Monitoring & Evaluation Framework for Malaria Elimination Digital Solutions

July 2018

CLINTON HEALTH ACCESS INITIATIVE

UiO : University of Oslo

Vital Wave℠

CONTENTS

1. Introduction and Objectives

As countries have moved to strengthen their surveillance systems as a core malaria intervention, new digital solutions have increasingly been adopted by malaria programs. These digital solutions are being used to centralize malaria data, increase reporting timeliness and completeness, enable field-based data collection, and generate automated analytics. These features are particularly important in malaria elimination settings, where fine-resolution data and rapid reporting are critical in the efforts to halt transmission.

As a core intervention, surveillance systems should be routinely analyzed and improved. As part of the Digital Solutions for Malaria Elimination (DSME) Grant, this document was developed to serve as a framework for Monitoring and Evaluation (M&E) of digital solutions for malaria surveillance. This framework was developed for the DSME grant, but can be applied to other digital surveillance systems.

To this end, the framework can be used by any surveillance programs using digital solutions to:

1. Understand how digital solutions are used for surveillance, in order to (a) inform system improvements, (b) inform supervision and trainings
2. Quantify the impacts of digital solutions, particularly on reporting rates and data quality
3. Understand the overall usefulness of digital solutions on the disease surveillance and health interventions

This framework is designed to function as a guide to support surveillance programs in developing a robust M&E plan. This framework does not cover plan specifics, which should be developed in country and include definition of indicator targets, identification of synergies with other M&E protocols, and identification of resources to support routine collection of M&E data.

The document is divided into different sections for guidance on M&E indicator categories that cover various parts of a surveillance system. Each indicator category entails distinct activities, effort and data collection methods. It is not expected that a program has to analyze each indicator in order to conduct successful M&E. Rather, individual indicators can be selected from each category based on relevance, priority and feasibility. This being said, ideally indicators from all categories should be represented in order to form an comprehensive evaluation. The indicators are described bellow and can also be found in Appendix A.

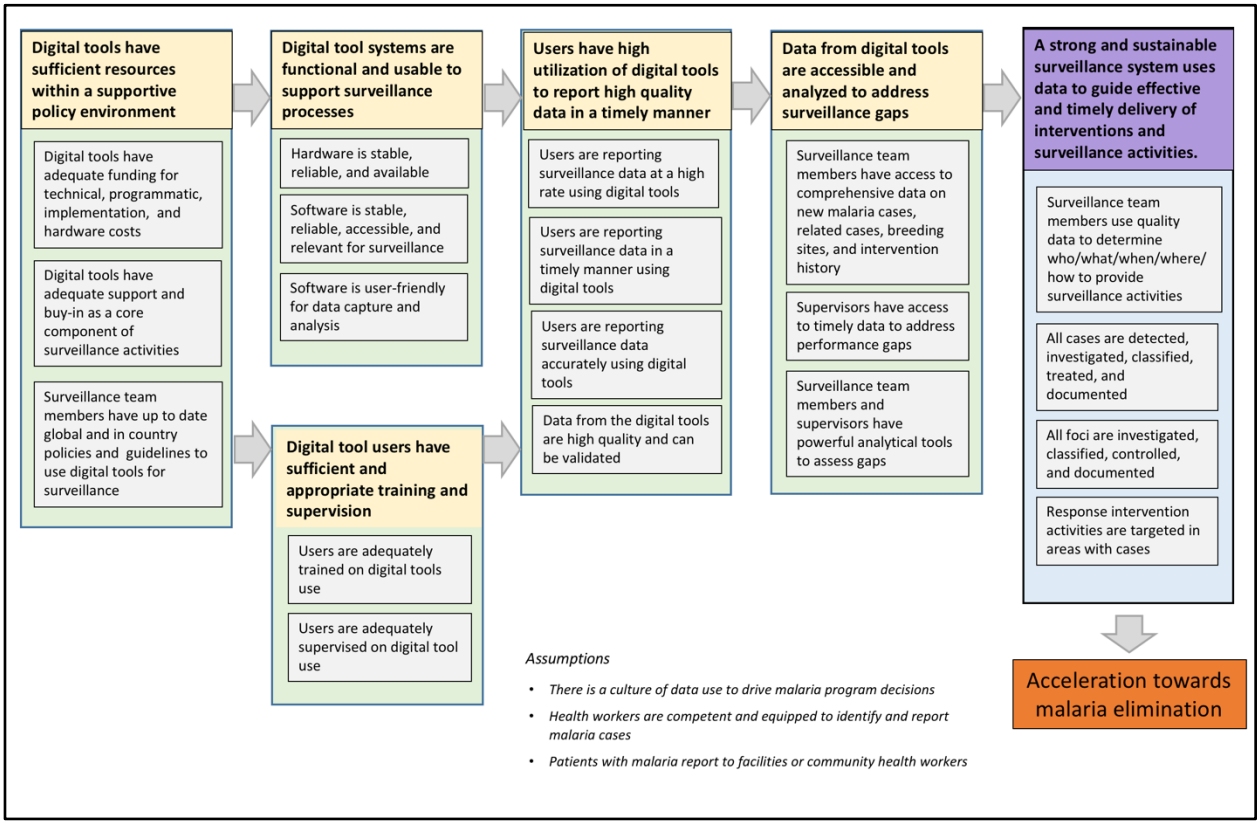
Resources leveraged include:

