessment of data systems of other departments. Necessary questions may need to be added to identify and measure if the data systems of other departments report data on all necessary indicators of scheme/sector performance. Specifically, questions

may need to be added to identify what outputs and outcomes are being tracked by the MIS systems. Further, if there are any non-schematic interventions such as e-governance services administered by the departments, the questionnaire may need suitable customization to cover such interventions. For instance, in case of e-services, the user-friendliness of the portal and efficacy of grievance redressal mechanisms are some of the additional themes that may need to covered.

Use Case 2: By IT Department of Central or State Governments

As mentioned in Section 3, the IT Department of governments presently face challenges in ensuring harmonious compliance against national e-governance or data management related policies across departments.

The IT Departments can customize DGQI toolkit to mainly focus on measuring compliance against existing policies. As the present questionnaire only covers some of the policies/guidelines relevant for Ministry/Department level MIS systems, State IT Departments can suitably modify the same depending upon the policies that are applicable to them and include them exhaustively in the questions.

Use Case 3: By Finance Department of Central or State Governments

As mentioned in Section 3, the Finance Department of governments presently face challenges in having access to good quality reliable data that may be used for budgeting purposes.

The Finance Departments can hence customize DGQI toolkit to add questions on the kind of outputs and outcome indicators are being tracked by the MIS systems and check if data is being collected on all indicators at the required granularity and frequency for performance budgeting as well as for inter alia fund allocation between States/Districts and timely releases.

Use Case 4: Any other department of Central or State Governments / Public Sector Undertakings (PSUs) with multiple divisions or agencies under the department / PSU

As mentioned in Section 3, other departments or PSUs with multiple divisions also face constraints with synergistic use of data for sector/organizational goal tracking due to siloed data systems.

Such stakeholders may also need to suitably modify the DGQI toolkit before being able to use the same. If there are any non-schematic interventions such as e-governance services administered by the department, the questionnaire may need suitable customization to

cover such interventions. For instance, in case of e-services, the user-friendliness of the portal and efficacy of grievance redressal mechanisms are some of the additional themes that may need to covered. The National e-Governance Service Delivery Assessment (https://nesda.gov.in/publicsite/) by Department of Administrative Reforms and Public Grievances may be used as reference for this customization.

In case of PSUs which are involved in business operations, enterprise resource planning (ERP) systems need to be deployed and digitized to ensure integrated management of business processes such as financials, supply chain management, trade operations etc. Hence, the questionnaire may need to be significantly overhauled to measure the data maturity of such ERP systems.

In addition, specifically considering the data analysis, dissemination and use theme of the questionnaire, there would be merit in customizing the uses for which data is analyzed. As funding decisions may not be applicable at divisional level, use of data for funding decisions may need to be removed from the questionnaire. Instead, more specific use of how data is used to improve implementation processes may need to be added. For instance, how is data used to identify leakages and fix them, how is data used to manage supply chains and demand, how is data used to improve operational efficiency of involved human resources at the ground level etc.

Further, data strategy and data driven outcomes need synergies at departmental/organizational level and hence measured only at the departmental/organizational level. Hence, the department/PSU may need to ensure that these themes are assessed only once – at the central level and not separately at each divisional level while customizing the questionnaire.

Use Case 5: By local government bodies/ district administrations

As mentioned in Section 3, local government bodies / district administrations need to collect a lot of data for reporting purposes. In addition, this data can be used by them for overseeing and improving day-to-day implementation of schemes and delivery of public services.

They can use the DGQI toolkit to effectively enable the above after necessary customization. As local government bodies are often the closest unit involved in ensuring smooth delivery of public services to citizens, a lot of focus would be required to customize the questionnaire to include interventions from this angle. For instance, in case of e-services such as filing of property tax return, the user-friendliness of the portal, last mile digitization of the process and efficacy of grievance redressal mechanisms are some of the additional themes that would need to covered. The National e-Governance Service Delivery Assessment by Department of Administrative Reforms and Public

Grievances may be used as reference for this customization.

In case of essential public goods and services provided by these bodies, the use of administrative data by these bodies in ensuring efficient administration of the delivery of these services may need more focus. For example, urban municipal corporations can use night light data to identify pockets which are not well-lit and require installation of street lights to ensure citizen security. Installation of GIS systems in waste collection vehicles can generate a lot of data which can be then used to monitor and optimize the process.

Use Case 6: By any other government division/agency/unit implementing or monitoring a single scheme/initiative/programme at any tier

As mentioned in Section 3, even singular divisions responsible for any scheme implementation or service delivery can use DGQI toolkit after necessary customization to self-assess and improve their data systems so that they can easily use it for monitoring and decision making.

If the division undertakes any non-schematic intervention such as sector monitoring or an e-governance service, the questionnaire may need suitable customization to cover such interventions. Further, in order to improve their data use, they can add specific questions on what kind of data needs to be collected from decision making perspectives vis-à-vis what kind of data is being actually collected. They can also re-evaluate the data quality protocols and data generation mechanisms listed in the DGQI questionnaire to identify which are more relevant for them to use for their intervention. Data analysis use cases may also need necessary customization – while allocation of funding to the division may not be applicable, there may be a need to add details on how is collected data used by the division to route funding to implementing agencies. Similarly, more specific use of how data is used to improve processes may need to be added. For instance, how is data used to identify leakages and fix them, how is data used to manage supply chains and demand, how is data used to improve operational efficiency of involved human resources at the ground level etc.

Further, data strategy and data driven outcomes need synergies at departmental/organizational level and hence must be measured at the departmental/organizational level. Hence, it may not be very useful for a single division to include all themes within these two pillars in their self-assessment. While the division must have its own data unit and action plan to improve its system after baseline measurement, the data unit may need to be significantly smaller given the smaller scope. Within data driven outcomes, synergistic data use and inter-agency collaboration may need suitable customization to only cover those steps which can be undertaken by a division. For instance, a division must exchange its data with other divisions via APIs and

undertake public-private partnerships; however, it cannot set up department level data lakes on its own.

A summary of the use cases along with required customizations to and benefits of the DGQI toolkit is provided below in Table 5.

User Type	Use Case	How to use DGQI toolkit	Benefits of using DGQI toolkit	Impact
Planning Deptts / Apex M&E Deptts of Centre / States	Need for synergistic data use to track outcomes and make strategic planning decisions (for meeting NDA / SDG goals)	Customize to include if Deptts collect data on all outputs and outcomes required for decision making	• Peer learning and inter Deptt. collaboration • Easier-one-stop access to data allowing for better evidence-based policymaking	• Data driven decision making embedded in the DNA of governance • Increase in efficiency and effectiveness of public administration • Improved government to citizen engagement • Agile, quicker and responsive public
IT Deptt of Centre / States	Need to ensure harmonious compliance against e-governance or data policies/ guidelines	Customize on the basis of existing policies/ guidelines where compliance needs to be ensured	• Faster adoption of existing policies and guidelines across Deptts	
Finance Deptt of Centre / States	Need for high quality reliable data to inform budgeting decisions (for outcomes based / performance budgeting by utilizing OOMF like farmeworks)	Customize to check if Deptts collect data on all output and outcome indicators required for budgeting purposes, and with what quality	• Demonstration of accountability and transparency in govt budgeting processes and expenditure	
Any other Deptt of Centre / States / PSUs with multiple divisions	Need to ensure high quality de-siloed data systems for sector / department wide goal tracking	Customize to cover non-schematic interventions undertaken by the Deptt and to cover ERP systems for business operations in case of PSUs	• Intra-department collaboration • Better MIS designs allowing for integrated data use within the department • Easier-one-stop access to sector-wide data allowing for	

			better evidence based policymaking	service delivery to citizens • Identification of capacity gap allowing for eventual development of adequate data analytical capacities in governments
Local bodies / Distt administrations	Need to collect high quality data for upward reporting while simultaneously being able to make use of this data to improve implementation of schemes and delivery of services to citizens at the last mile	Customize to add what kind of data is being collected vis-à-vis what kind of data is required for day-to-day decision making; Customize to include non-schematic interventions; Customize to cover how can this data be also used by local bodies for improving public service delivery/ scheme implementation at the last mile	• Ensuring good quality data collection at entry • Optimizing data reporting burden by streamlining data collection requirements and processes	
Any divisions/ agencies implementing a single scheme/ intervention	Need for timely and reliable data for day-to-day decision making, implementation and monitoring of the intervention to make mid-course corrections	Customize to cover non-schematic intervention undertaken by the division; Customize to cover how can this data be used by the division for decision making such as fund releases, process management etc.	• Better MIS design • Ensuring good quality data at entry • Optimizing data reporting burden • Better performance of the scheme/ intervention	

Table 5 Use Cases of DGQI Toolkit

2. **Adopt & Measure**: Once the questionnaire and methodology is finalized, begin with the self-assessment exercise with participating agencies. It is recommended that the users begin with only covering data systems pillar in the first round of self-assessment, which can then be used to define data strategy reform actions and then include data strategy and outcomes in the next round. The questionnaire can be canvassed on an online portal with dedicated credentials to ensure skipping patterns are adequately followed and all questions are responded to. Active participation from all relevant stakeholders who need to fill up the self-assessment questionnaire would need to be focused upon, which may require top-down push approaches as well as bottom-up capacity building initiatives. It is equally crucial to ensure that the participating agencies are well versed with the questionnaire and can easily understand the same. Training resources and videos in multilingual formats are encouraged to be produced for this purpose. Use the responses provided in this first round of measurement to generate baseline scores.
3. **Identify & Reform**: Use the baseline measurement to identify specific reform areas for each participating agency with respect to reaching the broader common vision of improved data-driven outcomes. To drive structural reforms, participating agencies can be encouraged to devise a detailed action plan to identify and improve upon their weak areas and set up an institutional architecture to implement this action plan. This institutional setup can also be made responsible for capacity development of the respective participating agency so that data governance is ingrained in the culture of the organization in the long run. The commissioning agency can provide indicative guidance to participating agencies on how to set up these processes and institutions. They must also develop a mechanism to regularly track the implementation of action plans made by participating agencies via the same digital platform used for the self-assessment questionnaire to allow for continuous momentum. Independent third-party audit of responses and action plans may also be undertaken to ensure veracity of responses.
4. **Reward & Upscale**: Embed the compliance against the timely implementation of the reform actions within the measurement and scoring process next round onwards by increasing the scope of the self-assessment questionnaire to include strategy and outcomes also. Undertake these exhaustive measurements on a regular basis and have regular reviews with participating agencies to discuss their performance. This way, institutionalize the thrust towards the exercise with regular progress reviews and incentivizing top performers in varied manners.

