# DATA QUALITY ASSESSMENT HANDBOOK

# TABLE OF CONTENTS

## LIST OF ACRONYMS

| | |
|---|---|
| CMAM | Community Management of Acute Malnutrition |
| DQ | Data Quality |
| DQA | Data Quality Assessment |
| IT | Information Technology |
| KPI | Key Performance Indicators |
| M&E | Monitoring & Evaluation |
| MS | Microsoft |
| MUAC | Mid Upper Arm Circumference |

## GLOSSARY OF TERMS

- **Data Quality Assurance** - A process for defining the appropriate dimensions and criteria of data quality, and procedures to ensure that data quality criteria are met over time. It involves a process of data profiling to unearth inconsistencies, outliers, missing data interpolation and other anomalies in the data.
- **Data Quality Assessment** – A review of programme or project M&E/IM systems to ensure that quality of data captured by the M&E/IM system is acceptable.
- **Data Quality** - refers to the condition of a set of values of qualitative or quantitative variables. It is a perception or an assessment of **data's** fitness to serve its purpose in a given context, be it in operations, decision making or planning.

# 1 INTRODUCTION TO DATA QUALITY ASSESSMENTS

A Data Quality Assessment (DQA) is a periodic review that helps donors and the implementing partner determine and document “how good the data is” and also provides an opportunity for capacity-building of implementing partners. This is a strategy that is used by organizations to assess the strengths and weaknesses of the data in relation to data quality dimensions (e.g. accuracy, reliability, consistency, relevance, accessibility and timeliness). A DQA is usually performed to fix subjective issues related to professional processes, such as the generation of accurate reports, and to ensure that data-driven and data-dependent processes are working as expected. It is important for an organization to conduct DQA on a regular basis at all stages of project cycle. DQAs can be used for purposes of;

1) verifying the quality of reported data for key indicators at selected sites,
2) the ability of data management systems to collect, manage and report quality data.
3) putting up corrective measures with action plans for strengthening the data management and reporting system and improving data quality. This gives an organization opportunity to make necessary adjustments on how they are implementing the project.
4) capacity improvements and performance of the data management and reporting system to produce quality data.

DQAs expose technical flaws in data and allow the organization to properly plan for data cleansing and enrichment strategies. This is done to maintain the integrity of systems, quality assurance standards and compliance concerns. Generally, technical quality issues such as inconsistent structure and standard issues, missing data or missing default data, and errors in the data fields are easy to spot and correct, but more complex issues should be approached with more defined processes.

The objective of this DQA initiative is to provide a common approach for assessing and improving overall data quality. The tool helps to ensure that standards are harmonized and allows for joint implementation between partners and with UNICEF supported program. The tool allows for programs and projects auditing to assess the quality of data and strengthen data management and reporting systems.

## 1.1 DATA QUALITY: ASSURANCE AND ASSESSMENT

Data quality assurance is the process of data profiling to discover inconsistencies and other anomalies in the data and it also assist in performing data cleansing activities (e.g. removing outliers, missing data interpolation) to improve the data quality. The data quality assurance suite of tools and methods include both data quality auditing tools designed for use by external audit teams and routine data quality assessment tool designed for capacity building and self-assessment. This allows implementers of programs to make necessary adjustments to the design of the program and how implementers can best execute the

whole exercise. A good data quality assurance plan, outlines strategies in the routine monitoring system to reduce:

- Estimation error and bias
- Measurement error and bias
- Transcription errors
- Data processing error

A data quality assurance plan also describes how and when internal data quality assessments will be implemented

On the other hand, a typical Data Quality Assessment approach might be identifying which data items need to be assessed for data quality and typically this will be data items deemed as critical to program operations and associated management reporting. DQA also assess which data quality dimensions to use and their associated weighting. In our proposed approach, each data quality dimension has its values or ranges representing good and bad quality data defined. It is important to note that a data set may support multiple requirements, therefore a number of different data quality assessments may need to be performed. At the same time, some assessment criteria should be applied to the data items while reviewing the results and determining if data quality is acceptable or not. Where appropriate corrective actions should to be taken like cleaning the data and improve data handling processes to prevent future recurrences. It is important to repeat these processes on a periodic basis to monitor trends in data quality.

