KPMG

DGQI 2.0 Questionnaire

01

Data System Pillar (60% weightage)

PART B

Data System Pillar (60%) - Data Generation Theme (11%)

Q1 - Are the data requirements of the scheme well defined and documented?
(Yes/No)

Q2 - Is data collected for all identified data requirements?
(Input/output/outcome points)

Q3 - Is collected data reported digitally? (i.e., is there is a digital electronic database/MIS)?
(Yes/No)

Q4 - At what granularity is data reported digitally for the scheme?
(State/District/subdistrict/tehsil/block/village/facility/project)

Q5 - At what frequency is data reported digitally for the scheme?
(Realtime/Daily/weekly/monthly/quarterly/half-yearly/annually)

Q6 - How is this data collected at the ground level?
(Collected on pen and paper then fed on digital system or collecting using digital modes)

Q7 – Are any of the following technologies used?
(CAPI Surveys/Geotagged/Geofenced information/others)

Data Generation Theme - Explained

Q1- Select 'Yes', if the MD has prepared such a document and following aspects are considered in the document (but not limited to):

a. The document clearly outlines the need for collecting data on the intervention

b. The document clearly identifies and defines the data points/indicators on which data needs to be collected i.e., input, output and outcome indicators. For example, for Samagra Shiksha scheme, one of the indicators at input stage can be the funds released to upgrade schools. Associated output and outcome indicators can be the number of schools upgraded and Gross enrolment rate, respectively. So, the Department should clearly specify such indicators on which data needs to be collected.

c. The document clearly defines the granularity at which the data for above mentioned indicators would be collected so that data can be used in the intended manner. For example, granularity for beneficiary-oriented schemes must be at beneficiary level, for firms it must be at facility level etc.

d. The document clearly defines the frequency at which the data for above mentioned indicators would be collected. For example, frequency for data on scholarship schemes must be the frequency at which scholarships are disbursed.

e. The document clearly defines the mode and method of data collection to be used for collecting data. For example, how will data be generated by ground staff (on paper or via tablets or transactionally), how will it be reported at higher levels (manual entry or transactionally, on paper or digitally, aggregated or disaggregated etc.) and if any other technologies or data sources will be used to collect any data points.

Q2- For each sub-question (a), (b) and (c):

Select 'Yes' if the scheme division is collecting data on all identified data points or indicators on which data needs to be collected (as per the data requirements document).

Select 'Partial' if the scheme division is collecting data only on some (not all) of the data points or indicators on which data needs to be collected (as per the data requirements document).

Select 'No' if the scheme division is not collecting data on any of the identified data points or indicators on which data needs to be collected (as per the data requirements document).

Data Generation Theme - Explained

Q3 - Select 'Yes' if the scheme division collates and reports collected data digitally via a MIS (Management Information System). This also includes scheme which may be collecting data on paper first but have mechanisms to later convert and report this data digitally.

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.

Q4 - This is a multi-select question. Scheme division to select 'Yes' for all the granularity levels at which the data is being reported on the MIS.

For e.g.: if a scheme MIS has district, state as well as national level data, scheme division will select all three options. And, if the scheme is collecting data at beneficiary level but MIS has granularity till village level only, then scheme division will select 'Village' (not 'Individual/Household').

Q5 - This is a multi-select question. Scheme division to select 'Yes' for all the frequency levels at which the data is being updated on the MIS.

Q6 - This is a multi-select question.

Select 'Collected on paper by human resources and then fed on digital systems' if the ground-level data is being collected by human resources who are using paper to do the same and then feeding the collected data into digital systems.

Select 'Collected using digital modes (tablets/phones etc.) by human resources', if humans are collecting the ground-level data and are using digital modes such as tablets/phones to do the same.

Select 'Transactional Data' if the collected data is transactional in nature i.e., it is collected in the process of a digital transaction requiring no separate human resources to separately collect this data. For example, bank account transfers in DBT scheme are transactional data.

Data Generation Theme - Explained

Q7 - This is a multi-select question.

Select 'CAPI surveys' if the scheme division is collecting ground level data using CAPI. Computer-assisted personal interviewing (CAPI) refers to survey data collection by an in-person interviewer (i.e., face-to-face interviewing) who uses a mobile/ tablet to administer the questionnaire to the respondent and captures the answers onto the device.

Select 'Geotagged information' if the scheme division is collecting the same in process of other data collection. 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. For example, this is used by the Department of Rural Development in the PM- Awaas App to monitor and track the physical progress of construction of houses under PMAY-G scheme.

Select 'Geofenced information' if the scheme division is collecting the same. 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

PART B

Data System Pillar (60%) - Data Quality Theme (11%)

Q1 - Are there pre-defined documented mechanisms to assess quality of incoming data?

(Yes/No)

Q2 - How is data quality assessment done?

(Automatically/Manually/Hybrid/Not Done)

Q3 - 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.

b) Summary statistics of incoming data are generated and checked for errors/abnormalities

c) Existence and accuracy of metadata for all the scheme's data is periodically checked (Schema is well defined)

d) There is a system for identifying duplicate data and removing redundancies

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)

Q4 - Are following feedback mechanisms/backchecks also leveraged for data quality control?

(Social Audits/Telephonic backchecks/verification with beneficiaries/ Multimedia data – citizen voice, video, images as evidence/ Sample inspections based on data/ Third party data verification/ data audits)

Data Quality Theme - Explained

Q1 - Select 'Yes', if the MD has prepared such a document and following aspects are considered in the document (but not limited to):

A. The document clearly outlines **the data quality dimensions** to be assessed by the scheme division. These include dimensions such as: accuracy of data, completeness, timeliness, validity of data, consistency etc.

B. The document clearly outlines all **the protocols to be followed to assess data quality of incoming data**. Examples include:

i. Data validation: Placing automatic checks at data input stage to ensure that data is uploaded in right format. For e.g., Mobile number has to only be a number of 10 digits. It cannot be alphanumeric or alphabets.

ii . Data filtering: Data is checked for missing values or logical flaws. Data with errors is sent back for verification or eliminated for further processing. For e.g., if scholarship scheme benefits are only given to students of a certain age, data of students with ages beyond the threshold limit is resent to field staff for verification (in case data is received from any agency) or the application is rejected (in case data is transactional in nature)

iii. Summary statistics: Summary statistics of data such as mean, median, mode may be regularly calculated to quickly see if there is any outlier or abnormality In the data from usual trend. For e.g., mean age of students applying for scholarships during each quarter may be tracked. If, at any particular point of time, there is a huge decrease or increase in the mean, it may flag some concerns with actual database which needs to be further checked.

iv. Data deduplication: Data is checked for duplicates using unique identifiers. For example, Aadhaar Id may be used as unique identifier of students in scholarship schemes database. If another application is received from same Aadhaar ID, the application should be rejected, either automatically or manually.

C. The document clearly outlines the frequency at which the scheme division should be assessing data quality using above-mentioned protocols. For e.g., the frequency of checking if the data input is of valid format can be real-time.

D. The document clearly outlines the backchecks to be undertaken by the scheme division to again check data quality, such as telephonic checks. Other feedback mechanisms/backchecks are mentioned in Question 4 of this theme.