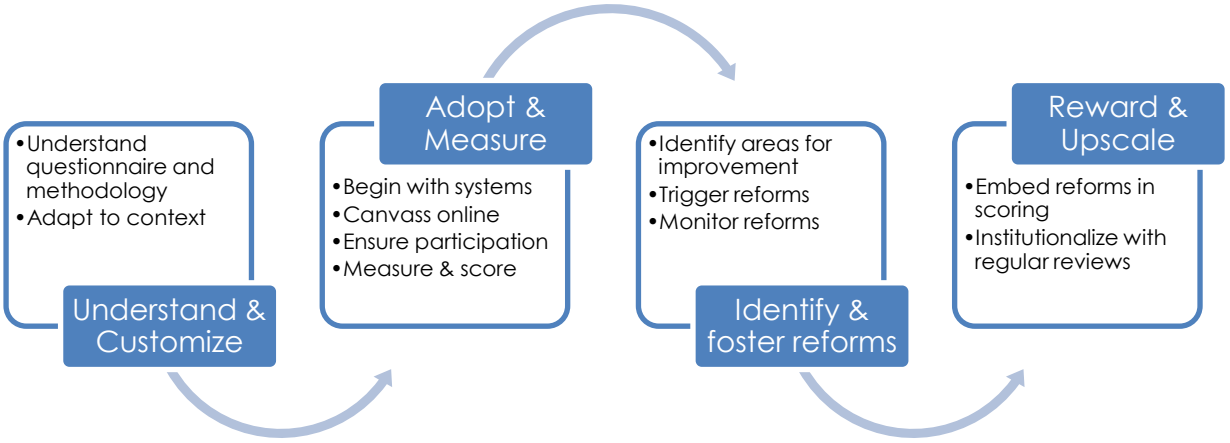

Figure 3 How to use DGQI toolkit

Some Do's and Don'ts while adopting the framework have been outlined below for the benefit of any user wishing to use this toolkit:

Do's

- Editing questionnaire and methodology to fit appropriate scenario by the user
- Regular training sessions with participating agencies
- Transparent sharing of questionnaire and methodology with participating agencies
- Online mode of conducting the entire exercise
- Focus on self-assessment of data preparedness rather than on external evaluation
- Action oriented exercise rather than sole focus on measurement
- Regular tracking of reform actions triggered by the exercise
- Embedding reforms in scoring to ensure reform momentum
- Focus on self-improvement than comparison; with built-in incentives for good performance

Don'ts

- As-is use of the framework without checking for customization
- Inadequate training to respondents
- Opaque methodology not shared with participating agencies
- Sole focus on external evaluation than on self-assessment

- Sole focus on measurement without reforms
- Changes in submitted responses for a round
- Drawing unnecessary comparisons between incomparable agencies or interventions

To summarize, the toolkit can be easily used by any government agency for conducting their own self-assessment of their present data preparedness levels and trigger necessary reforms in the direction be leveraging this framework. As mentioned earlier, throughout the above four steps that need to be followed to use this toolkit, there must be a unique focus on close partnership with participating agencies with a focus on self-reflection rather than on comparisons. Technology must be actively leveraged to conduct the exercise in a seamless manner. Regular handholding support must be provided by the commissioning agency via continuous engagement, webinars, training resources and guiding documents. Finally, the structural measure-and-reform approach must be the focal point for ushering in new wave of digital transformation across the government agencies.

8 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 who designed the DGQI toolkit can be reached at *dgqi-dmeo-niti@gov.in*.

Annexure 1: Indicative outline of action plan for Ministries/Departments

Better data preparedness would help improve the monitoring and consequently the outcomes of the schemes and interventions of Ministries/Departments. An indicative outline of roadmap to improve data preparedness and improve DGQI scores is given below. Ministries /Departments may use the indicative outline to have discussions to develop a roadmap to achieve higher, systemic, ministry-wide data preparedness levels and make provisions for human and financial resources for using technology and analytics to improve service delivery.

1. **Background**

   1.1. Brief overview of the M/D's business allocation, roles and responsibilities *(in 2-3 lines)*

   1.2. Current degree of digitization of administrative data systems in the M/D *(in brief)*

      1.2.1. Scheme-wise initiatives for digitization (for CS/CSS schemes)

      1.2.2. Other initiatives for digitization (Other Central Sector Schemes (OCS)/Other MIS/Dashboards)

   1.3. Ministry/Department's reflections on DGQI Scores 2020 *(in one page)*

      1.3.1. Major takeaways from the exercise

      1.3.2. Areas for improvement based on previous performance and DGQI methodology

      1.3.3. Limitations in achieving DGQI frontier scores (Dependence on states or other executing agencies for execution, human resource/financial constraints etc.)

2. **Vision, Mission & Objectives**

   2.1. Vision statement for achieving data driven decision making within the Ministry/ Department *(in 2-3 lines)*

   2.2. Mission statement for achieving DGQI frontier scores and going beyond DGQI by 2022 *(in 4-5 lines)*

   2.3. Objectives of the roadmap to achieve DGQI frontier scores

   (in half page – explaining Ministry/Department specific goals to achieve high levels of data preparedness; for instance, ensuring end-to-end digitization for high-quality, near real-time data generation across all schemes at project/beneficiary level, ensuring user-friendly MIS and dashboard systems for all scheme and non-schematic interventions, establishing an administrative system for human capability and technological development to enable data driven policy making etc.)

3. **Strategy to achieve DGQI Frontier Scores**

   3.1. Scope of the strategy *(in 1-2 pages)*

3.1.1. Schemes to be covered under the road map along with their contribution to overall M/D scheme budget (Encouraged to include all CS/CSS schemes of the M/D including schemes executed by other partnering agencies/states/PSUs aligned with the M/D)

3.1.2. Non-schematic interventions to be covered under the roadmap (Other MIS/Dashboards of the M/D which are not related to schemes; for instance, a sector-level MIS/dashboard used for monitoring the overall sector performance, separate MIS/dashboards for PSUs/Other Central Expenditure/any other purposes. Kindly note that administrative interventions for digitization within the office such as E-Office is not to be included here as it is outside the scope of the strategy. This strategy aims to implement digitization to improve monitoring and accountability of government expenditure on schemes and policies.)

3.2. Overall Approach *(in 1-2 pages)*

3.2.1. Principles to be followed while developing the roadmap (For instance, accuracy in information, relevance/utility to the strategy, transparency in processes, privacy of personal information, openness in disseminating non-personal information, inclusiveness in digitization, interoperability, integration of uses, etc.)

3.2.2. Integrated approach (Outline of an integrated and well-coordinated approach to be taken by the M/D to improve digitization across the board. The approach should target end-to-end digitization of all levels of information – Scheme level MIS/Dashboards, M/D Sector level and finally linking it to digitization of necessary information needed for achieving SDG goals/national priorities relevant to the M/D. Similarly, how data collection frequency, quality and timeliness-at-entry will be ensured on the field and during subsequent stages of data flow at the district and national levels. Also, the approach should focus on across the board interventions – Capacity development at M/D, technological overhaul at M/D, coordination between various divisions of the M/D, setting up of administrative systems at M/D to lead the effort, carrot-stick approaches to improve uptake etc.)

3.3. Scheme-wise Strategy *(2-3 pages per scheme)*

3.3.1. Scheme 1 (A short assessment of current system to be provided along with areas identified for improvement. Subsequently, the strategy should entail detailed steps to improve on each theme of the DGQI as shown below.)

3.3.1.1. Data Generation Strategy (Should cover steps for identifying data requirements of the scheme to have data on all relevant inputs, outputs and outcomes of the scheme; increasing granularity (beneficiary/project level) and frequency (near real-time) of digitization using latest sources of information; use of location

tracking devices for data collection; using GIS mapping/geo-coding/geo-fencing/mobile devices for data generation)

3.3.1.2. Data Quality Strategy (Should cover steps for ensuring rigorous data quality protocols for profiling/filtering incoming data, ensuring deduplication and redundancy removal within data, enforcement of data integrity, use of metadata standards for proper classification of data; use of mobile phones or other technologies for data quality control such as multimedia evidence, telephonic surveys etc.)

3.3.1.3. Use of technology Strategy (Should cover steps for linking M/D MIS/data systems with other platforms such as PFMS for finances and JAM trinity for beneficiary-oriented schemes; use of alternative data sources to complement M/D data such as private sector or GIS data; use of emerging technologies to improve scheme processes/delivery such as Machine Learning, Artificial Intelligence, IoT etc.)

3.3.1.4. Data Analysis, Use & Dissemination Strategy (Should cover steps for improving use of data by M/D to use it for policy making purposes; dissemination of data via websites/dashboards/social media/mobile apps; user-friendly visualizations; multilingual interfaces and compatibility features for differently abled etc.)

3.3.1.5. Data Security & HR Capacity Strategy (Should cover steps for improving data security, compliance requirements and privacy; capacity development for developing data analytics capabilities in the M/D to improve use of data in policymaking etc.)

3.3.1.6. Data Management Strategy (should cover steps for managing data across various stages right from generation to its use; devising strategies for integrated data storage and data disposal; ways and means of dealing with personal data using techniques like encryption, de-identification, etc., ensuring proper data classification using good-quality meta data to enable better reporting, analytics, and use; fixing accountability for data management by fixing intra-ministry and inter-ministry data ownership and other responsibilities for dissemination and use of data)

3.3.2. Scheme 2 and so on.. (Strategy for each scheme under the purview as per section 3.1 to be framed and the strategy should entail detailed steps to improve on each theme of the DGQI as shown in section 3.3.1.)

3.4. Non-schematic Strategy *(2-3 pages per intervention)*

3.4.1. Intervention 1 (A short description of the purpose and scope of the intervention to be provided with areas identified for improvement. Subsequently, the strategy should entail detailed steps to improvise on each theme of the DGQI as shown in section 3.3.1.)

3.4.2. Intervention 2 and so on..(Strategy for each intervention under the purview as per section 3.1 to be framed and the strategy should entail detailed steps to improvise on each theme of the DGQI as shown in section 3.3.1.)

3.5. Operational Execution Plan *(After strategy is formed, execution plan to be laid down for institutional development).*

3.5.1. Organizational Structure – Breaking the silos (To have a central unit leading the efforts to build, implement and revise the roadmap, it is recommended that a Data and Strategy Unit is established within the M/D and is placed directly under the Secretary. After setting up the unit, strategy for intra-ministerial coordination to be framed to ensure that the unit is able to work in conjunction with other scheme divisions and NIC.)

3.5.2. Human Resource Capacity Development (Should include steps for in-house capacity building to develop IT and data analytical capabilities, acquaint them with new tools/techniques, hire technical experts as per requirements if necessary, spread awareness about evidence-based policy making etc.)

3.5.3. Technological Development (Should include steps for overhaul of IT hardware and software systems in line with identified data generation, storage, management, and analytical needs including a procurement plan, development of data warehouses/ open data websites to create integrable data sources, creation of singular metadata standard/data classification norms to be followed across the M/D to create integrable datasets etc.)

3.5.4. Partnerships (Should include the nature of partnerships being planned with private sector or research organizations for developing capabilities, scope the possible partner landscape and areas of engagement, inter-ministerial coordination for synergies in data collection on common indicators, state-level engagements to help build adequate data systems at state level including CSS schemes)

3.5.5. Resource Allocation (Should include assessment of required financial resources to implement the roadmap and plans to make provisions for the same in scheme and M/D budget in the next EFC/SFC/Budget cycle; assessment of human resources to be deployed to implement the roadmap and provisions for the same; any other resources)

3.6. Consolidated roadmap *(Consolidated plan to be provided for all schemes and interventions listed in Section 3.3. and 3.4. as well as steps to be undertaken for institutional development in Section 3.5 with quarterly timelines against key strategy steps)*

Annexure 2: Indicative Terms of Reference for Data & Strategy Unit at Ministries/Departments

Purpose of DSU

In order to create better mechanisms for digitization of processes related to implementation and monitoring of Central Sector/Centrally Sponsored Schemes and other non-schematic interventions of Ministries/Departments, an institutional mechanism in the form of a **"Data and Strategy Unit"** may be set up within each Ministry/Department. The Data and Strategy Unit shall drive the process of building and harnessing existing as well as augmenting the monitoring, statistical, technological and data analytics capabilities of the respective Ministry/ Department.

The key roles of the DSU shall include breaking silos within the Ministry/Department to enable creation of well-integrated monitoring and data systems while ensuring adequate focus on data quality and security and creating mechanisms for regular data analysis within the Ministry/Department to inform policy decisions. Coordinating with scheme divisions within the Ministry/Department as well as with required external partners such as States, other Ministries/Departments, research organizations, leading private players and academic institutions for taking necessary steps in the direction shall also be one of their key responsibilities.

Organization Structure of DSU

To fulfill this purpose, The Ministries/Departments can augment their present institutional setup to create DSU.