- WHO M&E Toolkit for Digital Health Interventions: http://apps.who.int/iris/bitstream/10665/252183/1/9789241511766-eng.pdf
- Methodology for Measuring Usability: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html
- Data Quality Assessment Guidelines: http://pdf.usaid.gov/pdf_docs/Pnadw118.pdf
- mHealth Evidence Reporting and Assessment (mERA checklist): http://www.bmj.com/content/352/bmj.i1174

2. Theory of Change

A theory of change for how digital solutions can improve malaria surveillance is captured in Figure 1.

Figure 1. Theory of Change



With the end goal of improved surveillance systems that support acceleration towards malaria elimination, this theory of change sets up key learning questions to assess the inputs, outputs, outcomes, and impact of digital solutions in achieving that ideal surveillance system along the way. Working backwards from the outcome of a strong and sustainable surveillance system, key learning questions include:

1. Does the use of data from a strong and sustabile surveillance system lead to improved delivery of interventions and surveillance activities?
2. Do improvements in quality, accessibility and analytics of data lead to the use of data for surveillance?
3. Do utilization of digital solutions lead to improvements in quality, accessibility and analytics of data?
4. Do improvements in functionality and usability lead to increased utilization of digital tools?
5. Does sufficient and appropriate training and supervision lead to increased utilization of digital tools?
6. Do sufficient resources within a supportive policy environment lead to improved functionality and usabily of digital tools?

7. Do sufficient resources within a supportive policy environment lead to additional investments in proper training and supervision on the use of digital tools?

The result of each learning question can then identify constraints along the theory of change that may need targeted improvements in order to achieve overaching surveillance goals.

This framework proposes nine categories of indicators to measure the outcomes of using digital solutions for malaria surveillance: *Outcomes, Surveillance, Data Quality and Reporting, User Profile and Usage, User Engagement and Usability, System Performance, Hardware Performance, Governance,* and *Sustainability*. These indicator categories are designed to answer each learning question (Table 1).

Table 1. Indicators to answer key learning questions

Key Learning Questions	Indicator Categories
Does the use of data from a strong and sustabile surveillance system lead to improved delivery of interventions and surveillance activities?	• Outcomes • Surveillance
Do improvements in data quality and accessibility lead to the use of data for surveillance?	• Surveillance • Data quality and reporting
Do utilization of digital solutions lead to improvements in quality, accessibility and analytics of data?	• User profile and usage • User engagement and usability • Data quality and reporting
Do improvements in functionality and usability lead to increased utilization of digital tools?	• System performance • Hardware performance • User engagement and usability
Does sufficient and appropriate training and supervision lead to increased utilization of digital tools?	• User engagement and usability • Sustainability and scalability
Do sufficient resources within a supportive policy environment lead to improved functionality and usabily of digital tools?	• Sustainability and scalability • Governance
Do sufficient resources within a supportive policy environment lead to additional investments in proper training and supervision on the use of digital tools?	• Sustainability and scalability • Governance

3. Data Collection and Analysis

A. Indicator Prioritization

The list of proposed indicators within each category can be found in Section 4, with suggested levels of priority based on how effective each indicator can be in assessing digital system performance:

- High: Assessment of an indicator listed as high priority will be very important to identify whether digital systems are functioning as expected, and what components of the systems will need to improve in order to ensure the digital systems are usable and useful.