Data Quality Theme - Explained

Q2 - Select 'Automatically' if the data quality assessment of collected data against the defined data quality protocols can be undertaken automatically by advanced digital systems with no human intervention. For example, data validation and deduplication can be easily automated if digital modes such as CAPI surveys or transactional data is used to collect data on the ground. Similarly, generation of summary statistics can be automated on the MIS.

Select 'Manually' if the assessment is undertaken manually without using any digital systems. Select 'Hybrid' if the assessment is undertaken by using a hybrid of both manual and automated systems. Also, please elaborate on this hybrid mode, by typing in the details in the text box.

Select 'Not Done' if the scheme division is not undertaking any assessment to check data quality

Q3 - This is a multi-select question. Select 'Yes' for all the protocols that are regularly followed (executed) by the scheme division to assess its data quality:

A. If the collected data is cleaned by removing missing values, eliminating data with incorrect responses etc., select 'Yes'

B. If the collected data is used to generate summary statistics (like mean, median, trends etc.) to check for outliers, select 'Yes'

C. If the scheme division has clearly defined the metadata in a document and take steps to ensure that it is regularly updated, select 'Yes'. Metadata is data about data – containing details on variables covered in the data, frequency and granularity of their generation, number of observations, units of measurement of variables etc.

D. If the scheme division take steps to check collected data for duplicate values (this duplication may be in old data or new data) and remove any such redundancies (manually or automatically), select 'Yes'

E. If the scheme division has clearly defined data integrity and take steps to ensure the same, select 'Yes'. Data integrity means that no data is mistakenly deleted, lost or altered in this process of data processing. For example, the same data that is entered by field workers on paper should be reflected on the MIS. Another example would be to ensure that data is updated as of last month at one page of the MIS but it is updated as of last year on another page or on paper.

Data Quality Theme - Explained

Q4 - This is a multi-select question. Select 'Yes' for all the backcheck mechanisms that are regularly followed (executed) by the scheme division to further assess its data quality:

A. Social audit is a form of citizen participation that focuses on government performance and accountability. The findings from the social audits can then be compared with the official records. For example, People submitting photos of quality of roads built near their locations and this feedback data being used to cross check the progress of development of roads as reported on MIS. If social audits are being used to improve scheme data, select 'Yes'.

B. Telephonic backchecks can be undertaken based on collected data to verify that data is accurate. For example, telephonic calls to random sample of beneficiaries of PDS system using beneficiary level data reported on its MIS to check if they have actually received their entitlements. If such telephonic backchecks are being used to improve scheme data, select 'Yes'.

C. If there are provisions for citizens to submit multimedia evidence which is then used to improve the quality of data, select 'Yes'. For example, geotagged images of construction of houses submitted by citizens are used to validate the use of funds for intended purposes and for fund disbursements under PM Awaas Yojana.

D. Based on reported data on MIS, if random inspections are made by Ministry officials to verify data on ground, select 'Yes'. For example, SBM (G) scheme guidelines has suggested states to conduct such sample inspections.

E. If the scheme division is getting scheme's data on the MIS verified/audited by third parties, select 'Yes'. This does not include security audits of the website. For example, UIDAI uses a third-party data verification methodology to cross check Aadhaar data fed on the ground and ensure its accuracy.

PART B

Data System Pillar (60%) - Data Analysis, Use and Dissemination (11%)

Q1 - What types of data analysis is undertaken on collected data?

a. **Descriptive data analysis (e.g., basic cross tabulation, frequency distribution, mean, median etc.)**
b. **Exploratory data analysis (e.g., correlation etc.)**
c. **Inferential data analysis (Using a small sample of data to infer about a larger population)**
d. **Predictive analysis (Using historical or current data to find patterns to make predictions about the future)**
e. **Causal analysis (Looks at the cause and effect of relationships between variables, focused on finding the cause of a correlation)**
f. **Mechanistic Analysis (Understand exact changes in variables that lead to other changes in other variables)**

Q2 - Is cross-schematic/sectoral data also analyzed, wherever needed?

(Yes/No)

Q3 - How often is this data analysis well-documented (in reports/notes/publications)?

(Real time/Quarterly/Half-yearly/Annually/Never)

Q4 - 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.**
b) **To do mid-course corrections through design or implementation changes?**
c) **To guide intra-scheme funding decisions like interstate allocations, inter-component allocations, etc.?**
d) **To guide inter-scheme budgetary allocations?**
e) **To decide quarterly releases to implementing agencies?**
f) **For fraud management and analysis?**
g) **Day to day delivery and monitoring of implementation/ performance of the scheme**

NEXT

PART B

Data System Pillar (60%) - Data Analysis, Use and Dissemination (11%) Cont.

Q5 - What other modes are used to disseminate the MIS/ paper-based data and related data analysis?

(Social Media/ Dashboard/ Mobile App /SMS /Newspaper Magazines /Outdoor Media / Events/ TV Radio/ Others)

Q6 - 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

b. Capturing trends over time and identifying preempt trends

c. Measure efficiencies/inefficiencies in processes

d. User friendly one stop access to multiple automated reports

Q7 - What types of Data Visualizations are used?

(Bar Charts/ Histograms/ Pie Charts/ Scatter plot/ Heat Maps/ Tree Maps/ Gantt Chart/ Specialized visualizations- Stripe graphics, streamgraph, Others)

Q8 - Does the Dashboard visualize information on maps?

(Yes/No/NA)

Q9 - Does the information system (of MIS) support multilingual features as per GIGW norms?

(Yes/No/Partially)

NEXT

PART B

Data System Pillar (60%) - Data Analysis, Use and Dissemination (11%) Cont.

Q10 - Does the information system (of MIS) support features for differently abled as per GIGW norms?
(Yes/No/Partially)

Q11 - 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)

Q12 - 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)

Q13 - Is the MIS data available on 'data.gov.in' (non-sensitive data which can be shared)?
(Yes/No)

NEXT

Data Analysis, Use and Dissemination Theme - Explained

Q1 -This is a multi-select question. Select 'Yes' for all the types of data analysis that is regularly undertaken by the scheme division and associated data analysis reports are available:

A. Descriptive Analysis- The goal of this analysis is to describe a certain set of data. This is the most basic level of analysis that is done and generates simple summaries about the data population. Common descriptive statistics include mean, median, mode, variability, frequency, etc. For e.g., COVID-19 statistics page and line graph on Google, with summary information about the total number of cases, total deaths, variation in cases/deaths over time/regions and description of the population of a particular country infected by the virus.

B. Exploratory Analysis- The goal is very similar to descriptive analysis apart from the fact that it tries to also find relationships between variables present in the dataset. It is useful for discovering new connections and identifying gaps. For e.g., correlating Covid-19 deaths in a region with the number of public hospitals in the region or the income levels of the people residing in the region would be exploratory analysis.

C. Inferential Analysis- The goal is to use a small sample of data to make inferences about a larger population. For e.g., sample surveys are often undertaken by different scheme divisions to evaluate if the said scheme has saturated in a particular state/ region and achieved the required objectives.

D. Predictive Analysis- The goal is to use historical/ current data to find patterns to make predictions about the future. The accuracy of the predictions depends upon input variables. For e.g., predicting Covid-19 onslaught in regions based on past/present trends of Covid-19 caseload and other factors such as new variants, intensity of lockdowns etc.