The outputs of different data quality checks may be required in order to determine how well the data supports a particular program need. Data quality checks will not provide an effective assessment of fitness for purpose if a particular program need is not adequately reflected in data quality rules. Similarly, when undertaking repeat data quality assessments, organizations should check to determine whether business data requirements have changed since the last assessment (DAMA-UK, October 2013).

## 2 DIMENSIONS OF DQA

A *Data Quality (DQ) Dimension* is a recognized term used by data management professionals to describe a *feature* of data that can be measured or assessed against defined standards in order to determine the quality of data. (DAMA-UK, October 2013).

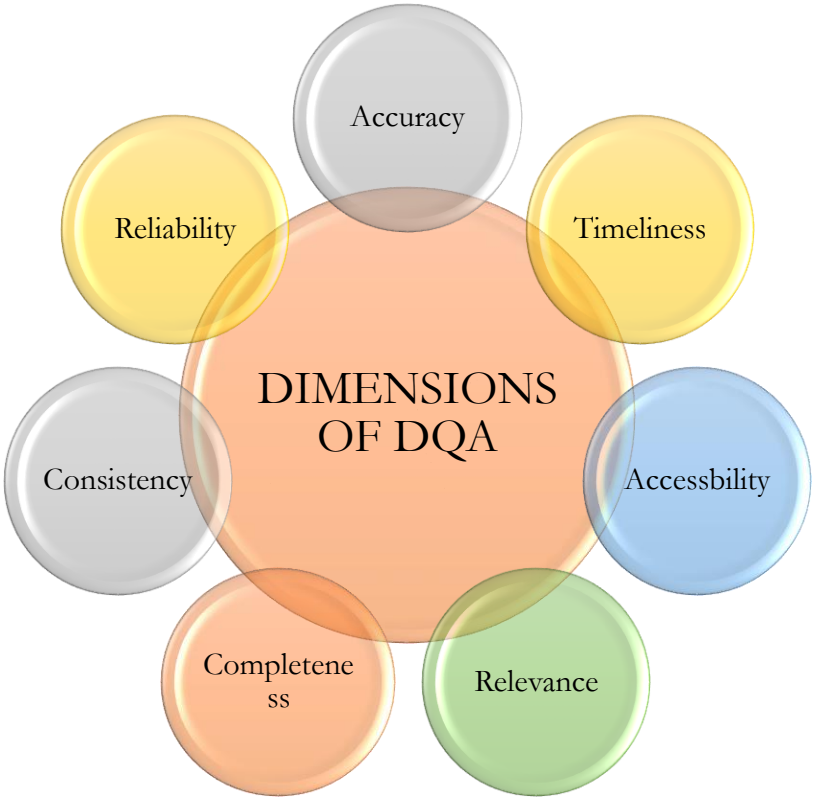


Figure 1: Dimensions of DQA

It is important to note that these dimensions are not always 100% met, for example, data can be accurate but incomplete, or it can meet all criteria except for timeliness. As managers have to make decisions based on data, it is very important to perform a short audit of data before compiling key performance indicator (KPI) results in a performance report, based on the quality dimensions presented above. Therefore, if data is not complete or there is a uniqueness issue, data users must be informed in order to keep this in mind when deciding.

