



Malaria Quarterly Report Format Standardization

Updated July 2022

The [Malaria Quarterly Report](#) (QR) is the primary data resource in [M-DIVE](#) for understanding malaria incidence, burden, and progress metrics in [PMI](#)-supported countries. It includes data provided by countries each quarter, based on what they have available from their existing data collection systems. This document describes how data submitted from different country systems and formats are pre-processed into one consistent format in M-DIVE, so that additional processing, quality control, and reporting steps can be applied consistently and efficiently.

Format standardization:

Data that are submitted to M-DIVE come from many different countries and source data systems. Data come to M-DIVE in different layouts (e.g. “long” format, where each row represents one time point per subject, versus “wide” format, where a subject’s responses are represented by different columns within the same row) and file formats including Excel spreadsheets or JSON data pushed by an in-country team through a [DHIS2](#) to M-DIVE application programming interface (API) connection.

In order for the M-DIVE automated tools to work with these data, the data must first be standardized into a consistent format for all countries, indicators, and time periods. The process is as follows:

1. The Civis team manually inspects the data to identify file format, define key parameters and note any non-standard submissions.
2. The data then continues through the [Civis-developed automated pipeline](#) to a standardized format.
3. These standardized data are stored in a table that houses data in a long format where each row represents one time point per geographic area, which allows Civis to append rows as variable_name + variable_value pairs in a way that we wouldn't be able to with wide format tables.

To conduct format standardization, a Civis team member is designated as the “manual data processor” responsible for taking the data for a given country and QR round through two stages of data preparation that must occur before the automated parts of the M-DIVE pipeline can proceed: 1) import the “raw” data into the M-DIVE Platform, and 2) reshape the data and appending it to a warehouse table to prepare it for other transformations in the pipeline.

Import Submitted Data to M-DIVE

Reshape Data



- Visually inspect the submitted data
- Modify extraction code
- Run the code to extract the selected ranges for headers and columns
- Extract the data from excel files and save tables to M-DIVE

- Check the content and format of the data to select the steps needed to reshape the data.
- Reshape the data by defining a set of values needed for reshaping
- Run gatekeeper function to flag potential issues, fix those issues, and to re-confirm that all the checks are cleared
- Add the data to the data warehouse table. Add metadata elements to track data lineage and priority

Import “Raw” Data to M-DIVE

The first step is to import the raw, unprocessed tables into the Civis platform where they are stored as backup and source of truth. The tables are extracted from Excel sheets submitted by each country and processed via code in M-DIVE.

Data may be submitted with varying formats and structures. Common differences include multi-index rows where multiple columns act as a row identifier, varying hierarchical data where multiple rows act as a header identifier, QR indicator variables are listed in rows while months are the columns, etc.

To ensure that automated processing will proceed properly, the designated manual data processor will confirm that the submitted data have been properly imported by:

1. Visually inspecting the submitted data for any variations from the standard QR template.
2. Modifying the extraction code to ensure the correct range of spreadsheet cells is selected and adjust to accommodate different hierarchical structures
3. Running the code to extract the selected ranges for headers and columns
4. Extracting the data from Excel files and saving as a database table in M-DIVE. Each distinct combination of country + QR round + topic (e.g., HMIS, LMIS, population) is stored as its own table with a uniquely identifiable name.

Reshape Data

The tables for each country and QR round of data submission are subsequently reshaped (i.e. converted from a 1 row per value to a 1 row per indicator format), to ensure consistency among the tables so that they can enter the pipeline and undergo other standardizations and transformations.



1. Civis checks the content and format of the data in order to select the steps needed to reshape the data. The three types of data inputs that are handled by manual data processors are: Campaign Insecticide-Treated Net (ITN) data, Campaign Indoor Residual Spraying (IRS) data, and Non-Campaign data (HMIS, LMIS, SBC, Meta, and Elimination data).
2. Once Civis identifies the type of data they are dealing with, Civis defines a set of values needed for reshaping such as source data table, temporal variables, geographical variables, and others based on the data input.
3. After reshaping, a “gatekeeper” function is run to highlight common data format issues, such as null values, mismatched data types, and/or duplicated values, and to re-confirm that all the checks are cleared. In some cases in-country teams are contacted to resolve data issues.
4. If the “gatekeeper” function checks do not raise any significant warnings (e.g. columns have the incorrect data type, time step values are out of bounds), then the manually processed data are ready to be loaded into the automated processing stage of the QR pipeline. The data are exported and appended to the data warehouse table. During this step some metadata are appended as well to help us track data lineage (e.g., source schema and source table to determine where the data is coming from) and priority (e.g., date when data were reported and date when data were submitted to the warehouse table to determine the most recent version of data to surface in the QR).

If you have any questions, please submit them to Alaine Knipes (aknipes@cdc.gov) and Misun Choi (mchoi@usaid.gov) on the PMI Surveillance & Informatics team, or to support@civisanalytics.com. If you would like to learn more about additional topics on M-DIVE, please click the links below for further reading.

Important Resources

- [Malaria Quarterly Report Data Pipeline Overview: How Quarterly Report Data Are Processed in M-DIVE](#)
- [Malaria Quarterly Report Quality Control Processes Overview in M-DIVE](#)
- [Malaria Quarterly Report Time Standardization](#)
- [Malaria Quarterly Report Indicator Standardization](#)
- [Malaria Quarterly Report Geographic Standardization](#)
- [Malaria Quarterly Report](#) (M-DIVE access required)
- [M-DIVE Help Center](#) (M-DIVE access required)