E. Causal Analysis- Causal analysis looks at the cause and effect of relationships between variables and is focused towards finding the cause of a correlation. For. eg. – Randomized Control Trials (RCTs) have been undertaken by different scheme divisions or States at pilot stage before rolling out the programme to understand the potential impact of schemes/ interventions, especially in beneficiary-oriented schemes which have the behavioural change component.

F. Mechanistic Analysis- The goal is to understand the exact changes in variables that lead to other changes in other variables. In some sense, mechanistic analysis is in some ways a predictive analysis, but modified to tackle studies that require high precision and meticulous methodologies for physical or engineering science. For e.g., drug trials which helps researchers understand the exact mechanism of action of the intervention and contribute to the understanding of the disease process.

Data Analysis, Use and Dissemination Theme - Explained

Q2 - Select 'Yes' only if in addition to schematic data, data from other schemes or sector level data from other sources (such as survey data etc.) is used to inform scheme level decision making on a regular basis.

Q3 - The scheme division is required to select the frequency at which the various types of analysis undertaken are documented in reports to help understand how is data being analyzed and used to inform decision making at the scheme level.

Q4 - This is a multi- select question, here the scheme division is required to select 'Yes' for all the uses for which data analysis is undertaken by them.

Q5 - This is a multi- select question, the scheme division is required to select 'Yes' for all modes used for disseminating schematic data and/or its analysis.

A. Dashboards are an essential tool to display Key Performance Indicators (KPIs) from schematic data and key insights can be drawn using interesting visualizations available on interactive dashboards. For. e.g. - PM Svanidhi scheme of the Ministry of Housing & Urban Affairs has a real time updated dashboard which can be accessed for understanding the progress of the scheme in terms of people benefitted, funds disbursed, beneficiary distribution on various socio-economic categories etc. State and implementing agency wise data, reports and visualizations are available to quickly compare them on KPIs.

B. Mobile Apps have been utilized by various scheme divisions to share data with citizens and interact with them. For e.g.- The mParivahan app developed by the Ministry of Road Transport and Highways has information concerning all Regional transport Offices (RTOs) and vehicles on them and helps citizens locate the nearest RTO and pollution checking centre as well as apply for mock driving license tests.

C. Social Media outlets have been used by multiple ministries and scheme divisions to share data with citizens and interact with them on a large scale. For e.g., most Ministries/Departments have social media presence on major platforms which are regularly used by them to share progress and success stories of various schemes/ interventions using certain KPIs/datapoints. The Ministry of Railways often responds to queries/ support requests posted on their handle.

D. SMS is often used by scheme divisions to send details regarding transaction/ process updates to users/ beneficiaries of certain schemes. For e.g.- Notification of DBT transactions to individual beneficiaries are often sent using SMS to the registered mobile numbers.

E. Newspapers/Magazines, Outdoor Media, Events, TV/Radio are effective mass communication methods which may be leveraged by scheme divisions from time to time to disseminate scheme related information and share key datapoints with citizens. It may also be used to spread awareness about detailed MIS portal maintained by these schemes for citizen use.

Data Analysis, Use and Dissemination Theme - Explained

Q6 - This is a multi- select question, the scheme division is required to select 'Yes' for all purposes the scheme dashboard is used for. KRA stands for Key Result Areas and KPI stands for Key Performance Indicators of the scheme.

Q7 - This is a multi- select question, the scheme division is required to select 'Yes' for all modes of data visualization available on the scheme dashboard.

Basic visualizations (options a, b and c) include Bar Chart/Histogram, Pie charts and Scatter Plots.

Heatmap uses a system of color-coding to represent different values and helps visualize density. By observing how cell colours change, one can observe if there are any patterns in value for one or both variables. For example, DGQI report visually presents DGQI scores of Ministries/Departments in a heatmap with light green background colour used for low scores and the colour keeps getting darker as score improves.

Treemap helps visualize hierarchical data that helps capture relative sizes of data categories allowing for quick perception of the items that are large contributors to each category. Colour coding of the items can help identify the items that are underperforming or over performing compared to the other items in the same category. For example, home page of OOMF Dashboard of DMEO shows the Ministries/Departments in a treemap. The box represents the budget of the Ministry/Department – the larger the size, the larger the budget. The colour represents the compliance of Ministries/Departments on the dashboard – Red for low, Yellow for medium and green for high compliance.

Gantt Charts help track progress over time and help allocate resources. They depict the relationship between the start and end dates of tasks, milestones and other dependent tasks. For example, the PMG dashboard of DPIIT uses Gantt Charts (of sort) to show the progress of an infrastructure project over time.

Stripe Graphs use a series of coloured stripes chronologically ordered to visually portray long term trends. For example, if climate change trends over decades needs to be shown visually, the temperature can be shown as stripes in different colors, with the color getting warmer as temperature increases.

Stream Graph displays the changes in data over time through the use of flowing, organic shapes which resemble a river like stream. For example, changes in sales of a product over time can be visualized via a stream graph.

Data Analysis, Use and Dissemination Theme - Explained

Q8 - Scheme division to select 'Yes' only if the dashboard includes visualizations on maps such as those available on the PM Svanidhi dashboard developed by the Ministry of Housing and Urban Affairs or the Poshan Tracker maintained by the Ministry of Women and child Development.

Q9 - Scheme division to select 'Yes' only if scheme MIS supports this multilingual feature for all pages of the MIS based on the Guidelines for Indian Government Website developed by NIC.

Q10 - Scheme division to select 'Yes' only if scheme MIS supports features to make MIS accessible to differently abled (screen reader, font size options) for all pages of the MIS based on the Guidelines for Indian Government Website developed by NIC.

Q11 - The scheme division to select the option most suitable to them based on the nature of the dashboard. If anyone (public) can access the dashboard without any login system, select Option 1. If anyone can access the dashboard but only after login, select Option 2. If it is not at all accessible, select Option 3.

Please note that with differential access rights, sensitive data can still be hidden for public access. Hence, the question is pertaining to how public information is disseminated on the dashboard.

Q12 - The scheme division is to select 'Yes' only if there is an option to download all MIS data in machine readable formats (for instance excel, CSV, dta file formats) by users on the MIS. Select 'Partial' if this option is available only for certain types of non-sensitive data available on the dashboard. Select 'No' otherwise.

Q13 - Scheme division to select 'Yes' only if all non-personal MIS data is available on data.gov.in as per the National Data Sharing and Accessibility Policy.

PART B

Data System Pillar (60%) - Use of technology (6%)

Q1 - Does the MIS of the scheme have linkages with PFMS?

(Yes/No/NA)

Q2 - Is PFMS integration completed till the field-level implementation agency?

(Yes/No/Partially)

Q3 - Does the MIS of the scheme have linkages with:

Aadhar /Mobile numbers / Bank Accounts / GSTN / Udyog Aadhar / Others.

Q4 - Does the scheme use any of the following:

Remote Sensing / Night Light Data / Social Media Data / Private sector generated data / Others.

Q5 - Is the MIS compliant with Local Govt Directory (LGD)?