Table 1*: **Data Quality Dimensions***

| S/N | Data Quality Dimension | Definition |
|---|---|---|
| **Main dimensions of data quality** | | |
| 1. | Accuracy | A measure of the correctness of data. Accurate data should represent what was intended or defined by the original source of the data. Data measure what they are supposed to measure. |

| S/N | Data Quality Dimension | Definition |
|---|---|---|
| | | -Also known as validity. Accurate data are considered correct: the data measure what they are intended to measure. Accurate data minimize error (e.g., recording or interviewer bias, transcription error, sampling error) to a point of being negligible. |
| 2 | Reliability | The data generated by a program's information system are based on protocols and procedures that do not change according to who is using them and when or how often they are used. The data are reliable because they are defined, measured and collected in the same. |
| **Sub dimensions of data quality** | | |
| 3. | Consistency | Data are consistent when the value of any given data element is the same across applications and systems. |
| 4. | Completeness | The extent to which the expected attributes of data are provided; all required data elements are captured in the database system.<br><br>-Completeness may mean that an information system from which the results are derived is appropriately inclusive: it represents the *complete* list of eligible units and not just a fraction of the list. |
| 5. | Relevance | The extent to which data are applicable and useful for the task at hand. |
| 6. | Accessibility | Accessibility is the extent to which data are available or easily retrievable. |
| 7. | Timeliness | The degree to which data are current and available for use as specified and in the time frame in which they are expected.<br><br>Data are timely when they are up-to-date (current), and when the information is available on time. Timeliness is affected by: (1) the rate at which the program's information system is updated; (2) the rate of change of actual program activities; and (3) when the information is actually used or required. |
| **Other dimensions of data quality** | | |
| 8 | Integrity | Data have integrity when the system used to generate them are protected from deliberate bias or manipulation for political or personal reasons. |
| 9 | Confidentiality | Confidentiality means that clients are assured that their data will be maintained according to national and/or international standards for data. This means that |

| S/N | Data Quality Dimension | Definition |
|---|---|---|
| | | personal data are not disclosed inappropriately, and that data in hard copy and electronic form are treated with appropriate levels of security (e.g. kept in locked cabinets and in password protected files. |
| 10 | Uniqueness | points out that there should be no data duplicates reported. Each data record should be unique, otherwise the risk of accessing outdated information increases. Looks at issues to do with a single view of the data sets and not having cases that are duplicates |

## 2.1 ACCURACY

**Table 2:** *Accuracy of Data*

| Title | Accuracy /validity |
|---|---|
| **Definition** | The degree to which data correctly describes the "real world" object or event being described.<br><br>The degree to which the data item correctly describes the object in context of appropriate real-world context and attributes. The real-world context may be identified as a single version of established truth and used as a reference to identify the deviation of data items from this reference. Specifications of the real-world references may be based on business requirements and all data items that accurately reflect the characteristics of real-world objects within allowed specifications may be |
| **Reference** | Ideally the "real world" truth is established through primary research.<br><br>However, as this is often not practical, it is common to use 3rd party reference data from sources which are deemed trustworthy and of the same chronology. |
| **Measure** | The degree to which the data mirrors the characteristics of the real world object or objects it represents. |
| **Scope** | Any "real world" object or objects that may be characterized or described by data, held as data item, record, data set or database. |
| **Unit of Measure** | The percentage of data entries that pass the data accuracy rules. |

| | |
|---|---|
| **Type of Measure:**<br><br>• **Assessment**<br>• **Continuous**<br>• **Discrete** | Assessment, e.g. primary research or reference against trusted data.<br><br>Continuous Measurement, e.g. age of students derived from the relationship between the students' dates of birth and the current date.<br><br>Discrete Measurement, e.g. date of birth recorded. |
| **Related Dimension** | Validity is a related dimension because, in order to be accurate, values must be valid, the right value and in the correct representation. |
| **Optionality** | Mandatory because - when inaccurate - data may not be fit for use. |
| **Applicability** | |
| **Example(s)** | A European school is receiving applications for its annual September intake and requires students to be aged 5 before the 31st August of the intake year.<br><br>In this scenario, the parent, a US Citizen, applying to a European school completes the Date of Birth (D.O.B) on the application form in the US date format, MM/DD/YYYY rather than the European DD/MM/YYYY format, causing the representation of days and months to be reversed.<br><br>As a result, 09/08/YYYY really meant 08/09/YYYY causing the student to be accepted as the age of 5 on the 31st August in YYYY.<br><br>The representation of the student's D.O.B.–whilst valid in its US context–means that in Europe the age was not derived correctly and the value recorded was consequently not accurate. |
| **Pseudo code** | ((Count of accurate objects)/ (Count of accurate objects + Counts of inaccurate objects)) x 100 |

| | |
|---|---|
| | Example: (Count of children who applied aged 5 before August/YYYY)/ (Count of children who applied aged 5 before August 31st YYYY+ Count of children who applied aged 5 after August /YYYY and before December 31st/YYYY) x 100 |