The DSU may be headed by an Additional Secretary/Joint Secretary/DDG level officer who would be directly reporting to the Secretary of the Ministry/Department. As shown below, it is proposed to have the following four verticals within the DSU

1. Monitoring Unit – For integrating siloed monitoring initiatives across the Ministry/Department
2. Statistics Unit – For identifying overall statistical needs of the Ministry/Department and ensuring coordination with necessary agencies to meet the same
3. Technology Unit – For ensuring 100% digitization and integrating siloed MIS/dashboards/data systems of the Ministry/Department
4. Data Analytics Unit – For undertaking and promoting data analysis on collected data to drive decisions

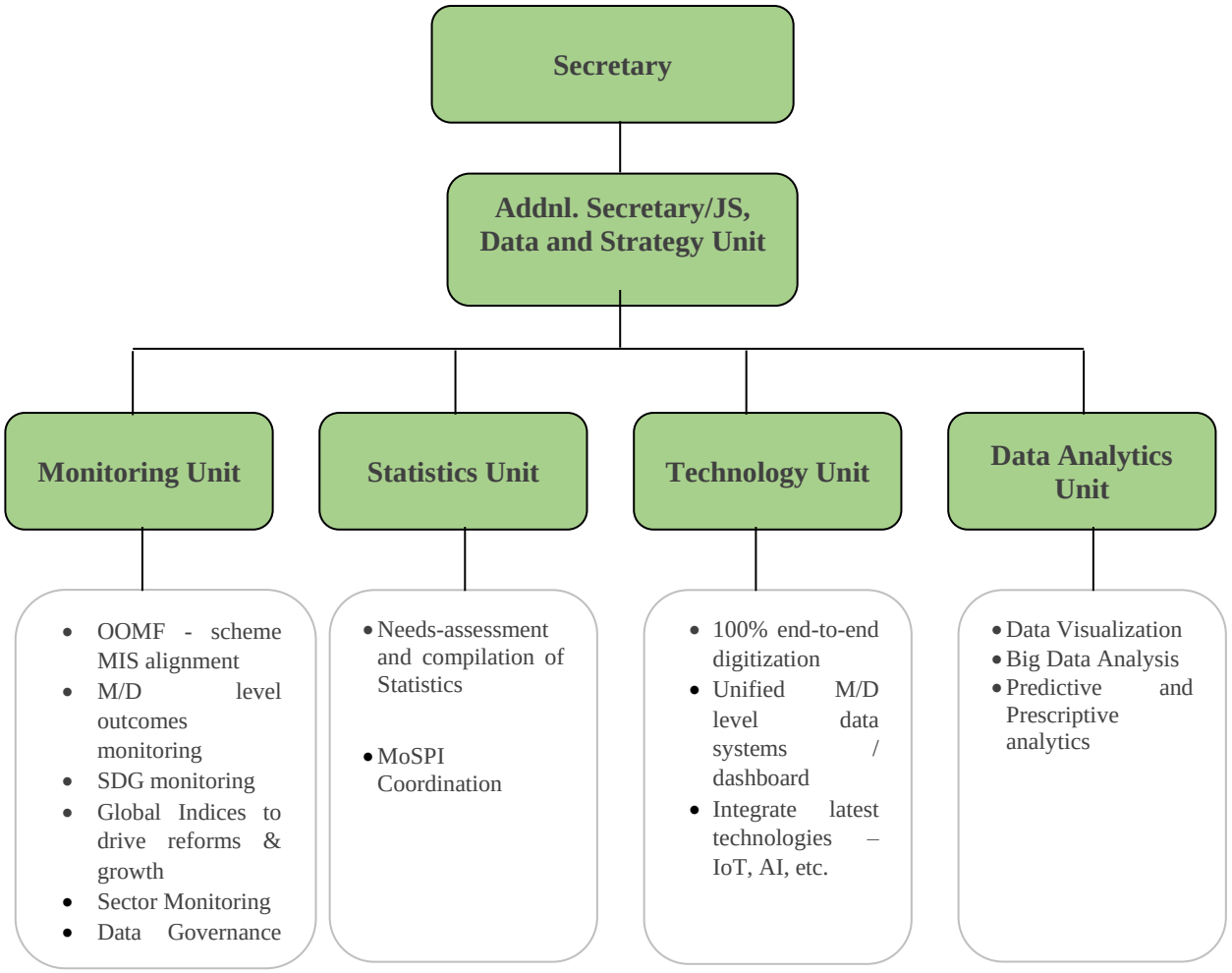

Figure 4 Data & Strategy Unit at Ministries/Departments

A snapshot of the organization structure of the four verticals of the DSU has been provided below in next figure.

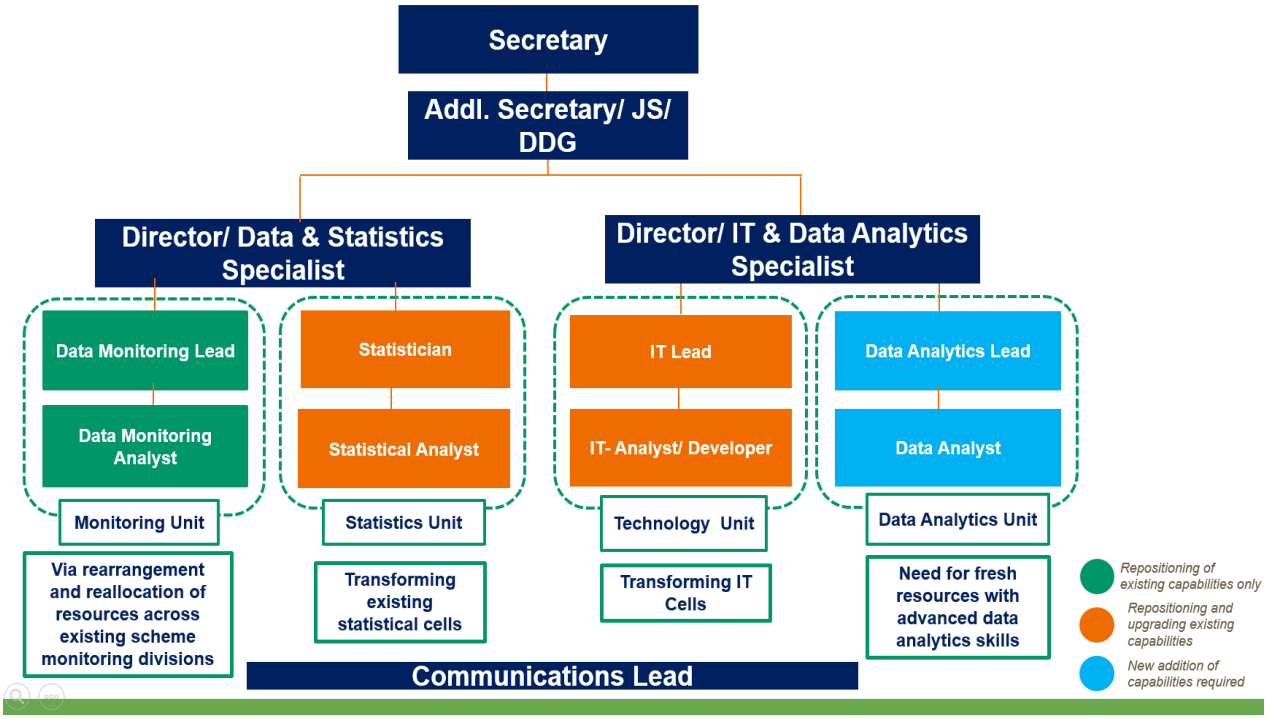

Figure 5 Organizational Structure of Data & Strategy Units at M/Ds

The sub-units within the DSU should be headed by a Director level officer. In order to make the DSU lean and more responsive, it is suggested that sub-units with similar tasks may be headed by one Director level officer i.e., one Director for Monitoring and Statistics sub-units and one Director for Technology and Data Analytic sub-units.

Each sub-unit is recommended to be composed of leads who would be domain experts and provide direction to the efforts to be undertaken and skilled analysts who would have domain knowledge and be responsible for carrying out the implementation efforts.

It is envisioned and suggested that majority of these positions in these sub-units can be filled by repositioning of existing staff as explained below.

It is suggested that the Monitoring sub-unit may be almost entirely created by reallocating some of the existing staff among different scheme divisions which are already regularly monitoring scheme implementation to DSU. Similarly, current resources engaged in OOMF, Global Indices and SDG monitoring in Ministries/Departments can be integrated to form the Monitoring sub-unit of the DSU. One, this would aid in ensuring that resources from different scheme divisions work in tandem with each other, helping in breaking the siloes as they get to understand how different schemes may be monitoring similar initiatives. Two, with different types of monitoring needs (OOMF, GI, SDG etc.) being looked at by a central team, the Ministry/Department would be able to understand how these efforts can actually be synergized, reducing the administrative reporting burden on Ministries/Departments.

Similarly, Statistics sub-unit may also be largely created by reallocating roles in the existing statistical cells to the DSU. The statistical cells at present are already responsible for dissemination and compilation of Ministry/Department statistics. To be able to meet newer expectations from the Statistics sub-unit of the DSU such as needs assessment for more

statistics that need/need not to be collected and coordination with other partners (private agencies/Ministries/others) for synergistically collecting some statistics, their capabilities may require some upgradation. In order to transform existing statistical cells for this purpose, every Ministry/Department may undertake an assessment at their own level and consider the need for a few additional resources with required skills based on their present status.

The Technology sub-unit may similarly be created by transforming existing IT cells present in the Ministry/Department. Some of the resources can be re-casted to play the role of the technology arm of the DSU. However, if it is felt that there may be a need for upgradation of their capabilities in certain domains such as integration of several dashboards or creating single metadata architecture for all scheme MIS/dashboards, then the Ministry/Department can again undertake an assessment at their own level and consider the need for a few additional resources with required skills based on their present status. For e.g. If the existing IT cell in the department does not have a suitable person with required skills of IT analyst, he/she may have to be recruited as a lateral entrant.

The Data Analytics sub-unit is one sub-unit where it is believed that most Ministries/Departments may not have enough capabilities or resources at present and hence may require people with advanced data science and analytics skills to be freshly recruited to complete this missing link. This would play a crucial role in completing the vision of moving to evidence-based policymaking across Ministries/Departments.

To summarize, it is suggested that the creation of DSU may be principled on reorganization of existing organizational structure and roles of Ministries/Departments. The key idea should be to bring together existing resources with skills, experience and passion for these tasks together within the DSU to break the siloes of the present structure, and thereafter, only for required roles, recruitment may be done to fill the skill gaps, wherever necessary. The same is also depicted above in above figure where monitoring sub-unit is colored green to show that it only requires repositioning, statistics and technology sub-units are colored amber to show that they majorly need repositioning but some Ministries/Departments may also require upgradation and finally, data analytics unit is colored blue to show that it is the major place where fresh talent acquisition may be needed.

Indicative strength of DSU

This section highlights the indicative strength of manpower that may be required for the 'Data and Strategy Unit'. As already mentioned above, the unit is proposed to be headed by an Additional Secretary/ Joint Secretary level officer of the M/D, reporting directly to the Secretary. A Director level officer called Data & Statistics Specialist can head the two sub-units, Monitoring Unit and Statistics Unit. Another Director level officer called IT & Data Analytics Specialist can head the remaining two sub-units, Technology Unit and Data Analytics Unit.

For leads and analysts, indicative strength of manpower required in the Data & Strategy Unit has been arrived at in below table. The M/Ds under have been classified into three main categories: Small, Medium and Large on the basis of the number of interventions (CS schemes +CSS schemes +Non-Schematic Interventions of the M/D). M/Ds with upto 10 interventions are called small, 11-30 interventions are termed medium and M/Ds with above 30 interventions are termed large. Further, two bifurcations have been created on the basis of average outlay of

M/D's interventions (for the size of interventions), depending on whether it is above or below INR 500 Crores. The same has been done keeping in mind that manpower requirements would rise in line with a greater number of interventions or increase in average outlay.