(Yes/No/NA)

Q6 - Does the scheme apply/use any of the following:

(Machine Learning / Artificial Intelligence / Block Chain / Internet of Things / Big Data Analytics / Drones)

NEXT

Use of Technology Theme - Explained

Q1 - The Scheme Division to select 'Yes' only if PFMS and MIS system of the scheme are automatically linked to each other. Hence, the latest status of funds being routed through PFMS is directly available on the MIS of the scheme without any manual data update. There is hence a single point of data generation, and it gets recorded automatically in both the PFMS portal and the MIS of the scheme.

It means that there should not be two separate lists based on two different databases which results in duplication and increases the chances for error. Also, just adding the hyperlink to the PFMS portal on the MIS of the scheme does not mean that the MIS of the scheme is linked.

Q2 - The scheme division is required to report if PFMS linkage is completed till the field level implementation agency or the last mile.

For e.g.- For a scheme implemented by city level agencies, if PFMS integration is only completed till state implementing agency level (i.e., MIS only reflects fund transfers till state level directly from PFMS) but fund flow below states i.e., to cities is not available on the MIS directly from PFMS, then integration is not completed till the last mile.

Q3 - This is a multi- select question. Scheme division to select 'Yes' for all options based on the nature of the schemes**.**

Options a, b and c are applicable to **beneficiary-oriented schemes.**

Options d and e are applicable to **industry/firm oriented schemes.**

Linkages here imply that these numbers are used as unique identifiers in the scheme MIS.

Use of Technology Theme - Explained

Q4 –

a) Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted and day time radiation at a distance, which is typically from a satellite or aircraft. Special cameras collect remotely sensed images, which help researchers 'sense' multiple things about the earth.

For e.g.- large forest fires can be mapped from space which can help plan mitigating strategies, clouds are regularly tracked by agencies to help predict the weather which may be used by the Department of Agriculture and Farmers welfare to plan for crop cultivation.

b) Night light data is the data emerging from night time lights emanating from earth captured by satellites from outer space. Different types of night light include moonlight, light directly emitted by a source (for example- building, transport) and light reflected by the ground.

For e.g. - Night light data may be useful in estimating economic activity which can be leveraged by the Ministry of Road Transport & Highways or the Ministry of Housing and Urban Affairs to plan for infrastructure projects for maximising potential.

c) Data collected from social media can help scheme division understand how people are engaging with topics of interest.

For e.g., social media data may be useful for scheme division to understand the trends in behaviour change and gain insights into the sentiments of the target beneficiaries.

d) Alternative sources of data including data generated from the private sector may be used scheme divisions to inform decision making.

For e.g.- Mobility data from private cab aggregators could be used to form better policies to regulate and ensure safety of the passengers. Economic activity data from e- commerce websites could help plan for building supply chain infrastructure in Tier 2 and 3 cities based on demand.

Use of Technology Theme - Explained

Q5 – Scheme division to select 'Yes' if they are using LGD codes as unique identifiers for storing data on geographies in their scheme MIS.

Local Government Directory (LGD) are unique codes created for each state, district, sub- district, block, village and local body by the Ministry of Panchayati Raj. All MIS must use the same codes to ensure that data is easily integrable in the future.

Q6 –

a) Machine Learning gives computers the ability to learn and predict from data without being explicitly programmed.

For e.g. The Department of Health Research can get information and target hygiene inspections by data mining of online restaurant reviews available online.

b) Artificial Intelligence (AI) refers to intelligence demonstrated by machines and can be used for enhancing the delivery of schemes.

For e.g.-The Ministry of Home Affairs can monitor social media to identify emergency situations and respond in real time.

c) Blockchain refers to having distributed ledgers or blocks of transactional data that are linked together. This structure helps in providing enhanced data security.

For e.g. - This technology can be used by the scheme division to protect sensitive data such as health records or personal information of beneficiaries of social registries.

d) Internet of Things (IoT) refers to a network of objects embedded with sensors and technologies for collecting and exchanging data over Internet.

For e.g.- The Ministry of Railways leverages IoT to automate train movement data real timewhich facilitates efficient train control functions including train movement forecasting based on accurate real time data.

e) Big data analytics refers to the use of advanced analytic techniques against very large and diverse data sets that include structured, semi- structured and unstructured data from different sources to arrive at findings for improving efficiency of the scheme/intervention. This doesn't include basic descriptive or exploratory data analysis on schematic data.

f) Drones can be used for monitoring various sectors like agriculture, infrastructure projects, commerce and logistics etc.

For e.g.- The Department of Land Resources can leverage drones to help with the digitization of land records across the country at minimal investment.

PART B

Data System Pillar (60%) - Data Security and HR Capacity (11%)

Q1 - Does the MIS follow regular antivirus updates?

(Yes/No)

Q2 - Is the MIS regularly assessed by third party auditors for the online security?

(Yes/No)

Q3 - Does the MIS/ website uses SSL certificate?

(Yes/No)

Q4 - Is the SSL certificate at least 2048-bit SHA 256 encryption or higher?

(Yes/No)

Q5 - Does the MIS use firewalls to secure access to data?

(Yes/No)

Q6 - All external communication/ 3rd party integration/ API integration for the MIS is done through encrypted channel?

(Yes/No/No external communication is established)

NEXT

PART B

Data System Pillar (60%) - Data Security and HR Capacity (11%) Cont.

Q7 - What measures are undertaken to secure sensitive/personally identifiable information?

(Single factor/ multi-factor authentication / Access control list is maintained / Data is encrypted / Data is anonymized / No such data)

Q8 - 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

Q9 - Is permission taken from user to collect, store and use their personal data?

(Yes/No)

Q10 - Is there a dedicated data quality assessment and management team for the scheme?

(Yes/No)

Q11 - Is there a dedicated data analysis team for the scheme?

(Yes/No)

NEXT

Data Security and HR Capacity Theme - Explained

Q1 - **Select 'Yes' if** the scheme division uses regular antivirus updates for the MIS software/hardware as per NIC norms.

Q2 - **Select 'Yes' if** the scheme division gets regular security audit certificate for the MIS as per NIC norms.

Q3 - **Select 'Yes' if** the scheme division has **an SSL certificate** for the MIS as per NIC norms.

Q4 - **Select 'Yes' if** the scheme division has **an SSL certificate** for the MIS as per NIC norms.

Q5 - **Select 'Yes' if** the scheme division uses **firewalls** for the MIS as per NIC norms**.**

Q6 - **Select 'Yes'** if the scheme division follows **data encryption** for communication of MIS with other digital systems as per NIC norms.

Data Security and HR Capacity Theme - Explained

Q7 - This is a multi-select question. Select all those options which are practiced by the scheme division to safeguard personal/sensitive data on its MIS. If the scheme doesn't have any such data, please select the last option only.

The personal information contained in administrative data has the potential to be used to sufficient to identify an individual. The ability to identify individuals or groups could be, for example, used to target them and lead them to be politically, socially or commercially targeted with malicious intent. To prevent the same, it is crucial to safeguard personal / sensitive data using following ways:

A. Access control is an imperative to ensure the protection of personal data. Only individuals who are authorized should be able to access the data and access should be restricted to everyone else. It is hence advised to go for authentication while authenticating access to the data when a user tries to log in. Single-Factor Authentication (SFA) is a method of providing access to users by having them present only one way of verifying their identity (usually, username and password). Multi-factor authentication uses more than one way of authentication in addition to password-based access such as OTP, lock question etc.