## 2.2 RELIABILITY

**Table 3:** *Reliability of Data*

| | |
|---|---|
| **Title** | Reliability |
| **Definition** | The degree to which the result of a measurement, calculation, or specification can be stable consistent or repeatable over time or by different people. |
| **Reference** | Established through primary research in relation to data quality. |
| **Measure** | It is a measure of the stability or consistency of test scores. |
| **Scope** | Characterized or described by data, held as data item, record, data set or database. |
| **Unit of Measure** | |
| **Type of Measure** | • Overall consistency of a measure |
| **Related Dimension** | Consistency and validity are related dimension because, in order for data to be reliable it has to be consistent and accurate. |
| **Optionality** | Mandatory because - when data is not reliable - data may not be fit for use. |
| **Example(s)** | Frontline workers in health use MUAC tapes to measure malnutrition risks in under-five children and pregnant women.<br><br>This means that different frontline workers are expected to come up with same results after measuring same child. |
| **Pseudo code** | N/A |

## 2.3 CONSISTENCY

**Table 4***: Data Consistency*

| **Title** | **Consistency** |
|---|---|
| **Definition** | The absence of difference, when comparing two or more representations of a thing against a definition.<br><br>This measure represents the absence of differences between the data items representing the same objects based on specific information requirements |
| **Reference** | Data item measured against itself or its counterpart in another data set or database. |
| **Measure** | Analysis of pattern and/or value frequency. |
| **Scope** | Assessment of things across multiple data sets and/or assessment of values or formats across data items, records, data sets and databases. Processes including: people based, automated, electronic or paper. |
| **Unit of Measure** | Percentage. |
| **Type of Measure:**<br><br>• **Assessment**<br>• **Continuous**<br>• **Discrete** | Assessment and Discrete. |
| **Related Dimension(s)** | Validity, Accuracy and Uniqueness |
| **Optionality** | It is possible to have consistency without validity or accuracy. |
| **Example(s)** | School admin: a student's date of birth has the same value and format in the school register as that stored within the Student database. |
| **Pseudo code** | Select count distinct on 'Date of Birth' |

## 2.4 COMPLETENESS

**Table 5:** *Accuracy of Data*

| **Title** | **Completeness**[1] |
|---|---|
| **Definition** | The proportion of stored data against the potential of "100% complete" |
| **Reference** | Business rules which define what "100% complete" represents. |
| **Measure** | A measure of the absence of blank (null or empty string) values or the presence of non-blank values. |
| **Scope** | 0-100% of critical data to be measured in any data item, record, data set or database |
| **Unit of Measure** | Percentage |
| **Type of Measure:**<br>• **Assessment**<br>• **Continuous**<br>• **Discrete** | Assessment only |
| **Related dimension** | Validity and Accuracy |
| **Optionality** | If a data item is mandatory, 100% completeness will be achieved, however validity and accuracy checks would need to be performed to determine if the data item has been completed correctly |
| **Example(s)** | Parents of new students at school are requested to complete a Data Collection Sheet which includes medical conditions and emergency contact details as well as confirming the name, address and date of birth of the student.<br><br>Scenario: |

[1] Measure critical data for completeness first; incompleteness in non-critical data may not matter to the business.