- Medium: Assessment of an indicator listed as medium priority can be useful to assess strengths and weaknesses of a digital system. These indicators may not necessarily suggest need for immediate changes, but can point to areas for additional investigation.

- Low: Assessment of an indicator listed as low priority can be useful for monitoring, but should not be used in isolation as it may be a minor component of the system as a whole.

Collection effort is also noted with High, Medium, and Low but are not quantified by hours due to differences that may exist across countries and systems. The rating provided for collection effort can be assessed relative to effort for other indicators on the list.

B. Data collection and analysis methods

Data collection methods will need to leverage both quantitative and qualitiative techniques to capture a wide range of perspectives on the inputs, outputs, outcomes, and impact of the digital solutions. Quantiative methods may include:

- **Database Analysis:** Data collection via database extraction can allow for analysis of data already being collected within the database (e.g. – case level data, number of form submissions, etc) to understand system performance and impact. Examples of indicators that can be collected using this method can be found in the sections on User Engagement and Usability (Section 3D), Data Quality and Reporting (Section 3E), and Surveillance Process Impact (Section 3F). The data extraction method can vary from system to system. There may be a simple export process within the system, or the evaluator may be required to use some basic SQL commands. Once extracted, the data can be analyzed using whatever platform the user prefers, such as Excel. This method assumes that the data elements of interest needed to calculate the corresponding indicators are already being routinely collected and stored in the database. If not, additional data elements can always be added for the benefit of future and ongoing M&E efforts.
- **Sample Test:** A sample test entails collecting data across a representative sample. Examples of indicators using this method can be found under Hardware (Section 3B) whereby a sample of devices are tested to see how long it takes to charge them and how long their charge holds. This data collection method is useful in such scenarios where no routine data is being collected.
- **System Analytics**: System/User Analytics are designed to be routinely collected data points around the number and types of users who access the system, when users access the system and for what duration, and how users interact with the system. Examples of indicators that can be collected using this method can be found in the section on User Profile and Usage (Section 3C) and User Engagement and Usability (Section 3D). The system must be set up to collect these data points in advance, however the set-up effort is relatively low, and once completed these data points will be routinely available for all future and ongoing M&E efforts. During set-up, this data can be made readily available through the user interface or through a data export option, and third-party tools such as google analytics can be used to facilitate this process. In the absence of these more user-friendly access methods, database analysis can also be used to extract the relevant data points (see Section 2A).
- **System Logs**: System logs are files that record key events and transactions that occur within a system. They are typically automated files that record the trail of events and transactions that occur in a system. This collection method is particularly useful for collecting indicators on System Performance (Section 3A). System log files are typically readily available, or can be turned on if not already available. Collecting data from a system log file requires a user to

extract the pertinent data points from a lengthy text file, and as such requires some technical knowledge of the system. System log files are generally the best source of truth for understanding system crashes or errors that may be occurring.

Qualitative methods to collect user feedback and programmatic data may include:

- **Focus Group Discussions**: As a qualitative data collection method, Focus Group Discussions (FGD) entail speaking with a representative sample of users as a group, in a setting where users are free to openly discuss opinions and attitudes regarding different aspects of a digital solution or set of tools. This type of collection method can be particularly useful for measuring indicators on User Engagement and Usability (Section 3D) and Sustainability (Section 3H). In a FGD, participants are encouraged to discuss amongst themselves and interact with their peers. Evaluators can use a set of questions to guide the discussion, and should record the main points being revealed by the group. When using this data collection method, users with similar job roles should be grouped together, and multiple FGDs should ideally be conducted to gain representation at multiple levels of the health system. Focus Group Discussions are best conducted in person, and are often scheduled alongside supervisory visits for logistical ease. Example questions for a focus group discussion can be found in Appendix B.
- **Key Informant Interviews**: A Key Informant Interview (KII) is designed to be a qualitative in-depth interview with representative end users, and unlike an FGD, is often designed to be one-on-one. This being said, they are another way to collect information on User Engagement and Usability (Section 3D). KII's are meant to be loosely structured and more open-ended, with evaluators using a high-level list of issues and questions to guide the conversation. KII's should be conducted with end users across different levels of the health system and across different geographic regions in order to get a representative sample. A Key Informant Interview can be conducted in person, often during a supervisory visit, or over the phone if an in-person visit is not feasible. Evaluators should take notes during each interview and attempt to find common themes and underlying issues amongst the different interviewees. Example questions for a KII can be found in Appendix C.
- **Questionnaires**: Questionnaires generally take the form of a set of questions on paper or a digital medium to collect user feedback. They can be a useful qualititative tool to collect data on a number of indicators, including Hardware (Section 3B), User Profile and Usage (Section 3C), User Engagement and Usability (Section 3D), Sustainability (Section 3H). Evaluators should put together a key set of questions that are designed to gain the necessary insight desired from users. Questions can ask users to place their response on a scale in order to gain a numeric value, ask for simple Yes/No responses, or be more open-ended and allow for paragraphs. Evaluators should have an idea of how they intend to analyze the responses while designing the questionnaire. When site visits are being conducted, the in-person visit can be leveraged by distributing the paper questionnaires, having the users record their responses, and then collecting the questionnaires for analysis. In the absence of in-person visits, questionnaires can also be conducted via phone or through digital solutions such as Google forms, although response rates will generally be lower when collecting data remotely. As with other data collection methods directly involving end users, a representative sample across job roles, health