Using this classification and the following general thumb rules –

- For every **5 interventions** with average budgetary allocation **less than Rs 500 crores**, **one analyst is recommended**. M/D may hire/ allocate **one lead for every two analysts** in a sub-unit to guide and review the tasks assigned to them.
- For every **5 interventions** with average budgetary allocation **more than Rs 500 crores**, **two analysts are recommended**. M/D may hire/ allocate **one lead for every two analysts** in a sub-unit to guide and review the tasks assigned to them.

The indicative manpower strength has been arrived at in Table 6. However, it may be noted that this is only meant to act as a guidance for Ministries/Departments is by no means a mandatory requirement.

Type of M/D	Total number of interventions (CS+CSS+NSI)	Avg outlay > = INR 500 Crores			Avg outlay < INR 500 Crores		
		Leads	Analysts	Total	Leads	Analysts	Total
Small	0-10	2	4	**6**	1	2	**3**
Medium	11-30	4	8	**12**	2	4	**6**
Large	Above 30	6	12	**18**	3	6	**9**

Table 6 Indicative strength of Data & Strategy Unit at M/Ds

**Individual M/Ds may modify the numbers as per different combinations/ categories and beneficiary coverage*

Annexure 3: Reference Data Maturity Models

While developing the methodology for DGQI, DMEO had reviewed several existing frameworks for assessing data preparedness of organizations. Four data maturity models spanning both private and public context were shortlisted for a detailed study based on their relevance, exhaustiveness and representativeness: US Federal Government Data Maturity Model, Data Governance Maturity Model (IBM), Data Maturity Assessment Framework (SCM) and Data Maturity Management Model (CMMI). Based on the assessment of these models, three key pillars of data preparedness were identified viz., Data Strategy, Data Systems and Data Outcomes and this theory of change formed the basis for design of DGQI.

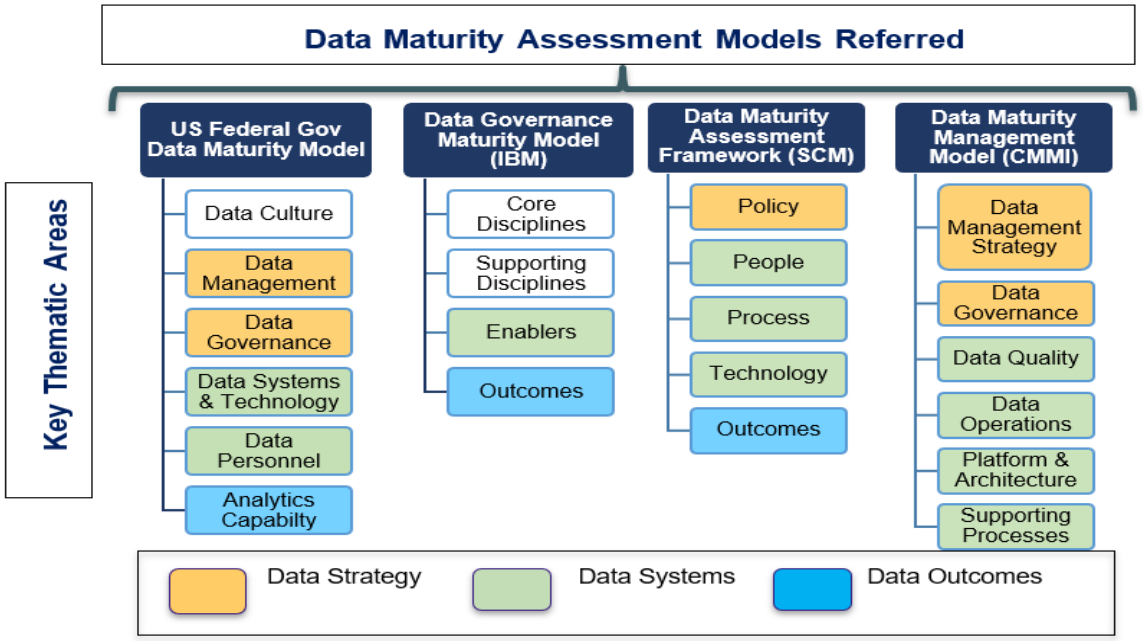

Figure 6 Reference Data Maturity Models

Annexure 4: DGQI 2.0 Self-Assessment Questionnaire

Detailed explanations for each question may be referred to after the questionnaire.

A. Background Information

1.	Ministry / Department(M/D) Name:	
2.	Name of the Central Sector (CS) Schemes of the M/D:	a.
		b.
		c.
3.	Name of Centrally Sponsored Schemes (CSS) of the M/D:	a.
		b.
		c.
4.	Please enter any other non-schematic intervention (NSI) to be included for DGQI self-assessment:	a.
		b.
5.	Details of the nodal officer responsible for verifying authenticity of information provided in this form:	
	a. Name	
	b. Designation:	
	c. E-Mail ID:	

B. Data & Strategy Unit

1.	Has the M/D constituted a Data & Strategy Unit (DSU) as a central unit for developing data strategy? (as per the D.O. letter from Sh. Bhaskar Khulbe, Advisor to PM dated 02.02.2021)	☐ Yes ☐ No
2.	*(Respond if answer to 1 is 'yes', else skip to Q1 of next section)* Who is the head of the DSU?	☐ AS and equivalent ☐ JS and equivalent ☐ Director and equivalent ☐ Below Director
3.	Please select the verticals established under the DSU of your Ministry/Department. (as per the D.O. letter from Sh. Bhaskar Khulbe, Advisor to PM dated 02.02.2021)	☐ Monitoring Unit ☐ Statistics Unit ☐ Technology Unit ☐ Analytics Unit

4.	Please provide the percentage of filled posts in DSU (number of posts filled up/ number of posts created by the Ministry/Department for the DSU) in the below provided table: **Enter % of posts filled up** Monitoring Unit Statistics Unit Technology Unit Analytics Unit Total	
5.	Is the terms of reference (ToR) for all units within DSU well defined and documented by the M/D to lay down their scope of work?	☐ Yes ☐ No ☐ Partial (to be selected if ToR development is in progress)
6.	Have any regular review meeting mechanisms at the level of the head of DSU and/or the Secretary been established for regular review of the work undertaken by the DSU (including implementation of action plan)?	☐ Yes ☐ No
7.	*(Respond if answer to 6 is yes, else skip this question)* What is the frequency of regular review meetings/review reports?	☐ Daily ☐ Weekly ☐ Fortnightly ☐ Monthly ☐ Quarterly ☐ Annually

C. Action Plan

1.	Has the M/D framed an action plan to improve its data preparedness levels? (as per the D.O. letter from Sh. Bhaskar Khulbe, Advisor to PM dated 02.02.2021)	☐ Yes ☐ No

2.	*(Respond if answer to 1 is yes, else skip to Q1 of next section)* Does the action plan have all the sections as per the outline shared with all M/Ds? (As per D.O. letter from Sh. Bhaskar Khulbe, Advisor to PM on 02.02.2021)	☐ Yes ☐ No ☐ Partially. If partially , please specify how: _____
3.	Does the action plan include data strategy for all CS/CSS schemes of the M/D?	☐ Yes ☐ No ☐ Partially (Some schemes included) If partially, please specify which schemes are not included: _____
4.	Are clear timelines for each action point identified under the strategy?	☐ Yes ☐ No ☐ Partially (For some actions) If partially, please specify how and why: _____
5.	Are the responsibilities for each action point clearly allocated to respective divisions for ensuring accountability?	☐ Yes ☐ No ☐ Partially (For some actions). If partially, please specify how and why: _____
6.	Please upload the action plan in PDF format.	
7.	Please enter action points along with date of completion and current status. Scores based on timely completion/compliance on the action points against the timelines set by the M/D will get auto-calculated and displayed here.	

D. Data Management

1.	Does the M/D have data management guidelines/architecture, explaining how generated data is to be processed, stored, exchanged, archived and destroyed?	☐ Yes ☐ No If yes, please briefly explain the scope implementation of these guidelines: _____
2.	*(Respond if answer to in 1 is 'yes', else skip this question)* Is there a dedicated senior-level officer responsible to check the compliance of the data management processes?	☐ Yes ☐ No
3.	Are data ownership norms clearly defined by the M/D?	☐ Yes ☐ No
4.	Is there a framework for assessing the risk and value of all the data collected by the M/D?	☐ Yes ☐ No

		If yes, please explain how is this done: _____
5.	Is there a framework governing the ethical use of data, including the use of predictive algorithms, machine learning etc. by the M/D?	☐ Yes ☐ No If yes, please explain how is this done: _____

Note: M/Ds may preferably fill up remaining sections of Part- A (given below from E-H) after completing Part – B of the questionnaire as these questions correspond to the third pillar of data driven outcomes.

E. Synergistic data use within the M/D

1.	Based on data analysis, has the M/D identified data gaps at M/D level that need to be plugged in from decision making/policy analysis perspectives?	☐ Yes ☐ No If yes, please specify how: _____
2.	*(Respond if answer to 1 is yes, else skip this question)* Has the M/D made any implementation plan to overcome these data gaps to aid in decision making?	☐ Yes ☐ No If yes, please specify how: _____
3.	Has the M/D created any systems for ensuring that data systems across scheme divisions are integrated so that data from different scheme divisions is shared with each other?	☐ Yes. ☐ No ☐ In progress ☐ N/A If yes or in progress, please specify how: _____ If "N/A", please provide reasons why inter schematic division data integration is not applicable: _____

F. Inter-Agency Data Collaboration

1.	Has the M/D collaborated with other agencies (other M/Ds, private agencies, research organizations etc.) for improving their data systems wherever possible?	☐ Yes ☐ No ☐In progress
2.	*(Respond if answer to 1 is yes, else skip this question)* Has the M/D undertaken any of the following steps to drive these inter-agency data collaboration initiatives?	
	☐ SoI, MoU, Partnerships with agencies	
	☐ API linking of MIS/Dashboards done to enable seamless data sharing between M/Ds	

	☐ Multiple data collection processes aimed at same target groups replaced by single synergistic process
	☐ Integrated data storage/warehouses
	☐ Collaboration with other M/Ds to use their data for developing own systems
	☐ Collaboration with M/Ds to develop joint systems for data gathering/use of non-conventional data sources/emerging technologies
	☐ Collaboration with private agencies for use of non-conventional data sources or emerging technologies
	☐ Jointly conducting analysis using data from multiple M/Ds
	☐ Partnerships/Collaborations for data security related measures
	☐ Partnerships/Collaborations for capacity building of human resources
	☐ Others - Please specify : _____

G. Prescriptive Analytics

1.	Has the M/D gone beyond exploratory data analysis to cross-functional prescriptive analytics?	☐ Yes ☐ In Progress ☐ No If yes or in progress, please specify how: _____
2.	*(Respond if answer to 1 is yes, else skip this question)* How often is this being undertaken?	☐ Annually ☐ Quarterly ☐ Monthly
3.	*(Respond if answer to 1 is yes, else skip this question)* What is the mode in which this is being practiced? (Multiselect)	
	☐ Mechanisms for regular prescriptive data analysis reports to be prepared and shared with decision makers at the highest level have been instated	
	☐ Committee formed to hold policy review meetings/review reports at regular frequencies	
	☐ Regular policy review meetings involving all scheme divisions/sections institutionalized	
	☐ Emerging actionables are undertaken, documented and disseminated via a separate newsletter/report/document/order etc. and tracked regularly	
	☐ Others - Please specify how: _____	

H. Good Practices - Please share any three good practices of how the M/D has taken measures to strengthen data-driven decision-making (non-schematic or scheme level) within the M/D along with its positive impact.