| | |
|---|---|
| | At the end of the first week of the Autumn term, data analysis was performed on the 'First Emergency Contact Telephone Number' data item in the Contact table.<br><br>There are 300 students in the school and 294 out of a potential 300 records were populated, therefore 294/300 x 100 = 98% completeness has been achieved for this data item in the Contact table. |
| **Pseudo code** | Count 'First Emergency Contact Telephone Number' where not blank in the Contact table/ count all current students in the Contact table. |

## 2.5 RELEVANCE

**Table 6:** *Accuracy of Data*

| | |
|---|---|
| **Title** | Relevance |
| **Definition** | an assessment of data's fitness to serve its purpose in a given context. |
| **Reference** | Established through primary research in relation to data and data quality |
| **Measure** | the degree to which something is related or useful to what is happening or being talked about<br><br>it is the closeness between data consumer need and data provider output. |
| **Scope** | An aspect of (-->) data quality: a level of consistency between the (-->) data content and the area of interest of the user |
| **Unit of Measure** | It is measured as percentage of all data required divided by all data provided. [100% is best] |
| **Type of Measure:**<br><br>• **Assessment** | Assessment, e.g. primary research or reference related to what is happening. |
| **Related Dimension** | |

| | |
|---|---|
| **Optionality** | |
| **Applicability** | |
| **Example(s)** | The ability to use data with maximum efficiency. Not having to sort through information you don't need.<br><br>Having beneficiary's database covering more communities and using queries one can be able to retrieve the exact data they are looking for |
| **Pseudo code** | |

## 2.6 ACCESSIBILITY

**Table 7:** *Accessibility of Data*

| **Title** | **Accessibility** |
|---|---|
| **Definition** | It is a generic term referring to a process which has both an IT-specific meaning and other connotations involving access rights in a broader legal and/or political sense. In the former it typically refers to software and activities related to storing, retrieving, or acting on data housed in a database or other repository. |
| **Reference** | |
| **Measure** | The ease at which data stored can be easily retrieved or manipulated. |
| **Scope** | Extent to which a consumer or user can obtain a good or service at the time it is needed. |
| **Unit of Measure** | Yes – Completely Accessible, No, Partially Accessible |
| **Type of Measure:** | Norminal |
| **Related Dimension** | Transparency |
| **Optionality** | |
| **Applicability** | |

| | |
|---|---|
| **Example(s)** | Users who have data access can store, retrieve, move or manipulate stored data, which can be stored on a wide range of hard drives and external devices. accessibility<br><br>At community level volunteers submit reports to frontline workers who later report to district officers. In this scenario if anyone in this chain holds data will affect the whole reporting process |
| **Pseudo code** | |

## 2.7 TIMELINESS

**Table 8:** *Timeliness of Data*

| **Title** | **Timeliness**[2] |
|---|---|
| **Definition** | The degree to which data represent reality from the required point in time. |
| **Reference** | The time the real-world event being recorded occurred. |
| **Measure** | Time difference |
| **Scope** | Any data item, record, data set or database. |
| **Unit of Measure** | Time |
| **Type of Measure:**<br>• **Assessment**<br>• **Continuous**<br>• **Discrete** | Assessment and Continuous |
| **Related dimension** | Accuracy because it inevitably decays with time. |
| **Optionality** | Optional dependent upon the needs of the business. |

[2] Each data set will have a different proportion of volatile and non-volatile data as time acts differently on static and dynamic records

| | |
|---|---|
| **Example(s)** | Tina Jones provides details of an updated emergency contact number on 1st June 2013 which is then entered into the Student database by the admin team on 4th June 2013. This indicates a delay of 3 days. This delay breaches the timeliness constraint as the service level agreement for changes is 2 days. |
| **Pseudo code** | Date emergency contact number entered in the Student database (4th June 2013) minus the date provided (1st June 2013) = a 3 Day delay. |

# 3 KEY COMPONENTS OF DQA

The DQA tool is composed of three components: (i) systems assessments (ii) compliance to data encoding and submission standards; (iii) data verification;

## 3.1 DATA MANAGEMENT AND REPORTING SYSTEM ASSESSMENT

This enables qualitative assessment of the relative strengths and weaknesses of functional areas of a data management and reporting system at all levels. The purpose of assessing the data management and reporting system is to identify potential threats to data quality posed by the design and implementation of data management and reporting systems. The systems assessment questions are asked to the persons responsible for managing data and preparing reports at the different levels.