B. There should be a systematic process of defining access levels for each dataset or module in the MIS, which can help protect the data from unauthorized access and use. This should be clearly documented with details on which user has which access type**.**

C. Data Encryption is a critical component of any data protection strategy. It is the process of converting sensitive data into a cipher text using algorithms. This cipher text can be converted back to the original data only using a decryption key. Thus, encryption protects sensitive data from malicious actors as it cannot be read by anyone who doesn't have the key. It hence creates an expensive barrier in trying to decrypt the data. This is often used for IDs such as tax identification numbers, phone numbers, Aadhaar numbers etc.

D. 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. It is advisable to always store sensitive administrative information in an anonymized format.

Data Security and HR Capacity Theme - Explained

Q8 - This is a multi-select question. Select all those options which are practiced by the scheme division to further safeguard anonymized data on its MIS.

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 even more important to protect personal data from re-identification risks. This can be achieved by following varied measures such as:

(i) instituting a stronger security system for databases that store anonymized information,

(ii) implementation of differential privacy while sharing data for research purposes i.e., sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals,

(iii) generating and using synthetic data that exhibits the statistical properties of the raw data, without identifying details of real individuals.

Q9 - If any type of personal data is collected by the scheme division, Select 'Yes' if they take permission from user before collecting, storing and using their personal data. This consent must be documented and available with the scheme division (preferably digitally) for all beneficiaries. Some common ways of obtaining consent include signed written/digital agreements, digitally asking users for consent to collect their data when they use any app/digital service etc. Apart from taking consent to collect data, data subjects must also be informed about how their data is stored, exchanged and used. Some common methods include written/digital disclosures on websites, maintaining a digital consent management and data exchange tracker which asks for individual's permission before every instance when their data is used or exchanged.

Q10 - The scheme division to select 'Yes' if there are dedicated human resources at the division who **assess data quality** at regular points of time.

Q11 - The scheme division to select 'Yes' if there are dedicated human resources at the division who **analyze data** at regular points of time.

PART A

Data System Pillar (60%) - Data Management (3.78%)

Q1 - Does the M/D have data management guidelines/architecture, explaining how generated data is to be processed, stored, exchanged, archived and destroyed?

(Yes/No)

Q2 - (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)

Q3 - Are data ownership norms clearly defined by the M/D?

(Yes/No)

Q4 - Is there a framework for assessing the risk and value of all the data collected by the M/D?

(Yes/No)

Q5 - 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)

NEXT

Data Management Theme (Part - A) - Explained

Q1 - Select 'Yes', if the MD has prepared such a document and following aspects are considered in the document (but not limited to):

A. Data generation

i. State the purpose for which MD divisions need to collect data (explain how is collected data useful for designing/monitoring/implementing the schemes or sector of the MD)

ii. Specify what all data is generated by the MD divisions and explain how is it collected (including its origin if it is sourced from other agencies or data sources)

iii. State the expected size of data generated by MD divisions (if known)

B. Data Processing

i. Specify the ways in which this data (if collected on paper) is converted to digital format

ii. Document the ways in which raw data is cleaned up and checked for errors (eliminating redundant, incomplete or incorrect data) to ensure only good quality data is reported ahead

iii. Specify metadata standards which need to be followed to describe processed datasets

iv. Specify the formats in which processed data and metadata needs to be stored (which machine readable format to be used etc.)

C. Data exchange

i. Specify the methodology to be used by divisions for classifying datapoints as personally identifiable, sensitive, non-sensitive and openly accessible data

ii. Document which datasets or datapoints collected by MD divisions lie in which of these categories: personally identifiable, sensitive, non-sensitive and open access

iii. For personally identifiable and sensitive data which should not be shared with other stakeholders, specify what methods will be used to control access to data and ensure its security/privacy

NEXT

Data Management Theme (Part - A) - Explained

Cont…

Q1 – C) Data Exchange

iv. For non-sensitive data which can be shared on request basis, specify how will requests be received and what kind of digital methods can be used to share data

v. For open access data which should be publicly available, specify how will open access be ensured, where, when and in which formats

vi. Specify how often this list will be reviewed and by whom

vii. Specify how will data exchange tracker be maintained and consent management will be undertaken before sharing/using any personal data

D. Data storage, archival and destruction

i. State the type/list of servers which should be used by divisions for storing data (in line with MEITY's policies)

ii. Document dataset wise list of servers being used for each dataset generated by the MD

iii. Specify the methods and repository to be used by divisions for archival of historical data which is no longer used

iv. Specify when will data be categorized as 'historical' (for example, when a scheme closes, it may be categorized as historical) and the time period for which historical data should be kept beyond the life of the project

v. Specify the procedures to be followed to destroy historical data beyond its archival period

vi. Document dataset wise historical data stored by the MD, where it is stored and for how long

vii. Specify how often this list of historical datasets will be reviewed and by whom

Data Management Theme (Part - A) - Explained

Q2 - Select 'Yes', if following aspects are implemented by the MD:

A. There is a dedicated officer who is assigned the responsibility to check compliance by MD divisions against the above-mentioned data management guidelines

B. The officer regularly checks compliance via different mechanisms including review meetings, verifying documentation, verifying processes etc.)

Q3 - Select 'Yes', if the MD has prepared such a document and following aspects are considered in the document (but not limited to):

A. MD has identified data owner(s): A data owner is an individual who is in overall charge of the data. This individual must ensure that the data within their purview is correctly maintained, is accurate and of high quality. Data owners are usually senior officers who have the authority to approve data requirements, quality assessment rules and set up other data processes for its use. However, they don't have to themselves do this activity on a day-to-day basis.

So, MD must define data owner(s) for each dataset collected by MD divisions as part of this document. These may be JS level officers heading the respective scheme/intervention for which data is collected.

B. MD has identified data steward(s): Data stewards are individuals who work with data owners. They work on data management related tasks on a day-to-day basis as delegated to them by a data owner. Their responsibilities include defining data requirements, defining data quality rules and protocols, defining data exchange mechanisms, detect data quality issues to data owner etc. These protocols are then verified and approved by data owner.

So, MD must define data steward(s) for each dataset collected by MD divisions as part of this document. These may be officers handling the scheme/intervention as well as the associated dataset on a day-to-day basis and must have necessary idea about the scheme/intervention as well as technical skills to undertake this activity.

C. MD has identified data custodian(s): Data custodians are individuals who have physical or direct control over the storage and security of dataset. They are usually IT professionals who manage the security and storage infrastructure according to the policy set by the organization/data owner.

So, MD must define data custodian(s) for each dataset collected by MD divisions as part of this document. These may be officers from IT divisions of the MD.

D. The **roles and responsibilities of data owners, stewards and custodians** is clearly defined by the MD in this document.