Good Practice 1
1a. Describe the problem statement faced by the M/D. (100 words)
1b. Describe how the M/D has used and implemented data systems and analytics to address the issue to drive smart, near real-time and granular decisions (100 words).

1c. Explain the positive impact generated with supporting evidence that indicated such impact due to the solution implemented (100 words).
Good Practice 2
1a. Describe the problem statement faced by the M/D. (100 words)
1b. Describe how the M/D has used and implemented data systems and analytics to address the issue to drive smart, near real-time and granular decisions (100 words).
1c. Explain the positive impact generated with supporting evidence that indicated such impact due to the solution implemented (100 words).
Good Practice 3
1a. Describe the problem statement faced by the M/D. (100 words)
1b. Describe how the M/D has used and implemented data systems and analytics to address the issue to drive smart, near real-time and granular decisions (100 words).
1c. Explain the positive impact generated with supporting evidence that indicated such impact due to the solution implemented (100 words).

Part -B (To be fed at CS/CSS/NSI Level)
To be fed by the Ministry/Department for each CS/CSS/NSI of the Department in Q.A of Part A of the Questionnaire

A. Data Generation

1.	Are the data requirements of the scheme well defined and documented?	☐ Yes ☐ No
2.	Is data collected for all identified data requirements?	
	a. Input Data Points	☐ Yes ☐No ☐Partial
	b. Output Data Points	☐ Yes ☐No ☐ Partial
	c. Outcome Data Points	☐ Yes ☐No ☐Partial
3.	Is collected data reported digitally? (i.e. is there is a digital electronic database/MIS)?	☐ Yes ☐ No i.e. On paper only If Yes, please provide the link: _____ If credentials are required for login, please provide some username and password: User - _____ Pw - _____
4.	*(Respond if answer to 3 is 'Yes', else skip to 1 of Q1 of data quality section)* At what granularity is data reported digitally for the scheme?	
	a. At the M/D (National)	☐ Yes ☐ No ☐ NA
	b. State	☐ Yes ☐ No ☐ NA
	c. District / City	☐ Yes ☐ No ☐ NA
	d. Sub-District / Tehsil	☐ Yes ☐ No ☐ NA
	e. Block	☐ Yes ☐ No ☐ NA
	f. Village	☐ Yes ☐ No ☐ NA
	g. Individual / Household	☐ Yes ☐ No ☐ NA
	h. Facility	☐ Yes ☐ No ☐ NA
	i. Project	☐ Yes ☐ No ☐ NA
5.	At what frequency is data reported digitally for the scheme?	
	a. Realtime or near realtime	☐ Yes ☐ No ☐ NA
	b. Daily	☐ Yes ☐ No ☐ NA
	c. Weekly/Fortnightly	☐ Yes ☐ No ☐ NA
	d. Monthly	☐ Yes ☐ No ☐ NA
	e. Quarterly	☐ Yes ☐ No ☐ NA
	f. Half-yearly	☐ Yes ☐ No ☐ NA
	g. Yearly	☐ Yes ☐ No ☐ NA
6.	How is this data collected at the ground level?	
	☐ Collected on paper by human resources and then fed on digital systems	
	☐ Collected using digital modes (tablets/phones etc.) by human resources	
	☐ Transactional data	

7.	*(Respond if answer to 6 is 'second/third option', else skip this question)* Are any of the following technologies used?	
	a. CAPI Surveys	☐ Yes ☐ No ☐ NA
	b. Geotagged information	☐ Yes ☐ No ☐ NA
	c. Geofenced information	☐ Yes ☐ No ☐ NA
	d. Others - Please specify which technology	_____

B. Data Quality

1.	Are there pre-defined documented mechanisms to assess quality of incoming data?	☐ Yes ☐ No
2.	How is data quality assessment done? *(If answer to 2 of data generation section is 'No', please select 'Manually')*	☐ Automatically ☐ Manually ☐ Hybrid ☐ Not done If Hybrid, please specify how: _____
3.	*(Respond if answer to 2 is not "not done", else skip to Q1 of next section)* Are following protocols followed during data quality assessment?	
	a. Incoming data is filtered/cleaned after checking for missing values, logical flaws in data, incorrect values etc.	☐ Yes ☐No
	b. Summary statistics of incoming data are generated and checked for errors/abnormalities	☐ Yes ☐No
	c. Existence and accuracy of metadata for all the scheme's data is periodically checked (Schema is well defined)	☐ Yes ☐No
	d. There is a system for identifying duplicate data and removing redundancies	☐ Yes ☐No
	e. There is a system to ensure data is accurate, consistent and traceable to origin/source, whenever it is reproduced by any agency (data integrity)	☐ Yes ☐No
4.	Are following feedback mechanisms/backchecks also leveraged for data quality control?	
	a. Social audits	☐ Yes ☐ No ☐ NA
	b. Telephonic backchecks/verification with beneficiaries	☐ Yes ☐ No ☐ NA
	c. Multimedia data – citizen voice, video, images as evidence	☐ Yes ☐ No ☐ NA
	d. Sample inspections based on data	☐ Yes ☐ No ☐ NA
	e. Third party data verification/ data audits	☐ Yes ☐ No ☐ NA

C. Data Analysis, Use and Dissemination

1.	What types of data analysis is undertaken on collected data?	
	a. Descriptive data analysis (e.g. basic cross tabulation, frequency distribution, mean, median etc.)	☐ Yes ☐ No

	b. Exploratory data analysis (e.g. correlation etc.)	☐ Yes ☐ No
	c. Inferential data analysis (Using a small sample of data to infer about a larger population)	☐ Yes ☐ No
	d. Predictive analysis (Using historical or current data to find patterns to make predictions about the future)	☐ Yes ☐ No
	e. Causal analysis (Looks at the cause and effect of relationships between variables, focused on finding the cause of a correlation)	☐ Yes ☐ No
	f. Mechanistic Analysis (Understand exact changes in variables that lead to other changes in other variables)	☐ Yes ☐ No
	g. Others - Please specify the name and the type of data analysis - _____	
2.	*(Respond if answer to any of the options in 1 is "yes", else skip to Q5)* Is cross-schematic/sectoral data also analysed, wherever needed?	☐ Yes ☐ No ☐ NA. If NA, please explain why: _____ _____
3.	How often is this data analysis well-documented (in reports/notes/publications)?	☐ Real-Time on a dashboard ☐ Quarterly ☐ Half-yearly ☐ Annually ☐ Never
4.	How often is this data analysis being used by the M/D officials for:	
	a. To re-design the schemes or activities undertaken under the scheme at the end of the tenure?	☐ Yes ☐ No
	b. To do mid-course corrections through design or implementation changes ?	☐ Yes ☐ No
	c. To guide intra-scheme funding decisions like inter-state allocations, inter-component allocations, etc.?	☐ Yes ☐ No
	d. To guide inter-scheme budgetary allocations?	☐ Yes ☐ No
	e. To decide quarterly releases to implementing agencies?	☐ Yes ☐ No
	f. For fraud management and analysis	☐ Yes ☐ No
	g. Day to day delivery and monitoring of implementation/ performance of the scheme	☐ Yes ☐ No
5.	What other modes are used to disseminate the MIS/ paper-based data and related data analysis?	
	a. Dashboard	☐ Yes ☐ No ☐ NA
	b. Mobile App	☐ Yes ☐ No ☐ NA
	c. Social Media	☐ Yes ☐ No ☐ NA
	d. SMS	☐ Yes ☐ No ☐ NA
	e. Newspapers/ Magazines	☐ Yes ☐ No ☐ NA
	f. Outdoor media (signages/ billboards)	☐ Yes ☐ No ☐ NA
	g. Events	☐ Yes ☐ No ☐ NA

	h. TV/ Radio	☐ Yes ☐ No ☐ NA
	i. Others - Please mention the mode - _____	
6.	*(Respond if 'Yes' in 5a, else skip to 9)* What purposes are dashboards used for by the M/D officials?	
	a. Visual presentation of KPI/KRAs with drill-down capability to lowest level to gain total visibility	☐ Yes ☐ No
	b. Capturing trends over time and identifying preempt trends	☐ Yes ☐ No
	c. Measure efficiencies/inefficiencies in processes	☐ Yes ☐ No
	d. User friendly one stop access to multiple automated reports	☐ Yes ☐ No
7.	What types of Data Visualizations are used?	
	a. Bar chart/Histogram	☐ Yes ☐ No
	b. Pie charts	☐ Yes ☐ No
	c. Scatter plot	☐ Yes ☐ No
	d. Heat maps	☐ Yes ☐ No
	e. Treemaps	☐ Yes ☐ No
	f. Gantt chart	☐ Yes ☐No
	g. Specialized visualizations- Stripe graphics, streamgraph, etc.	☐ Yes ☐No
	h. Others - please mention data visualizations used - _____	
8.	Does the Dashboard visualize information on maps?	☐ Yes ☐ No ☐ NA
9.	*(Respond if 'Yes' in 3 of Data Generation section, else skip to Q6 of next section)* Does the information system (of MIS) support multilingual features as per GIGW norms?	☐ Yes ☐ No ☐ Partially (some norms followed but not all)
10.	Does the information system (of MIS) support features for differently abled as per GIGW norms?	☐ Yes ☐ No ☐ Partially (some norms followed but not all)
11.	How is the MIS data (non-sensitive data which can be shared) accessible for general population?	☐ Openly accessible without credentials ☐ Accessible through credentials ☐ Not accessible
12.	Is there an option on the MIS to download bulk data (non-sensitive data which can be shared) in excel, csv, dta files (machine readable formats)?	☐ Yes ☐ No ☐ Partial data download allowed
13.	Is the MIS data available on 'data.gov.in' (non-sensitive data which can be shared)?	☐ Yes ☐ No

D. Use of Technology

1.	*(Respond if 'Yes' in 3 of Data Generation section, else skip to Q4)* Does the MIS of the scheme have linkages with PFMS?	☐ Yes ☐ No ☐ NA

2.	*(Respond if answer to 1 is yes)* Is PFMS integration completed till the field-level implementation agency?	☐ Yes ☐ No ☐ Partially
3.	Does the MIS of the scheme have linkages:	
	a. Aadhaar	☐ Yes ☐ No ☐ NA
	b. Mobile numbers	☐ Yes ☐ No ☐ NA
	c. Bank accounts	☐ Yes ☐ No ☐ NA
	d. GSTN	☐ Yes ☐ No ☐ NA
	e. Udyog Aadhaar	☐ Yes ☐ No ☐ NA
	f. Others – please specify	
4.	Does the scheme use any of the following:	
	a. Remote sensing data	☐ Yes ☐ No ☐ NA
	b. Night light data	☐ Yes ☐ No ☐ NA
	c. Social media data	☐ Yes ☐ No ☐ NA
	d. Private sector generated data	☐ Yes ☐ No ☐ NA
	e. Others – please specify	
5.	*(Respond if 'Yes' in 3 of Data Generation section, else skip to Q6)* Is the MIS compliant with Local Govt Directory (LGD)?	☐ Yes ☐ No ☐ NA
6.	Does the scheme apply/use any of the following:	
	a. Machine Learning	☐ Yes ☐ No ☐ NA
	b. Artificial Intelligence	☐ Yes ☐ No ☐ NA
	c. Blockchain	☐ Yes ☐ No ☐ NA
	d. Internet of Things (IoT)	☐ Yes ☐ No ☐ NA
	e. Big Data analytics	☐ Yes ☐ No ☐ NA
	f. Drones	☐ Yes ☐ No ☐ NA