Systems assessment sections are found as Part 2 in the Service Point, Governorate Level Aggregation Sites, and National Level M&E Unit sheets in the MS Excel workbook. The systems assessment section of the DQA tool includes the following five functional areas:

1. **M&E Structure, Functions and Capabilities:** Availability of M&E organizational structure, training plan, and trained data management staff.
2. **Indicator Definitions and Reporting Guidelines:** Availability of indicator definitions and guidelines on reporting i.e. when, where and to whom reports should be sent.
3. **Data Collection and Reporting Forms and Tools:** Availability, appropriateness and utilization of standard data collection and reporting tools.
4. **Data Management Processes:** Availability of data quality controls, data back-up procedures, confidentiality of personal data, and feedback on quality of reported data.
5. **Links with National Reporting System:** Use of / adherence to national reporting system i.e. data tools, reporting channel, reporting deadlines, and sites identification.

Using the Excel DQA tool, scores are generated for each functional area. The scores are an average for all responses to the qualitative questions in each functional area, with each question coded 3 for "yes, completely," 2 for "partly," and 1 for "no, not at all." The scores are intended to be compared across functional areas to guide program implementers on which systems strengthening activities to prioritize. It would be reasonable to consider investing more resources in an area whose score is low compared to that whose score is relatively high.

In order to complete both the Data Verification and Systems Assessment parts of the DQA tool, the assessment team will have to make some observations, do a recount, and ask questions to the appropriate respondents. The emphasis of DQA is to verify the quality of reported data and identify potential challenges to data quality created by the data management and reporting system. It is intended to improve the quality of reported data and systems but not to change already reported data (NTD-Guidelines, December 2013).

The Program Information Management System of a program/project provides the overall procedures in collecting, processing, and managing data. As mentioned earlier, quality data is highly dependent on the systems in place. Strong system should produce better quality of data. The *systems assessment* will look into the functionality and effectiveness of the following:

- Competencies of M&E officers and other staff involved in data collection and management
- Capacity building and technical assistance
- Data collection, processing and management
- Use of paper-based forms and computer-based templates
- Encoding and submission
- Internal quality control
- Data utilization and reporting
- Storage and retrieval

Moreover, the objective of this assessment is to help the management understand the underlying limitations and problems encountered during data collection, processing and management, determine possible area or source of data errors, identify measures to improve the capabilities of staff involved in the process and strengthen data management at all levels.

## 3.2 COMPLIANCE TO DATA ENCODING AND SUBMISSION STANDARDS

Compliance programs are seen as an effective mechanism to assure compliance with regulations and minimize risk of fraud. A coding compliance program should be a key component of any program. That is by complementing, not conflicting with, the compliance program. This provides timely and complete data to management, partners and other stakeholders thereby facilitating better and informed decision-making. This component deals with the completeness and timeliness of submission of data from the partner/district

up to the national level based on the reporting requirements and standards set and provided to all levels. The partner, district, regional and national teams will be evaluated on the level of their compliance to these standards.

## 3.3 DATA VERIFICATION

Data Verification is a process in which different types of data are checked for accuracy and inconsistencies after data migration is done. It helps to determine whether data was accurately translated when data is transferred from one source to another, is complete, and supports processes in the new system. During verification, there may be a need for a parallel run of both systems to identify areas of disparity and forestall erroneous data loss. An example of Data Verification is double entry and proofreading data. Proofreading data involves someone checking the data entered against the original document.

During project implementation different forms, templates and documents are used, completed and collected by stakeholders and other program staff to capture and document project activities implemented in various communities. Given the volume of data and information being collected, the management still gives high regard and importance to quality and providing accurate and consistent data to its stakeholders. Data verification will look into the accuracy and consistency of data from the source document, crosschecking the reported information with the paper-based forms, templates and other post documentations. It will identify, track and resolve inconsistencies and errors in the database. This exercise can be time consuming and costly.