system levels, and geographies should be included. An example questionnaire for a common usability test "System Usability Scale" -can be found in Appendix D. Another example questionnaire used specifically for malaria tools can be found in Appendix E.

- **Programmatic and Financial Analysis**: Indicators requiring programmatic and financial analysis may require data that is cannot be found in the digital system. These indicators may include Surveillance Process Impact (Section 3F), Governance (Section 3G), Sustainability (Section 3H). Evaluators must review human resource availability, training capacity, programmatic scheduling and activities, and budget data to perform this analysis.

C. Level of analysis

For each indicator, this document recommends a specific level of analysis. The levels of analysis and rationale for why this level is appropriate are described below, though it is important to note that this is software and context-dependent.

- **Aggregated Centrally**: Indicators recommended to be analyzed at an aggregated central level are indicators that are either specific to a software system as a whole, or a program as a whole. On the technical side, these are software indicators that typically can only be measured at an overall system level, and are not subject to regional variation (e.g. system downtime). On the programmatic side, these are indicators that must be evaluated at the central level by the national malaria program. While they will take regional variation into account (e.g. variations in regional operational budgets), the indicators' ultimate utility for decision making will be at the central level.
- **Aggregated Regionally:** Indicators recommended to be analyzed at an aggregated regional/sub-national level are indicators that can reveal important information regarding regional variation. An example of this would be reporting lag (time between an initial event and its entry into the system). By analyzing this information at the sub-national level, programs can identify problem areas to investigate and target (e.g. through additional support and supervision, or through additional training).
- **Per User:** Indicators recommended to be analyzed at the user level are those that are typically most informative when averaged at the per user experience. An example of this is data usage (how much data is required per user), as this would allow a program to estimate overall data needs and what the cost implications would be to maintain operations.

4. Indicator Categories

A. Outcome

Overview and purpose: These indicators broadly measure the progress towards malaria elimination. The performance of the surveillance system is only one of many aspects that will affect these indicators. However, outcome indicators provide important insight into whether the surveillance system is functioning as expected to inform each of these indicators.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
A1. Case burden	Number of confirmed cases	High	Aggregated centrally	Database analysis	Medium
A2. Incidence (per 1,000)	Percentage of population at-risk that are confirmed cases.	High	Aggregated centrally		Medium
A.3 ABER (annual blood examination rate)	Proportion of the at-risk population receiving a test	High	Aggregated centrally		Medium
A4. Test positivity rate	Number of cases tested that are confirmed for malaria.	Medium	Aggregated centrally		Medium
A5. Malaria free areas	Number of operational units with zero local incidence (or zero incidence) for 1 year.	High	Aggregated centrally		Medium

Outcomes and outputs: The availability of outcome data reflects the strength of a surveillance system and it may be possible that increases in these indicators may reflect surveillance improvements. Lack of change in outcome data over time may indicate a need to adjust out he surveillance system is used in order to improve quality or targeting of malaria control methods.

Data collection methods: Data for these indicators should be routinely collected through the case-based and aggregate-based surveillance systems and analyzed through database analysis.

Frequency: It is recommended that analysis of case burden, incidence, and test positivity rate should be conducted at least quarterly, or follow the analysis protocols that may already exist in country.