E. Data Security and HR Capacity

1.	*(Respond if 'Yes' in 3 of Data Generation section, else skip to Q10)* Does the MIS follow regular antivirus updates?	☐ Yes ☐ No
2.	Is the MIS regularly assessed by third party auditors for the online security?	☐ Yes ☐ No
3.	Does the MIS/ website uses SSL certificate?	☐ Yes ☐ No
4.	If "Yes" in previous question, is the SSL certificate at least 2048 bit SHA 256 encryption or higher?	☐ Yes ☐ No
5.	Does the MIS use firewalls to secure access to data?	☐ Yes ☐ No
6.	All external communication/ 3rd party integration/ API integration for the MIS is done through encrypted channel?	☐ Yes ☐ No ☐ No external communication established
7.	What measures are undertaken to secure sensitive/personally identifiable information? (Multiselect)	
	☐ Single-factor/ Multi-factor authentication	
	☐ Access control list is maintained	
	☐ Data is encrypted	
	☐ Data is anonymized	
	☐ No such data	

8.	(If anonymization is selected in previous question) How do you protect de-identified data from re-identification risks?	
	☐ No efforts made	
	☐ Tighter security for databases that store anonymized information	
	☐ Implementation of Differential Privacy	
	☐ Generation of Synthetic Data that exhibits the statistical properties of the raw data, without allowing real individuals to be identified	
	☐ Others - provide details – _____	
9.	*(Respond if answer to 7 is any option other than "no such data", else skip to Q10)* Is permission taken from user to collect, store and use their personal data?	☐ Yes ☐ No
10.	Is there a dedicated data quality assessment and management team for the scheme?	☐ Yes ☐ No
11.	Is there a dedicated data analysis team for the scheme?	☐ Yes ☐ No

F. Data Management

Respond if you have answered 'Yes' in 3 of Data generation section, else skip this section.		
1.	Where is MIS data stored?	☐ On separate servers for different schemes (distributed storage) ☐ On central server which is used for all schemes
2.	Are there mechanisms in place which can enable data sharing with other scheme divisions?	☐ Yes ☐ No If yes, please explain how: _____
3.	How is MIS data stored?	☐ Physical servers ☐ Cloud Storage ☐ Hybrid servers
4.	*(Respond if "Cloud Storage is selected in 3, else skip this question)* Which cloud service is being used?	☐NIC/ Gov cloud- Meghraj ☐ Cloud Services directly from CSP (Cloud Service Provider) (empanelled by MeiTY) / Cloud services through System Integrators (SI) after Standardisation Testing and Quality Certification / Cloud services through Managed Service Provider (MSP) after Standardisation Testing and Quality Certification ☐ Cloud Services from other CSPs which are not empanelled / from other MSPs or SIs which don't have Standardisation Testing and Quality Certification

5.	How is historical MIS data managed?	☐ Data is not backed up (i.e. it is destroyed) ☐ Data is backed up and data is archived ☐ Data history is well maintained including retention, destruction, and audit trail details

Detailed explanations to questions of DGQI Self-Assessment Questionnaire

Part	Section	Question	Explanation
A	A	1	M/D name would be automatically filled up when the M/D logins using their credentials.
A	A	2	A pre-populated list of CS schemes of the M/D would be visible here.
A	A	3	A pre-populated list of CS schemes of the M/D would be visible here.
A	A	4	M/Ds to enter any other non-schematic intervention such as sector dashboards, sector level MIS, any other MIS/dashboards etc. that they would like to include for DGQI assessment using the self-assessment questionnaire.
A	A	5	M/Ds to enter details of DGQI nodal officer. He/she would be assumed to have verified the correctness and authenticity of the information filled in this self-assessment form.
A	B	1	Constitution refers to establishing the unit, hence, even if its staffing is ongoing, M/Ds can select 'yes' if they have established the admin structure of the unit and some members have been assigned to it.
A	B	4	M/Ds to undertake calculations at their end based on how many posts they have proposed to create for the DSU based on their requirement and how many of these posts have been filled up. Total will be auto-calculated using values entered in the table.
A	B	5	Documentation of terms of reference here refers to the documentation of detailed objectives, roles and responsibilities of the DSU specific to the M/D. Indicative ToR for guiding M/Ds was shared by DMEO earlier.
A	B	6	M/Ds to select yes if guidelines for a standard system for regularly scheduling review meetings (via OM etc.) has been issued.
A	C	1	M/Ds to select yes if they have completed preparation of exhaustive action plan to improve data preparedness levels of the M/D.
A	C	2	Action plan is to have 3 sections with all sub-sections: 1. Background, 2. Vision, Mission & Objectives, 3. Strategy – Scope, Overall approach, scheme wise strategy, non-schematic strategy, operational execution plan.
A	C	3	M/Ds to select yes if the action plan has separate action points for all CS/CSS schemes of the M/D (as per the list on this portal).
A	C	4	M/Ds to select yes if every action point has a corresponding mm/yy timeline by which it is aimed to be completed, clearly documented in the action plan.
A	C	5	M/Ds to select yes if every action point is mapped to unit/personnel within the M/D by whom it is expected to be completed, clearly documented in the action plan.
A	D	1	Data management guidelines/architecture explains how data is to be managed across its lifecycle, i.e., how is it to be collected, stored, processed? How will it be exchanged? What will be done with historical data?

A	D	3	Data ownership norms would define who would be the owner of data when data is shared with other divisions or M/Ds or in public.
A	D	4	Understanding the value of the data collected by the M/D from utility perspectives and comparing it to the associated data security and privacy risks to ensure there is a balance between the two.
A	D	5	Data ethics refers to systemizing, defending, and recommending concepts of right and wrong conduct in relation to data, particularly personal data. With use of machine learning and predictive algorithms, it becomes even more important to protect sensitive data.
A	E	1	Data gaps refer to data that is required by the M/D from decision making point of view, however, for some reasons, such data is not available with the M/D. M/Ds to select yes if they have identified such data gaps based on analysis of their current data.
A	E	2	After identification of data gaps, M/Ds must take reform actions to develop data capture mechanisms/exchange mechanisms to fill up data gaps. M/Ds to select yes if they have started planning these actions.
A	E	3	For schemes with similar target groups, data collection can be done together rather than separately. This is an example of integrated data systems for collection. Similarly, if one scheme is collecting data on some indicator which is required by another division on its portal, it should be able to get this data from the scheme division via suitable exchange systems. M/Ds to select yes if this is possible currently.
A	F	1	If M/Ds are collecting similar data or running similar interventions, data collaborations can be undertaken. If private sector has some useful data (let's say e-commerce or traffic data), data collaborations can be undertaken by M/Ds.
A	F	2	If some data collaboration has been undertaken, M/Ds to select how it has been done from the given options.
A	G	1	Prescriptive analytics is the final stage of analytical capabilities. While predictive analytics answers what, when and why something will happen, prescriptive analytics builds on this further by specifying what present actions need to be undertaken to achieve the predictions and how will these decisions affect /impact other outcomes. Therefore, it helps in taking advantage of a future opportunity or mitigating future risks. It can also improve the accuracy of predictions by continuously taking in new data to re-predict and re-prescribe.
A	G	2	M/Ds to select the frequency of prescriptive analytics.
A	G	3	M/Ds to select the modes/mechanisms by which they have institutionalized prescriptive analytics, to ensure it is continuously undertaken to inform policymaking, and not just undertaken on random basis.
A	H	1,2,3	M/Ds to enter good practices of how they have used data for policymaking and/or set up systems for institutionalizing data driven policymaking.

B	A	1	Data requirements refer to various input, output, and outcome data points/indicators that need to be monitored. They must be clearly documented for each scheme. M/Ds to select yes if this is done.
B	A	2	After gathering of data requirements, scheme division to select the indicators for which it is collecting data also. For e.g.: If scheme has multiple outcome indicators documented but the division is collecting data on only some of them due to various reasons, it must select Partial.
B	A	3	After data is collected, it must be collated and reported via paper or digitally through a MIS. Scheme divisions to accordingly choose Yes/No. Regarding credentials, this is optional. However, M/Ds are encouraged to create dummy login credentials for DMEO with view-only rights. This shall stay confidential and not be used outside the government for unintended purposes.
B	A	4	Scheme division to select all the granularities at which data is reported on the MIS. For e.g.: if a scheme MIS has district, state as well as national level data, scheme division to select all three options.
B	A	5	Scheme division to select the frequency at which data is updated on the MIS.
B	A	6	This question is to essentially understand if the data reported by the M/D on the MIS is "collected" by humans or is it transaction-based collection. If it is collected by human resources, is it directly collected using digital tablets/mobiles etc. or is it the case that it is first collected on paper and then fed on computers by someone else.
B	A	7	If data is collected using digital modes or it is transactional in nature, use of survey tools and/or geotagging can improve the data reliability. Scheme division to select yes if the same is done.
B	A	7a	Computer-assisted personal interviewing (CAPI) refers to survey data collection by an in-person interviewer (i.e., face-to-face interviewing) who uses a computer to administer the questionnaire to the respondent and captures the answers onto the computer.
B	A	7b	A geotagged photograph is a photograph which is associated with a geographic position by geotagging. Usually this is done by assigning at least a latitude and longitude to the image, and optionally altitude, compass bearing and other fields may also be included.
B	A	7c	Geofencing is a location-based service which triggers some pre-programmed action like a survey when a mobile device or RFID tag enters or exits a virtual geographical boundary.
B	B	1	Data quality protocols and mechanisms should be clearly documented by the scheme division. Scheme division to select yes if the same is done.
B	B	2	Data quality assessment of collected data against data quality protocols can be undertaken automatically by advanced digital systems, manually or using a hybrid of both manual and automated systems.

B	B	3	Question to assess which protocols are included and followed by the scheme division in its data quality assessment.
B	B	3a	This is the first step of data quality where collected data is cleaned by checking missing values, incorrect responses etc.
B	B	3b	Next step is to generate summary statistics of data (like mean, median, trends etc.) to check for outliers
B	B	3c	Another important step is to ensure metadata is properly defined. Metadata is data about data – containing details on variables covered in the data, their number of observations, summary statistics, units etc. This must also be regularly updated if new data is collected.
B	B	3d	Next important protocol is to check collected data for duplicate values (this duplication may be in old data or new data) and remove any such redundancies
B	B	3e	Finally, ensuring data integrity. This means that if collected data is being reflected anywhere (on the MIS, on any other portal etc.), it must be ensured that the accurate and recent most value is reflected everywhere. It should not be the case that at one place, data is updated as of last month, but at other portal, it is updated as of last year or showing inaccurate value due to some error.
B	B	4	Apart from data qual assessment, backchecks may be deployed to further improve data quality and increase its reliability.
B	B	4a	Social audit is a form of citizen participation that focuses on government performance and accountability. If social audits are being used to improve scheme data, select yes.
B	B	4b	If telephonic backchecks are undertaken based on collected data to verify that data is correctly collected, select yes. E.g.: Based on PDS beneficiary data available on MIS, random sample of ppl are contacted on phone to validate data entries made on MIS.
B	B	4c	If there are provisions for citizens to submit multimedia evidence which is then used to improve the quality of data, select yes. For eg: People submitting photos of quality of roads built near their locations and this feedback data being used to reflect the quality of roads on MIS.
B	B	4d	Based on reported data on MIS, random inspections are made by MD officials to verify data on ground.
B	B	4e	Getting data on MIS verified/audited by third parties.
B	C	1	Scheme divisions to select all types of data analysis undertaken by them.
B	C	2	Apart from scheme data, if data from other schemes or sector level data is also used to complement scheme data for analysis purpose, select yes.
B	C	3	Data analysis must be documented in some manner. Select the frequency at which this is done.
B	C	4	M/Ds to select the uses for which data analysis is done.
B	C	5	Select different modes used for disseminating data and its analysis.
B	C	5a	DB is essentially a tool to display key KPIs from data and important analytics through interesting visualizations.