Data verification sections are found as Part 1 in the Service Delivery Site (SDSs), District Level, and National Level M&E Unit sheets in the DQA MS Excel Workbook.

In the Service Delivery Site sheets, the section is divided into three subparts, as shown below

- **Documentation Review:** qualitative description of availability and completeness of PC data sources for the indicators being assessed.
- **Recounting Reported Results:** for each indicator being assessed, a recount for all the sampled Service Delivery Sites (typically villages, schools, health centres) will be done and results compared with what was reported. It is very important to ensure that source documents used by all the community volunteers and or frontline workers e.g. teachers within the selected Service Delivery Sites in this case schools are included in the recount. In addition to using the officially recognized source documents, there are instances when the community volunteers or frontline workers, record services they provide using improvised tools e.g. exercise books for teachers. Data recorded using such tools should be included in the recount. Only documented information should be included in the recount (i.e., not verbal reports of the indicator values). In case of discrepancies between reported and recounted results, reasons for the differences will be noted. Recounting of reported

results will be done at all the intermediate aggregation levels and at national level. At least two individuals should carry out the recount separately and compare the recounted values, in order to ensure accuracy of the DQA results.

- **Cross-check reported results with other data sources:** reported results may be cross-checked against other data sources such as school registers and drug inventory. It may not be possible to carry out cross-checks for every indicator since in some cases there might be no data sources for conducting the cross-checks.

## 4 THE PROPOSED DATA QUALITY ASSESSMENT TOOL

The DQA MS Excel Tool is designed to verify that appropriate data management systems are in place at all levels, verify the quality of reported data for key indicators at selected site and contribute to M&E system strengthening and capacity building.

The DQA Data Collection Tool presents a **Data Verification and System Assessment tool** with – 1. *Data Verifications*; recounting reported results, and, reporting performance 2. *System assessment*; I – M&E capacities roles and responsibilities, training, indicator definitions, data reporting requirements, data-collection and reporting forms and tools, data management processes and data quality controls, and a data reporting system. in the form of detailed questions (Mulongo, 2017).

## 5 DETAILED STEPS FOR CONDUCTING DQA

Preparation for the DQA exercise is a very important phase of conducting the DQA. Preparation for the DQA exercise will involve, among other steps, identifying indicators to be assessed, selection of sites, assembling the field team, and putting together the necessary documentation. It is imperative that different stakeholders are involved in the preparations.

### STAGE I: PREPARATION WORK BY THE DQA TEAM

The processes involved in this stage can be described in a table below

| Step | Comment / level |
|---|---|
| 1. Select indicators to be assessed and their time period for which they will assessed | This should be done as part of the preparation. It is advisable to use the most recent PCA round for which treatment reports have been compiled. |
| 2. Obtain necessary authorization to conduct the assessment | Authorization could be from relevant national and sub-national / district authorities. |

| Step | Comment / level |
|---|---|
| 3. Prepare needed documentation | This is part of preparation. It is important to sort out all required documentation before carrying out the whole exercise. These could be registers, forms, or program/project database. |
| 4. Assemble assessment team | The assessment team are supposed to be briefed before rolling out the exercise so that they have an understanding of the organizational systems, methods of data collection and know what is expected of them. |
| 5. Select sites for assessment | Sites should be selected following guidance given in these guidelines. Avoid biased selection of sites. |
| 6. Prepare for on-site visits | Preparations may include timing, constitution of teams, training and logistics |

### Detailed Preparatory Steps

**1.** Determine purpose of the DQA

You can find more detailed explanation of the various reasons for conducting a DQA in the earlier section, but as a reminder, you could conduct an DQA for:

- Routine data quality checks as part of on-going supervision
- Initial and follow-up assessments of data management and reporting systems
- Strengthening program staff's capacity in data management and reporting
- Preparation for a formal data quality audit
- External assessment by partners of the quality of data (UNAIDS, July 2008)