B. Surveillance

Overview and Purpose: These indicators measure the impact of digital solutions on surveillance processes. The indicators listed below are specific to malaria surveillance activities, including elimination-specific activities, and focus on 1) the timeliness and completeness of the end-to-end case surveillance workflow, 2) intervention coverage, and 3) the actual utilization of the information by the program. Included are also indicators that measure how surveillance processes have changed in moving from a paper to a digital system. To effectively measure whether digital solutions have led to improvements on surveillance processes, a baseline assessment should be done through a desk review of data submitted by legacy paper systems.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
B1. Point-of-Care completeness	Proportion of points of care reporting cases out of all those expected to rport	High	Aggregated regionally	Database analysis	Medium
B2. Case notification rate	Percentage of total cases – as reported in the aggregate system – notified through case notification solution.	Medium			Medium
B3. Notification time lag	Time difference between date of case detection/diagnosis and case notification.	High			Medium
B4. Case notification timeliness	Percentage of cases notified within country guidelines timelines.	High			Medium
B5. Case investigation rate	Percentage of notified cases that are investigated.	High			Medium
B6. Case investigation time lag	Time between case notification and investigation.	High			Medium
B7. Case investigation timeliness rate	Percentage of cases investigated within country guidelines for timelines.	High			Medium
B8. Case classification rate	Percentage of notified cases that are classified.	High			Medium
B9. Foci investigation rate	Proportion of foci investigated out of all identified foci.	High			Medium
B10. Intervention coverage rate	Percentage of targeted household or individuals covered by the interventions.	High			Medium
B11. Supervision intervention rate	Percentage supervision visits out of those targeted based on surveillance information.	Medium			Medium

Outcomes and Outputs: Improved outcomes in core surveillance indicators are one of the ultimate goals of technology investments and enhancements to surveillance systems. While the software-oriented indicators discussed in other sections inform system use and usability, the above Surveillance Process Impact indicators reveal more program-specific end results regarding the overall impact of the digital solutions interventions. Programs should be looking for improvements in notification, investigation and classification rates for cases and foci, improved and finer resolution intervention

coverage data, and increased utilization of information using a digital system, as compared to a paper-based system. It is highly recommended that all of these indicators are assessed as part of this evaluation, and automated for routine evaluation where possible (i.e. case, foci and intervention indicators).

Data Collection Methods and Frequency: Surveillance and intervention-related indicators should already be readily available through the program's normal malaria reporting activities, and no additional data collection effort should be required as part of this evaluation. A major advantage that comes with the digitization of these reporting streams is that these indicators can be automated for routine monitoring with minimal effort. It is highly recommended that this automation is done for users in all levels of the health system given both the extreme importance of these indicators and ease of automation.

Note that baseline indicators are especially critical to evaluate impact of digital solutions on sureveillance processes. As a result, legacy data sets should be analyzed to calculate baseline indicators for comparison. This can be done through a desk review of data such as case notification rate, case investigation rate, and foci investigation rate under a paper-based system.

Frequency: The final two indicators above pertain to data usage (Visualization of Information and Supervision Intervention Rate) and should be collected from a review of program activities that were conducted. Indicators with collection method 'database analysis' should be analyzed routinely (once per month) when connected to automated analytics outputs. Indicators with data collection method 'programmatic analysis' should be analyzed once or twice per year.

C. Data Quality and Reporting

Overview and Purpose: These indicators measure data quality by assessing data validity, integrity, precision, reliability, and timeliness. This information can then be used to identify how effective a system is in its utility in collecting data. All of these data and reporting measures are critical in verifying the data is being reporting when it needs to be, and can be used to the full extent for decision-making with high confidence. As such, these indicators should be measured for each software system being used for data collection by the country.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
C1. Reporting lag	Time between date of the initial event and date of entry/syncing with the system database.	High	Aggregated regionally	Database Analysis	Low
C2. Data sync frequency	Time between data form completion and data form sync.	High			Low

C3. Data completeness	Proportion of expected cases that are reported in case-based reporting.	High			Medium/High
C4. Concordance rate	Proportion of malaria cases confirmed in aggregate reporting that are reported in case-based reporting systems.	High			Low
C5. Correct entry	Accuracy of data entry against a source of truth (in the absence of this, can assess accuracy of format, i.e. – number of digits in a phone number).	Medium			Medium
C6. Case geolocation rate	Proportion of cases in case-based system that are linked to a geolocated health facility, locality or household.	Low			Low
C7. GPS organizational unit accuracy	Percent of GPS points that fall within their expected administrative unit.	High			High
C8. Usable data	Percentage of records removed from analysis due to illogical and unusable data.	Medium			Medium