Data Management Theme (Part - A) - Explained

Q4 - Select 'Yes', if the MD has prepared such a document and following aspects are considered in the document (but not limited to):

A. Specify the parameters to be used for assessing the value of any dataset collected by the MD. Some examples include:

i. Assess how will the data help in understanding social and economic conditions of the target group of any scheme/intervention

ii. Assess how will the data help in understanding the overall performance of the sector handled by the MD

iii. Assess how will the data be useful in improving service delivery to target groups

iv. Assess how will the data be useful for research and development purposes

v. Assess how will the data be useful in increasing government transparency and accountability

vi. Assess if any datapoint or dataset may be redundant as similar data may already be collected by some other division/data source/external party

B. Specify the parameters to be used for determining the risks of any dataset collected by the MD. Some examples include:

i. Assess what datapoints / aspects in the dataset would be classified as sensitive information, how sensitive this data would be and efforts required to safeguard this data

ii. Assess if personally identifiable information would be collected as part of the dataset, of what type and efforts required to safeguard data privacy for personal data.

iii. Assess the efforts required to ensure data security to prevent data breaches etc.

iv. Assess the efforts required to ensure data backup for disaster recovery

v. Assess any other risks (national security, high processing costs, technology risks etc.) that may be faced during data collection and reporting

Next

Data Management Theme (Part - A) - Explained

Cont….

Q4 –

C. Document the risks and value of all datasets collected by the MD in this document. This documentation can be used by the MD to implement stronger (and expensive) security mechanisms for datasets with high risk in comparison to datasets with low/zero risk. This will also help MD in reducing data collection costs by focusing more on high value data instead of low value data.

Q5 - For each dataset containing sensitive and/or personally identifiable information, ethical use of data involves ensuring that data is only used for its intended purposes. **Select 'Yes', if the MD** has prepared such a document and following aspects are considered in the document (but not limited to):

A. Specify mechanisms that must be used by MD divisions for ensuring that personal data is being collected only with due consent. Some common ways of obtaining consent include signed written/digital agreements, digitally asking users for consent to collect their data when they use any app/digital service etc.

B. Apart from taking consent to collect data, data subjects must also be informed about how their data is stored, exchanged and used. The document must hence state methods by which this information will be transparently shared with the data subjects whose personal data is being collected. Some common methods include written/digital disclosures on websites, maintaining a digital consent management and data exchange tracker which asks for individual's permission before every instance when their data is used or exchanged.

C. Specify mechanisms which must be used by MD divisions to safeguard data privacy such as anonymizing personally identifiable information, data encryption etc.

D. Specify the purposes for which personal data can be collected and ethically used by MD divisions. If any dataset is ever found to be used for unethical purposes, it must be reported as per existing policies/regulations.

E. If any machine learning algorithms are being written on collected data, it may create intentional or unintentional bias. The document must hence suggest ways to ensure that the purposes for which algorithms are used must be transparently shared with data subjects. In addition, it must outline ways to ensure that algorithms are not biased. For example, using a fully representative dataset for training the algorithm, ensuring unbiased user feedback used to train algorithms etc.

PART B

Data System Pillar (60%) - Data Management (7.02%) Cont.

Q1 - Where is MIS data stored?

(On separate servers for different schemes (distributed storage) / On central server which is used for all schemes)

Q2 - Are there mechanisms in place which can enable data sharing with other scheme divisions?

(Yes/No)

Q3 - How is MIS data stored?

(Physical servers / Cloud Storage / Hybrid servers)

Q4 - (Respond if “Cloud Storage is selected in 3, else skip this question) Which cloud service is being used?

Q5 - 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)

NEXT

Data Management Theme (Part - B) - Explained

Q1 - The scheme division to select the type of server on which scheme data is stored.

Q2 - Select 'Yes', if following aspects are implemented (but not limited to)

A. Scheme division has identified datasets and datapoints which need to be exchanged with other datasets

B. Scheme division is using unique identifiers for common fields in such datasets, wherever necessary, so that these unique IDs can be used to merge datasets whenever required

C. Scheme division has harmonized the data formats in which these datasets are stored (For example, mandating use of machine-readable data formats that should be used such as TXT, XML, JSON, CSV etc.)

D. Scheme division is not relying on manual exchange of data (either on paper or on e-file or email)

E. Scheme division is undertaking this data sharing via digital automated means (primitive methods will include file transfer via internet, secondary methods will include API and event-based messaging, advanced methods will include specialized data sharing platforms such as India Urban Data Exchange, enterprise integration etc.)

F. Scheme division is tracking what data is exchanged, where, when and with which parties using a data exchange tracker

Q3 - The scheme division to select the **type of server on which scheme data is stored**. Cloud storage offers better disaster recovery options and hence are preferred over physical servers.

Q4 - The scheme division to select the **type of server on which scheme data is stored as per NIC norms.**

Q5 - The scheme division to select the type of archival strategy used for historical data. Historical data refers to data corresponding to previous time periods which may not be actively used at present. For this purpose, the database should be well equipped to generate audit trails that have detailed information on when was the data collected and by whom, what changes were made to this data at what stages and by which users.

02

Data Driven Outcome Pillar (20% weightage)

PART A

Synergistic data use within M/D (6%)

Q1 - 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)

Q2 - Has the M/D made any implementation plan to overcome these data gaps to aid in decision making?

(Yes/No)

Q3 - 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/NA)

NEXT

Synergistic data use within M/D Theme - Explained

Q1 - Select 'Yes', if following aspects are implemented by the MD (but not limited to):

A. The M/D has identified and documented key indicators (for each scheme as well as at sector level) for which it needs data from decision making perspectives as well as the form in which this data is required (granularity, frequency, agency providing data, data format etc.)

B. The M/D has documented, for which of these indicators, data is available with the M/D or not available with the M/D and whether it is available in the needed format or not

Q2 - Select 'Yes', if following aspects are implemented by the MD (but not limited to):

A. For the indicators where data is not available with the M/D in the required format, the M/D has identified if it can get this data from some other data source or not. If the source has been identified, the M/D has made a plan for fetching data from this source at required frequency.

B. If data is not available with any other data source also, M/D has made a timebound plan for developing a data capturing mechanism themselves to collect this data at required granularity and frequency.

Q3 - Select 'Yes', if following aspects are documented and implemented by the MD (but not limited to):

A. MD has identified datasets (data cataloguing) used by all divisions within the MD, datapoints which need to be exchanged among these datasets and by whom

B. MD has developed and mandated use of unique identifiers for common fields in datasets, wherever necessary, so that these unique IDs can be used to merge datasets whenever required

C. MD has harmonized the data formats in which all these datasets are stored (For example, mandating use of machine-readable data formats that should be used such as TXT, XML, JSON, CSV etc.)

Next

Synergistic data use within M/D Theme - Explained

Cont..

Q3 - Select 'Yes', if following aspects are documented and implemented by the MD (but not limited to):

D. MD is not relying on manual exchange of data (either on paper or on e-file or email)

E. MD has mandated that exchange of data among datasets must occur via digital automated means (primitive methods will include file transfer via internet, secondary methods will include API and event-based messaging, advanced methods will include specialized data sharing platforms such as India Urban Data Exchange, enterprise integration etc.)

F. MD is ensuring that data exchange is happening using one of these digital modes

G. MD is able to track what data is exchanged, where, when, how and with which parties using a data exchange tracker

H. MD is ensuring that this data exchange is occurring while following MEITY's policies such as NDSAP (2012), Open API Policy (2014) etc.