**2.** Select levels and sites to be included (depending on the purpose and resources available). Once the purpose has been determined, the second step in the DQA is to decide what levels of the data management and reporting system will be included in the assessment - service sites, intermediate aggregation levels, and/or central M&E unit. The levels should be determined once the appropriate reporting levels have been identified and "mapped" (e.g., there are 100 sites providing the services in 10 districts. Reports from sites are sent to districts, which then send aggregated reports to the M&E Unit). In some cases, the data flow will include more than one intermediate level (e.g. regions, provinces or states or multiple levels of program organizations).

It is not necessary to visit all the reporting sites in a given program to determine the quality of the data. Random sampling techniques can be utilized to select a representative group of sites whose data quality is indicative of data quality for the whole program. Depending on the volume of service of the program (e.g. number of children in CMAM program), the number of service delivery sites and the quality of the data, as few as a dozen sites can be assessed to obtain a reasonable estimate of data quality for the program.

Precise measures of data accuracy are difficult to obtain for an entire program using these methods. "Reasonable estimates" of data accuracy are generally sufficient for the purposes of strengthening data quality, capacity building or preparing for external auditing.

**3.** Identify indicators, data sources and reporting period. The DQA is designed to assess the quality of data and underlying systems related to indicators that are reported to programs or donors to measure success in program areas related to specific diseases during specific reporting periods.

## STAGE 2: COLLECT INFORMATION AND CONDUCT DATA VERIFICATION (FOR SELECTED INDICATORS)

**4.** Conduct site visits. Sites should be notified prior to the visit for the data quality assessment. This notification is important in order for appropriate staff to be available to answer the questions in the checklist and to facilitate the data verification by providing access to relevant source documents. During the site visits, the relevant sections of the appropriate checklists in the MS Excel file are completed (e.g. the service site checklist at service sites). These checklists are completed during or immediately following interviews of relevant staff and reviews of site documentation.

## STAGE 3: COMPLETING THE DQA - SUMMARIZE INFORMATION, PROVIDE JUDGMENT ON ASSESSMENT RESULTS, DOCUMENT NEXT STEPS TO STRENGTHEN DATA QUALITY

The third phase of conducting the DQA is to compile the results. The data should be entered into the Excel tool if the initial data collection was performed using paper-based versions of the tool. The Excel tool will generate summary information on the availability, timeliness and completion of the data at the various reporting levels; verification factors to describe the accuracy of the reported data; and evidence the quality of the program's data management and reporting system, such as precision, integrity, and confidentiality. These results should be incorporated into a DQA Action Plan and should be the basis of drafting preliminary findings and recommendations. After presentation of the preliminary findings, the report should be finalized and the DQA documentation compiled (NTD-Guidelines, December 2013).

**5.** Review outputs and findings. The outputs from the DQA described above should be reviewed for each site visited. Site-specific summary findings in the form of recommendations are noted at each site visited.

**6.** Develop a system strengthening plan, including follow-up actions. Given the findings and recommendations for each site, an overall action plan is developed.

## 7. TOOLS FOR CONDUCTING A DQA

The DQA Tool in Excel can be printed out or used on a computer for the proposed DQA

## 8 REFERENCES


DAMA-UK. (October 2013). *The Six Primary Dimensions for Data Quality Assessment-Defining data quality dimensions.* United Kingdom: The premier organisation for data professionals in the UK.

Mulongo, G. (2017). *Data Quality Assuarance- Data Collection Tool.* BNFB- Data management and reporting system.

NTD-Guidelines. (December 2013). *Data Quality Assessment for Neglected Tropical Diseases: Guidelines for Implementation- working draft for field testing.* NTD.

UNAIDS, P. U. (July 2008). *Routine Data Quality Assessment Tool-guidelines for implementation for HIV, TB and Malaria Programs.* GLOBAL FUND to fight AIDS, TB and Malaria.