B	C	5b	Mobile apps can be used to share data with citizens and interact with them.
B	C	5c	Social media outlets can be used to share data with citizens and interact with them.
B	C	5d	SMS are often used to send details to users/beneficiaries with respect to the scheme activities.
B	C	5e	Such mass communication methods may also be used to share data with citizens.
B	C	6	M/Ds to select the purposes for which dashboard are being used by them.
B	C	9	This is important to ensure data is accessible to all.
B	C	10	This is important to ensure data is accessible to all.
B	C	11	All MIS may not be in public domain. Hence, scheme divisions to enter details on how can public in general can access MIS data.
B	C	12	Scheme divisions to check if there is an option to download all MIS data in machine readable formats by users on the MIS and accordingly select.
B	C	13	As per NDSAP, all non-personal data should be available on data.gov.in to facilitate easy access to all govt. data at one place. Scheme division to select yes if non-personal data of their MIS is available on this platform.
B	D	1	MIS linkage with PFMS means that latest status of funds being routed through PFMS should be linked with MIS of the scheme.
B	D	2	The field-level implementation agency is the last agency to which funds are to flow. For eg: if PFMS integration is done till state implementing agency level but fund flow below states is not PFMS integrated for a scheme where projects are implemented by city level agencies, integration is not completed till last mile.
B	D	3a	Applicable for beneficiary-oriented schemes
B	D	3b	Applicable for beneficiary-oriented schemes
B	D	3c	Applicable for beneficiary-oriented schemes
B	D	3d	Applicable for industry/firm oriented schemes
B	D	3e	Applicable for industry/firm oriented schemes
B	D	4a	Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted day-time radiation at a distance (typically from satellite or aircraft). Special cameras collect remotely sensed images, which help researchers "sense" things about the Earth. For e.g.: large forest fires can be mapped from space, Tracking clouds to help predict the weather or watching erupting volcanoes, and help watching for dust storms, tracking the growth of a city etc.
B	D	4b	Night-light data is basically the data of night-time lights emanating from the earth captured by satellites from outer space. These sources include moonlight, light directly emitted by a source (e.g., buildings and transport), and light reflected by the ground. It has several use cases - aid in disaster mitigation, estimating economic activity etc.

B	D	4c	This data is collected from social media networks to see how people are engaging on specific topics of interest. Scheme divisions may use the same to check for behavior change etc.
B	D	4d	Scheme may use data generated by private sector also as per requirement. For eg: mobility data from private cab aggregators, economic activity data from e-commerce websites etc.
B	D	5	Unique LGD codes have been created for each state, distt, sub-distt, block, village and local body by GoI. All MIS must use the same codes so that data on different platforms is easily integrable.
B	D	6a	Machine learning gives computers the ability to learn and predict from data without being explicitly programmed. E.g.: predicting the probability that individuals commit crimes, targeting hygiene inspections by data-mining online restaurant reviews or estimating poverty levels based on satellite imagery.
B	D	6b	AI refers to intelligence demonstrated by machines and can have several use cases in governance and delivery of schemes. e.g.: Monitoring social media for public feedback on policies, Monitoring social media to identify emergency situations, Anticipating road maintenance requirements, Providing personalized education to students etc.
B	D	6c	Blockchain refers to having distributed ledgers or blocks of transactional data that are linked together. Using this structure, govt. can offer services with improved data security. For e.g.: electronic health records, e-registries etc.
B	D	6d	IoT refers to network of objects embedded with sensors and technologies for collecting and exchanging data over Internet. e.g.: IoT to measure air quality, IoT to monitor power consumption i.e., smart metering etc.
B	D	6e	The use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources.
B	D	6f	Drones can be used for monitoring of various sectors like agri, infra projects, commerce, logistics etc.
B	E	3	An SSL certificate is a digital certificate that authenticates a website's identity and enables an encrypted connection. SSL stands for Secure Sockets Layer, a security protocol that creates an encrypted link between a web server and a web browser.
B	E	5	a firewall refers to a network device which blocks certain kinds of network traffic, forming a barrier between a trusted and an untrusted network.
B	E	7	Sensitive/PII contains personal information of individuals, firms etc. which are not freely accessible to all.
B	E	7a	Single-Factor Authentication (SFA) is a method of logging users by having them present only one way of verifying their identity (usually, username and password). Multi-factor authentication uses more than one way – such as OTP, Captcha etc.
B	E	7b	List of users of MIS along with details of which user has access to which type of data is regularly maintained.

B	E	7c	Encryption refers to conversion of data from readable format to encoded format. Encrypted data can only be read and processed after its decrypted by recipient if they have the codes.
B	E	7d	Data anonymization refers to the process by which personal data is altered in a way that the data subject can no longer be identified directly by data user.
B	E	8	With advancements in machine learning and big data analytics, it is becoming increasingly easier to de-identify anonymized data using indirect means. Hence, it is important to protect personal data from re-identification risks.
B	E	8b	Includes provisions for mandatory audit trails, controlled access, only central server logins allowed etc.
B	E	8c	Sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals.
B	E	9	Before using and putting personal data in public domain such as photographs, names, other details of individuals or firms, their consent must be asked for and documented.
B	F	2	Before using and putting personal data in public domain such as photographs, names, other details of individuals or firms, their consent must be asked for and documented.
B	F	3	Data may be stored on physical servers or cloud servers. Cloud servers offer better disaster recovery.
B	F	4	Select the cloud server used by the scheme MIS.
B	F	5	Historical data refers to data corresponding to previous time periods which may not be actively used at present.

Annexure 5: Question wise scoring mechanism

Pillar	Theme	Question No.	Question	Scoring mechanism
Data Strategy	Data & Strategy Unit	Part A, B1	Constitution	If response is 'Yes' score '5', if 'No' score '0'.
		Part A, B2	Head	If response if 'AS' or 'JS' score '5', if response is 'Director' score '3', if response is 'Below Director' score '0'. If question is disabled, score '0'.
		Part A, B3	Verticals	If all four verticals/sub-units are selected score '5', if three/two verticals are selected score '3', if only one is selected score '1', else score '0'. If question is disabled, score '0'.
		Part A, B4	Strength	If total % is > 80% score '5', if total % is >60% and <=80% score '4', if total % is >40% to <=60% score '3', if total is >20% and <=40% score '2', if total is <=20% score '0'. If question is disabled, score '0'.
		Part A, B5	ToR	If response is 'Yes' score '5', if 'Partial' score '3', if 'No' score '0'. If question is disabled, score '0'.
		Part A, B6	Review mechanisms	If response is 'Yes' score '5', if 'No' score '0'. If question is disabled, score '0'.
		Part A, B7	Frequency of review	If response is 'daily'/'weekly'/'fortnightly/monthly' score '5', if response is 'quarterly' score '3', if response is 'annually' score '1'. If question is disabled, score '0'.
	Action Plan	Part A, C1	Action plan	If response is 'Yes' score '5', if 'No' score '0'.
		Part A, C2	Sections	If response is 'Yes' score '5', if 'Partial' score '3', if 'No' score '0'. If question is disabled, score '0'.
		Part A, C3	Schemes	If response is 'Yes' score '5', if 'Partial' score '3', if 'No' score '0'. If question is disabled, score '0'.
		Part A, C4	Timelines	If response is 'Yes' score '5', if 'Partial' score '3', if 'No' score '0'. If question is disabled, score '0'.
		Part A, C5	Responsibilities	If response is 'Yes' score '5', if 'Partial' score '3', if 'No' score '0'. If question is disabled, score '0'.
		Part A, C7	Compliance Scoring	If timely compliance is above 80% score '5', if between 60% to 80% score '4', if between 40% to 60% score '3', if between 20% to 40%

Pillar	Theme	Question No.	Question	Scoring mechanism
				score '2', if below 20%, score '0'. If question is disabled, score '0'. If none of the action points are due when scores are being calculated, a standard score of 1 is given (as no timelines are due, it suggests action plan is not detailed and granular enough - hence low score).
Data Systems	Data Generation	Part B, A1	Requirements gathering	If response is 'Yes' score '5', if 'No' score '0'.
		Part B, A2	Collection	If response to all three parts is 'Yes', score 5. If response to two is 'Yes' and one is 'Partial', score 4. If response to two is 'Yes' and one is 'No', score 3. If response to one part is 'Yes' and two is partial, score '3'. If response to one is 'Yes', one is 'Partial', one is 'No', score 2. If response to one part is 'Yes' and two is 'No', score 1. If response to two is 'Partial' and one is 'no', then score '2'. If response to one part is 'partial' and two is 'no', score '1'. If response to all three parts is 'Partial', score 3. If response to all three parts is 'No', score 0.
		Part B, A3	Digitization	If response is 'Yes' score '5', if 'No' score '0'.
		Part B, A4	Granularity	Lowest level of granularity to be used - '1' at national level, '3' at State level, '4' at district/sub-district/block level and '5' at village/individual/facility/ project level. If question is disabled, score '0'.
		Part B, A5	Frequency	Lowest level of frequency to be used - '1' at Yearly, '2' at halfyearly, '3' at Quarterly, '4' at monthly/fortnightly/weekly and '5' at daily/realtime/near realtime level. If question is disabled, score '0'.
		Part B, A6 and 7	Use of technologies in generation	If Q6 is disabled, score '0'. If response to Q6 is 'Option 1' score '1'. If response to Q6 is 'Option2' or 'Option 3', then use Q7 responses to score. If none of the responses to Q7 is 'Yes', score '3'. If any one responses to Q7 is 'Yes', score '5'.

Pillar	Theme	Question No.	Question	Scoring mechanism
	Data Quality	Part B, B1	QC mechanisms documentation	If response is 'Yes' score '5', if 'No' score '0'.
		Part B, B2	QC automation	If response is 'Not done' score '0', if 'Manually' score '2', if 'Hybrid' score '3', if 'Automatically' score '5'.
		Part B, B3	Data quality assessment	If no response is 'Yes', score '0'. If any one response is 'Yes' score '1', if any two responses are 'Yes' score '2', if any three responses are Yes, score '3', if any four responses are 'Yes' score '4', if all responses are yes, score '5'. If question is disabled, score '0'.
		Part B, B4	Use of mobile phones in QC	If no response is 'Yes', score '0'. If one or two responses are 'Yes' score '3'. if three or more responses are 'Yes' score '5'. If question is disabled, score '0'.
		Part B, C1	Types of data analysis	If no response is 'Yes' score '0'. If any 1/6 option is selected then score '1', if 2/6 options are selected then score '2'. If 3/6 options are selected then score '3'. If 4 or more options are selected then score '5'.
		Part B, C2	Cross sectoral analysis	If response is 'Yes' score '5'. If response is 'No' score '0'. If question is disabled, score '0'.
		Part B, C3	Documentation of data analysis	If response is 'Never' score '0'. If 'Annually', score '2'. If 'Half-yearly', score '3'. If 'Quarterly', score '4'. If 'Real time on a dashboard', score '5'. If question is disabled, score '0'.
		Part B, C4	Use of data analysis	If question is disabled, score '0'. If response is 'no' for all the sub-categories, score '0'. If response yes for 1-2/7 sub-categories, score '1'. If response is yes for 3-4/7 sub-categories, score 3'. If response is yes for 5-7/7 sub-categories, score '5'.
		Part B, C5	Modes of dissemination	If response is 'No' for all sub-categories, score '0'. If response is 'Yes' for 1-2 sub-categories, score '1'. If response is 'Yes' for 3-4 sub-categories, score '3'. If response is 'Yes' for 5-7 sub-categories, score '5'.