PART A

Inter-agency data collaboration (6%)

Q1 - 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)**

Q2 - Has the M/D undertaken any of the following steps to drive these inter-agency data collaboration initiatives?

a) **SoI, MoU, Partnerships with agencies**
b) **API linking of MIS/Dashboards done to enable seamless data sharing between M/Ds**
c) **Multiple data collection processes aimed at same target groups replaced by single synergistic process**
d) **Integrated data storage/warehouses**
e) **Collaboration with other M/Ds to use their data for developing own systems**
f) **Collaboration with M/Ds to develop joint systems for data gathering/use of nonconventional data sources/emerging technologies**
g) **Collaboration with private agencies for use of non-conventional data sources or emerging technologies**
h) **Jointly conducting analysis using data from multiple M/Ds**
i) **Partnerships/Collaborations for data security related measures**
j) **Partnerships/Collaborations for capacity building of human resources**

NEXT

Inter-agency data collaboration Theme - Explained

Q1 - Select 'Yes', if following aspects are documented and implemented by the MD (but not limited to):

A. MD has identified and documented external datasets (along with names of their agencies) which are useful for the MD from decision making perspective

B. MD is not relying on manual exchange of data (either on paper or on e-file or email) to get access to such external datasets

C. MD has entered into some agreements with agencies handling this external data to regularly get access to this data which must occur via digital automated means (primitive methods will include file transfer via internet, secondary methods will include API and event-based messaging, advanced methods will include specialized data sharing platforms such as India Urban Data Exchange, enterprise integration etc.)

D. MD is tracking what external data is used, where, when, how and exchanged with which parties using a data exchange tracker

E. MD is ensuring that this data exchange is occurring while following MEITY's policies such as NDSAP (2012), Open API Policy (2014) etc.

Q2 - This is a multi-select question. Select all the options which are being used by the MD for undertaking data exchange with other external agencies.

a) SoI, MoU, Partnerships with agencies
b) API linking of MIS/Dashboards done to enable seamless data sharing between M/Ds
c) Multiple data collection processes aimed at same target groups replaced by single synergistic process
d) Integrated data storage/warehouses
e) Collaboration with other M/Ds to use their data for developing own systems
f) Collaboration with M/Ds to develop joint systems for data gathering/use of nonconventional data sources/emerging technologies
g) Collaboration with private agencies for use of non-conventional data sources or emerging technologies
h) Jointly conducting analysis using data from multiple M/Ds
i) Partnerships/Collaborations for data security related measures
j) Partnerships/Collaborations for capacity building of human resources

PART A

Prescriptive Analytics (2%)

Q1 - Has the M/D gone beyond exploratory data analysis to cross-functional prescriptive analytics?
(Yes/In-progress/No)

Q2 - (Respond if answer to 1 is yes, else skip this question) How often is this being undertaken?
(Annually/Quarterly/Monthly)

Q3 - (Respond if answer to 1 is yes, else skip this question) What is the mode in which this is being practiced?

a) Mechanisms for regular prescriptive data analysis reports to be prepared and shared with decision makers at the highest level have been instated
b) Committee formed to hold policy review meetings/review reports at regular frequencies
c) Regular policy review meetings involving all scheme divisions/sections institutionalized
d) Emerging actionables are undertaken, documented and disseminated via a separate newsletter/report/document/order etc. and tracked regularly
e) Others - Please specify how:

NEXT

Prescriptive Analytics Theme - Explained

Q1 - Select 'Yes' only if the MD regularly implements the following:

There are 4 main stages of data analysis:

a) Descriptive data analysis: which helps to understand what is presently happening

b) Diagnostic data analysis: which helps to understand why is the present condition happening

c) Predictive data analysis: which helps to understand what is likely to happen in future

d) Prescriptive data analysis: which helps to understand what present actions need to be undertaken to achieve or prevent what is going to happen in future

Prescriptive Analytics is hence the final stage of data analysis, which builds on descriptive, diagnostic and predictive analysis. It helps in preparing for future opportunities and building a plan for mitigating future risks. The accuracy of prescriptive analysis builds over time by continuously feeding and analysing new data. It relies on statistics techniques as well as artificial intelligence techniques such as machine learning.

Examples of prescriptive analytics include:

a) Adjusting the curriculum or mode of training program in real-time based on how the trainees feedback on how they are responding to each lesson

b) Using health records, analyzing which hospital patients have the highest risk of readmission so that health care providers can have special mechanisms to help such patients via patient awareness and other interventions

c) Adjusting ticket prices in railways or civil aviation sector based on availability, demand and other factors such as weather, oil prices etc.

NEXT

Prescriptive Analytics Theme - Explained

Cont…

Q1 - Real life examples include:

a) Expenditure forecasting & planning: A country used a projection model to forecast their health care expenditure. Combining various factors like population growth, the volume of services per treated case and health price inflation, the model was able to provide projections for 20 disease groups. It also provided estimates of change in funding by the national and state governments as well as the private sector.

b) Early interventions for youth offenders: By collating and analyzing data across different organizations such as the police and youth services, a country's Department allows early intervention teams to identify at-risk youths more effectively. It enables them to flag high priority cases and to inform relevant agencies who can provide targeted support and keep more children safe.

c) Emergency Planning: During floods, a country used predictive analytics to optimize their emergency response. Using a mapping program, city officials were able to see the potential damage each additional centimeter of water could bring to their community. Instead of relying on assumptions about where water was accessing properties, the city's emergency planning team used the mapping tool to input different water levels and predict which areas would be affected.

Prescriptive analytics can hence help in:

a) Better understanding communities' needs

b) Improving allocation of scarce government resources

c) Improving decision making and reducing human error

The MD should hence select 'Yes' only if it is undertaking such advanced data analysis using multiple data sources to predict future outcomes and are deciding present actions based on these predictions. They must have necessary analysis reports and decision taken orders to support the same.

Prescriptive Analytics Theme - Explained

Q2 - The MD should select the frequency at which prescriptive analytics is undertaken by them.

Q3 - This is a multi-select question.

- The MD is required to select the modes/mechanisms by which they have institutionalized prescriptive analytics and ensure that is continuously undertaken to inform policy making at the MD level. This does not include prescriptive analytics initiatives undertaken by the Ministry/Department on a one-off basis.
- The MD must hence select the first option only if they have proper mechanisms (team, processes, resources etc.) to undertake prescriptive analytics on regular basis and these analysis reports are regularly shared with policymakers in the MD.
- The MD must select the second option only if they have formed a committee of policymakers to review prescriptive analysis reports prepared by the team in the above mentioned point.
- The MD must select the third option if the committee or any policymaker is regularly holding meetings to review prescriptive analysis reports prepared by the team and decide policy action points based on this analysis**.**
- The MD must select the fourth option if the policy actionables emerging from the third point are regularly documented and disseminated across the MD for feedback and information. The status of implementation of these action points emerging from prescriptive analytics is also tracked to ensure timely execution.

PART A

Good Practices (6%)

Q1 - Describe the problem statement faced by the M/D. **(100 words)**

Q2 - 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).**

Q3 - Explain the positive impact generated with supporting evidence that indicated such impact due to the solution implemented **(100 words).**

NEXT

Good Practices Theme - Explained

Answer these questions considering the following aspects (but not limited to):

Q1 - Describe the problem statement faced by the MD. (100 words)

a) Brief context of the scheme/programme/project for which the good practice is being implemented/used

b) Challenges faced by the MD for which technology intervention was needed

Q2 - Describe how the MD has used and implemented data systems and analytics to address the issue to drive smart, near real-time and granular decisions (100 words).

a) What is the technology intervention?

b) What issue/challenge does it cater to?

c) How does it resolve the issue/challenge?

d) Who uses it and how?

e) Its features (Include details like geographical coverage, granularity, frequency, features (like geofencing etc.) as per relevance)

f) For example, the following details may also be provided to make the response exhaustive (including but not limited to): (i) Date of implementation (ii) Geographical coverage (iii) Key information captured (iv) How is it made available and who are the users?