Outcomes and Outputs: High data quality helps to ensure that information users of the surveillance system remain confident in their inputs, which in turn helps improve their confidence in using data for decision-making. Furthermore, because all malaria surveillance and intervention activities are dependent on data at a fine resolution, high quality data is necessary in order to make progress towards elimination. The adoption of new digital solutions for data collection, particularly when transitioning from a paper-based system, is often marked by major improvements to data quality due to the ability to incorporate input constraints and data validation rules. These indicators serve as one of the strongest argument for the adoption of digital solutions for surveillance activities, and as a result it is highly recommended that baseline measurements be collected from the legacy systems for reference. This will provide major insight into the true impact of the introduced digital solutions.

Data Collection Methods: Data collection for the above Data Quality and Reporting indicators is best conducted though a desk review and analysis of the data. As mentioned in Section 2 (Data Collection Methods), a formal Data Quality Assessment* using information from the database can be conducted in a structured manner using available published guides. This involves statistically evaluating collected form data in order to determine whether it meets the quality necessary to inform processes and decision-making. DQA's focus on 5 key data quality standards: validity, reliability, precision, integrity and timeliness. A variety of approaches can be used to assess these data quality standards, but a desk review and database analysis effort is always required. Evaluators can leverage existing DQA guides (such as the USAID guide in the Resources section) to assist in the planning and execution of the assessment.

Data sync frequency can be calculated by analyzing timestamps recorded in the database, provided timestamps for both data entry and data submission are present. Adding timestamps to records requires minimal effort from developers; as a result, if these timestamps are not already being recorded by the system, it is recommended that steps are taken to include them. Data accuracy can be measured against a source of truth such as a paper form; in the absence of this it can be assessed based on accuracy of format (i.e. – number of digits in a phone number).

Frequency: Because of the central role in data quality and their ability to inform supervisory visits, it isF recommended that these are obtained on a quarterly basis.

D. User Profile and Usage

Overview and Purpose: These indicators provide a profile of users and how they use the system. By revealing who is using the system, how often and from what device (e.g. desktop vs. mobile) a program can identify poorly engaged users and administrative units, and use this insight to inform follow-up plans to address reporting and usage gaps. These indicators are meant to guide supervision visits by region as well as possible usability improvements to the system. They should be measured for each software system in the country.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
D1. Monthly or daily active user	Number of users on a given day or month compared to expected number of users (based on case load).	Medium	Aggregated regionally	System Analytics, Question-naires	Medium
D2. Device metrics	Devices used to access the system.	Low			Medium
D3. Geographic distribution	Geography of where the system is used.	Low			Medium

Outcomes and Outputs: The end result of these indicators is to verify how many of the expected users are accessing the system, and make decisions accordingly. By comparing expected vs. actual results, programs can identify gaps and plan supervision visits, follow-up calls and refresher trainings to improve usage. Furthermore, if transitioning from an older software system or version, having a baseline for these indicators would help a program understand whether engagement has been impacted positively or negatively by newer versions.

Data Collection Methods: Some products, like DHIS2, have built-in user analytics dashboards but can also be connected with an external analytics product. Although some initial setup may be required (depending on the software), these indicators are easily automatable once set up and can thus be integrated into routine analysis. Other data collection methods like questionnaires can also be used to gather this information in a self-reported fashion, however if possible the objective and quantitative data made possible by automated user analytics is highly preferred. Programs should use these

indicators to compare expected system usage vs. actual system usage, and take steps to understand and address gaps.

Frequency: These indicators should ideally be automatically calculated from system analytics and analyzed routinely (on a monthly basis). If resorting to questionnaires, these indicators should be obtained at least twice a year.