Pillar	Theme	Question No.	Question	Scoring mechanism
		Part B, C6	Use of dashboards	If question is disabled, score '0'. If response is 'No' for all sub-categories, score '0'. If response is 'Yes' for 1-2 sub-categories, score '3'. If response is 'Yes' for 3-4 sub-categories, score '5'.
		Part B, C7	Data visualization types	If question is disabled, score '0'. If response is 'No' for all sub-categories, score '0'. If response is 'Yes' for 1-2 sub-categories, score '1'. If response is 'Yes' for 3-4 sub-categories, score '3'. If response is 'Yes' for 5-7 sub-categories, score '5'.
		Part B, C8	Data visualization on maps	If question is disabled, score '0'. If Yes, score '5', else '0'.
		Part B, C9 and 10	Data Accessibility for all	If question is disabled, score '0'. If response to both Q9 and Q10 is 'No', score '0'. If response to one is 'No' and one is 'partially' score '1'. If response to both are 'partially', score '3'. If response to one is 'Yes' and one is 'partially', score 4. If both are 'yes' score '5'.
		Part B, C11	Open data	If question is disabled, score '0'.If response is 'Not accessible', score '0'. If response is 'Accessible through credentials', score '3'. If response is 'Openly accessible', score '5'.
		Part B, C13	Open data	If question is disabled, score '0'. If response is 'Yes', score 5, else '0'.
		Part B, C12	Machine readable data	If question is disabled, score '0'. If response is 'Yes', score '5', if 'Partially' score '3', if 'No' score '0'.
	Use of technology	Part B, D1	Linkage with PFMS	If response is 'Yes' score '5', else '0'. If question is disabled, score '0'.
		Part B, D2	Last mile linkage of PFMS	If previous question was disabled, this will also be disabled and scored '0'. If 'yes' is selected in previous question, score based on response provided to this question. If 'no' is selected as a response here, score '0', if 'Partially', score '3', if 'Yes', score '5'. If 'no' is selected in previous question, this question will be

Pillar	Theme	Question No.	Question	Scoring mechanism
				disabled and scored '0'. If 'NA' is selected in previous question, this question will be disabled and scored '5'.
		Part B, D3	Linkage with other platforms	If no option selected, score '0'. If one option is yes, score '3'. If more than two options are selected, score '5'. If question is disabled, score '0'.
		Part B, D5	Linkage with LGD Codes	If response is 'Yes' score '5', else '0'. If question is disabled, score '0'.
		Part B, D4	Use of alternative data sources	If no option selected, score '0'. If any one option is yes, score '5'.
		Part B, D6	Use of emerging technologies	If no option selected, score '0'. If any one option is yes, score '5'.
	Data security & HR capacity	Part B, E1	Antivirus updates	If response is 'Yes', score 5, else '0'. If question is disabled, score '0'.
		Part B, E2	Security audits	If response is 'Yes', score 5, else '0'. If question is disabled, score '0'.
		Part B, E3 and 4	SSL certification	If response to Q3 is 'No', score '0'. If response to Q3 is 'Yes', use responses for q4 to score further. If response to Q4 is 'No', score '3'. If response to Q4 is also 'Yes', score '5'. If Q3 is disabled, score '0'.
		Part B, E5	Firewalls	If response is 'Yes', score 5, else '0'. If question is disabled, score '0'.
		Part B, E6	External communication	If response is 'Yes' or 'No external communication established', score 5, else '0'. If question is disabled, score '0'.
		Part B, E7, 8	Personal data protection	If question is disabled due to no MIS, score '0'. First check if response is 'No such data', score '5'. If this option is not selected, check which of remaining four options are selected. If only first/second option is selected, score '1'. If both first and second option are selected (but not third and fourth) score '2'. If third option is selected (but fourth is not), score '3'. If fourth option is selected (either along with other options or only fourth option is

Pillar	Theme	Question No.	Question	Scoring mechanism
				selected) and 'No efforts made'/'others' selected in Q8, score '4'. Further, if fourth option is selected (either along with other options or only fourth option is selected) and any other option selected in Q8, score '5'.
		Part B, E9	Personal data protection - 2	If question is disabled due to no MIS, score '0'.If question was disabled due to 'No such data' response in Q7, score '5'. If response is 'Yes', score 5, else '0'.
		Part B, E10	Data QC team	If response is 'Yes', score 5, else '0'.
		Part B, E11	Data analysis team	If response is 'Yes', score 5, else '0'.
	Data management	Part A, D1,3,4,5	Data management architecture	If response to all four questions is 'No', score '0'. If response to only one question is 'Yes', score '1', if response to only any two questions is 'Yes', score '2'. If response to any three is 'Yes', score '4'. If response to all is 'Yes', score '5'.
		Part A, D2	Data management Compliance	If response is 'Yes', score 5, else '0'. If question is disabled, score '0'.
		Part B, F1 and 3	Distributed cloud storage	If 'Separate servers' in Q1 and 'physical servers' or 'hybrid servers' in Q3, score '0'. If 'Central server' in Q1 and 'physical servers' or 'hybrid servers' in Q3, score '1'. If 'Separate servers' in Q1 and 'cloud servers' in Q3, score '3'. If 'Central server' in Q1 and 'cloud servers' in Q3, score '5'.
		Part B, F2	Data sharing mechanisms	If 'yes', score '5', else score '0'.
		Part B, F4	Type of cloud storage	If first or second option is selected, score '5', else score '0'.
		Part B, F5	Historical data management	Is response is "Data is not backed up", score '0'. If response is "Data is backed up and data is archived", score '3'. If response is "Data history is well maintained including retention, destruction, and audit trail details", score '5'. If

Pillar	Theme	Question No.	Question	Scoring mechanism
				question is disabled, score '0'.
Data driven outcomes	Synergistic data use within M/D	Part A, E1 and 2	Identification of data gaps	Is response to Q1 is "No", score 0. If response to Q1 is "Yes", use Q2 to score further. If response to Q2 is "No", score '3', If response to Q2 is also 'Yes', score '5'.
		Part A, E3	Data exchange	If response is "No", score 0. If response is "In- progress", score 3. If response is "Yes", score 5. If "Yes" is the response, its veracity will be validated from the subjective descriptions and hence responding to descriptions is also important.
	Inter-agency data collaboration	Part A, F1	Collaborations	If response is "No", score 0. If response is "In- progress", score 3. If response is "Yes", score 5. If "Yes" is the response, its veracity will be validated from the subjective descriptions and hence responding to descriptions is also important.
		Part A, F2	Types of collaborations	If none of the options are selected, score '0'. If one to five options selected, score '3'. If more than five options are selected, score '5'. If question is disabled, score '0'.
	Prescriptive Analytics	Part A, G1	Prescriptive analytics	If response is "No", score 0. If response is "In- progress", score 3. If response is "Yes", score 5. If "Yes" is the response, its veracity will be validated from the subjective descriptions and hence responding to descriptions is also important.
		Part A, G2	Frequency	If response is "Annually", score 3. If response is "Quarterly/Monthly" score 5. If question is disabled, score '0'.
		Part A, G3	Modes	If none of the options are selected, score '0'. If any one option is selected (other than "others"), score '5'. If question is disabled, score '0'.
	Good Practices	Part A, H	Good practices	Each good practice will be assessed on 3 parameters – relevance of practice to DGQI exercise and objectives (40%), exhaustiveness of the case study (30%) and impact of the intervention (30%).” Then, a simple average of the three scores

Pillar	Theme	Question No.	Question	Scoring mechanism
				for each good practice will be taken to arrive at overall good practice dimension score.

Annexure 6: NA scoring mechanism

For scoring purposes, for a certain question, NA option is selected, to not penalize any entity for any requirement that is not applicable for them, its weight will be redistributed among other questions within the theme. However, if it is the case that only certain sub-parts (a, b,.) of a question are not applicable, a case-by-case mechanism of how they willbe taken care of in at the scoring stage has been devised in the following manner.

Q.No.	Question	Scoring mechanism	Way to handle NA
Part B, A4	Granularity	Lowest level of granularity to be used - '1' at national level, '3' at State level, '4' at district/sub-district/block level and '5' at village/individual/facility/ project level. If question is disabled, score '0'.	Scoring should not be changed since it's a range. Any one of the options from village/individual/facility/project must be applicable for all schemes. Hence, if the scheme is collecting data at any level not equivalent to these four options, scores should be decreased the way they have been done.
Part B, A5	Frequency	Lowest level of frequency to be used - '1' at Yearly, '2' at halfyearly, '3' at Quarterly, '4' at monthly/fortnightly/weekly and '5' at daily/realtime/near realtime level. If question is disabled, score '0'.	Only need to change scoring if daily/ realtime/ near realtime/ monthly/fortnightly/weekly - all of these options are not applicable (infra schemes with long gestation periods). In this case, quarterly to be scored as '5', half yearly as '3' and yearly as '1'.
Part B, A6 and 7	Use of technologies in generation	If Q6 is disabled, score '0'. If response to Q6 is 'Option 1' score '1'. If response to Q6 is 'Option2' or 'Option 3', then use Q7 responses to score. If none of the responses to Q7 is 'Yes', score '3'. If any one responses to Q7 is 'Yes', score '5'.	Only need to change scoring if none of the options in Q7 are applicable, otherwise scheme already gets full score. In this case, scoring will be done only on basis of 6 - 1 if first option is selected, 5 if other two options are selected.
Part B, B4	Use of mobile phones in QC	If no response is 'Yes', score '0'. If one or two responses are 'Yes' score '3'. if three or more responses are 'Yes' score '5'. If question is disabled, score '0'.	The options are such that if one is applicable, all others would also be applicable. Hence, only need to change scoring if none of the options are applicable (research/defence schemes). If this is the case, weights to be redistributed within data quality theme.
Part B, C2	Cross sectoral analysis	If response is 'Yes' score '5'. If response is 'No' score '0'. If question is disabled, score '0'.	If it is NA, its weight will be redistributed within data analysis theme.

Q.No.	Question	Scoring mechanism	Way to handle NA
Part B, C5	Modes of dissemination	If response is 'No' for all sub-categories, score '0'. If response is 'Yes' for 1-2 sub-categories, score '1'. If response is 'Yes' for 3-4 sub-categories, score '3'. If response is 'Yes' for 5-7 sub-categories, score '5'.	It should not be the case that all options are NA- If scheme enters so, it would be scored '0' as a disincentive for entering wrong responses. Otherwise, NA treated as yes and accordingly scored as per the method.
Part B, C8	Data visualization on maps	If question is disabled, score '0'. If Yes, score '5', else '0'.	If it is NA, its weight will be redistributed within data analysis theme.
Part B, D1	Linkage with PFMS	If response is 'Yes' score '5', else '0'. If question is disabled, score '0'.	If it is NA, its weight will be redistributed within use of technology theme.
Part B, D2	Last mile linkage with PFMS	If response is 'Yes', score '5', if 'Partially' score '3', if 'No' score '0'. If question is disabled, score '0'.	If it is NA, its weight will be redistributed within use of technology theme.
Part B, D3	Linkage with other platforms	If no option selected, score '0'. If one option is yes, score '3'. If more than two options are selected, score '5'. If question is disabled, score '0'.	If every option is NA - weight to be redistributed. Otherwise, ranges are defined in a manner that schemes get correct scores. If any one option is yes, by virtue of options, atleast two become applicable.
Part B, D4	Use of alternative data sources	If no option selected, score '0'. If any one option is yes, score '5'.	If every option is NA – weight to be redistributed. Otherwise, ranges are defined in a manner that schemes get full scores.
Part B, D5	Linkage with LGD Codes	If response is 'Yes' score '5', else '0'. If question is disabled, score '0'.	If it is NA, its weight will be redistributed within use of technology theme.
Part B, D6	Use of emerging technologies	If no option selected, score '0'. If any one option is yes, score '5'.	If every option is NA - weight to be redistributed. Otherwise, ranges are defined in a manner that schemes get full scores.
Part A, E3	Data exchange	If response is "No", score 0. If response is "In- progress", score 3. If response is "Yes", score 5. If "Yes" is the response, its veracity will be validated from the subjective descriptions and hence responding to descriptions is also important.	If it is NA, its weight will be redistributed within synergistic data use within M/D theme.