Q3 - Explain the positive impact generated with supporting evidence that indicated such impact due to the solution implemented (100 words).

a) Share the improvements in the outcome post implementation of the intervention. For example, any quantifiable impact may be included (e.g., reduction in time by 'X%' etc.)

b) Overall contribution to improving scheme/departmental outcomes

03

Data Strategy Pillar (20% weightage)

PART A

Data & Strategy Unit(10%)

Q1 - 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 Advisor to PM dated 02.02.2021)

(**Yes/No)**

Q2 - (Respond if answer to 1 is 'yes', else skip to Q1 of next section) Who is the head of the DSU?

a) **AS and equivalent**
b) **JS and equivalent**
c) **Director and equivalent**
d) **Below Director**

Q3 - Please select the verticals established under the DSU of your Ministry/Department. (as per the D.O. letter from Advisor to PM dated 02.02.2021)

a) **Monitoring Unit**
b) **Statistics Unit**
c) **Technology Unit**
d) **Analytics Unit**

PART A

Data Driven Strategy Pillar (20%) – Data & Strategy Unit(10%)

Q4 - Please provide the percentage of filled posts in DSU

Unit	Enter % of posts filled up
Monitoring Unit	
Statistics Unit	
Technology Unit	
Analytics Unit	
Total	

Q5 - 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)

Q6 - 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)

Q7 - What is the frequency of regular review meetings/review reports?

a) Daily **b) Weekly** **c) Fortnightly** **d) Monthly** **e) Quarterly** **f) Annually**

PART A

Action Plan (10%)

Q1 - Has the M/D framed an action plan to improve its data preparedness levels? (as per the D.O. letter from Advisor to PM dated 02.02.2021)

(Yes/No)

Q2 - Does the action plan have all the sections as per the outline shared with all M/Ds? (As per D.O. letter from Advisor to PM on 02.02.2021)

(Yes/No/Partially)

Q3 - Does the action plan include data strategy for all CS/CSS schemes of the M/D?

(Yes/No/Partially)

Q4 - Are clear timelines for each action point identified under the strategy?

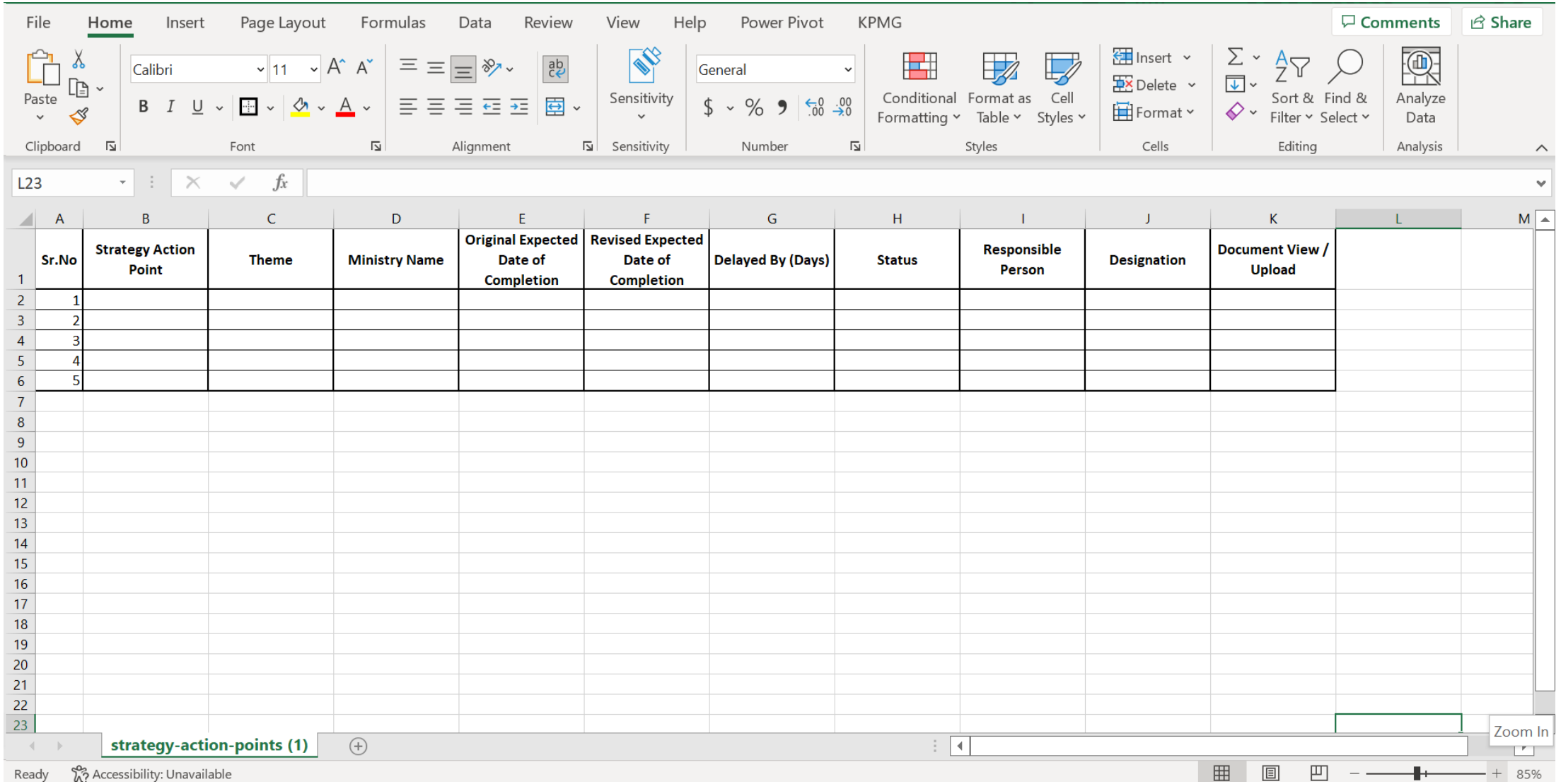
(Yes/No/Partially)

Q5 - Are the responsibilities for each action point clearly allocated to respective divisions for ensuring accountability?

(Yes/No/Partially)

Q6 - Upload the action plan in PDF format.

Action plan format



Sr.No	Strategy Action Point	Theme	Ministry Name	Original Expected Date of Completion	Revised Expected Date of Completion	Delayed By (Days)	Status	Responsible Person	Designation	Document View / Upload
1										
2										
3										
4										
5										

Thank you!

home.kpmg/in/socialmedia

KPMG Assurance and Consulting Services LLP, Lodha Excelus, Apollo Mills Compound, NM Joshi Marg, Mahalaxmi, Mumbai - 400 011 Phone: +91 22 3989 6000, Fax: +91 22 3983 6000.