E. User Engagement and Usability

Overview and Purpose: These indicators measure how users are engaging with the system and how easy they find the system to use. These indicators comprise a mix of user-reported data, data from supervisory visits, and user analytics data in order to provide a comprehensive picture of user engagement and usability. These indicators are intended to guide supervision visits and possible usability improvements to the system, and should be measured for each software system in the country. This category of indicators allows for the most direct user feedback, and as such should not be neglected in any evaluation of a surveillance system. In general, many valuable design and configuration improvements can come from this user feedback, while the user analytics indicators will provide a complementary and objective component to the engagement described by users.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
E1. System Usability Scale rating	SUS score using on standard questionnaire.	High	Aggregated regionally	Question-naire	High
E2. User competence	How well the user is able to navigate and complete tasks.	High		Sample Test	High
E3. User perception of performance	Feedback from the user on how well they perceive the system to be performing (e.g. speed, crash rates, errors, etc.) – which may affect user adoption.	Medium		Question-naires, KII, FGD	High
E4 User pain points	Feedback from the user on frustrations, pain points, and commonly encountered issues.	High			High
E5. User identified deficiencies	Feedback from the user on items/information they need that is absent from the system.	High			High
E6. User satisfaction	Feedback from the user on how satisfied they are with using the system for their work.	High			High
E7. Data entry time	Average length of time to complete data entry from start to completion of form.	High		Sample Test, Error Logs,	High

E8. Object hits	Proportion of users on a given page that interact with specific objects of interest. Programs should break down this indicator for each individual object they want to measure.	Medium	Aggregated regionally	Database Analysis	Medium
E9. Error rates	Frequency of users encountering system errors (can be indicative of usability or training).	Medium		System Analytics	Medium
E10. Session length	Time between opening and closing the system.	Low			Medium
E11. Session interval	Length of time between users' first session and their next one.	Low			Medium
E12. Retention rate	Percentage of users who return to the system based on the date of their first visit.	Low			Medium
E13. Page hits	Proportion of users who have accessed a page, for each page in the system. Programs should break down this indicator for each individual page they want to measure.	Low			Medium
E14. Time spent per page	Amount of time spent by users on each page of the system (generally in minutes)	Low			Medium

Outcomes and Outputs: The combination of positive user feedback, high user competence, and measurable active user engagement are crucial in ensuring digital solutions used in the surveillance system are positively received. This positive reception helps to ensure the tools are institutionalized and a culture around their active and routine use is well established, fostering high reporting rates, quality data, and routine information use. When rolling out new tools (or new versions of tools) it is extremely valuable to compare these indicators with baselines from the legacy systems (or older versions) to gain insight into how users view the change. If no baseline values exist, at a minimum users can be asked the same set of questions for both the legacy and new systems while evaluating the new system and collecting data through questionnaires, KII's and FGD's. Alternatively, they can be asked to rate how the new system compares to the old one in a series of questions.

Data Collection Methods: Indicators calculated from user-reported data within this category is often best gathered during supervisory visits, in the form of questionnaires, Key Information Interviews (KII) or Focus Group Discussions (FGD). Data for this category should be collected for users at all levels of the health system hierarchy and across diverse geographic regions in order to get a representative sample

of all users. Data is ideally collected in person, particularly for KII's and FGD's. Questionnaires can easily be administered remotely as well if resources are limited, and a variety of free form products (e.g. Google Forms) can be used for this purpose. Similarly, interviews can be conducted via telephone if an in-person visit is not feasible. For some indicators such as user competence, data collection via observation is extremely valuable as it allows an evaluator to directly witness areas of weakness or gaps in knowledge without user bias. This can be done during supervisory visits, observing a user's ability to complete a predefined list of tasks in the system.

Indicators collected by system analytics are complementary and provide a window into how users truly engage with the system, and are typically collected by setting up a user analytics page or dashboard. Once set up however, the indicators are routinely available and no further effort is needed. For evaluation purposes, appropriate benchmarks should be established so that programs can clearly define what they determine to be high-performing, medium-performing and low-performing areas and user groups. Furthermore, programs should be breaking down higher-level indicators into metrics specific to their system for measurement (e.g. for object hits, programs should define what the objects of interest are and track each one separately.)

Frequency: Technical and quantitative indicators that can be routinely collected within the system should be analyzed on a regular (monthly) basis. Indicators requiring user interaction are ideally calculated every 6 months.

F. System Performance

Overview and Purpose: These indicators measure how the core software is performing, and inform performance updates to the software. They provide quantitative insight into the system stability and reliability that would be experienced by a user. As a result, these indicators can provide an important objective measurement for user experience quality, and should be measured for each software system being used in the country.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
F1. System Crash Rate	How often the system experiences an unexpected exit.	High	Aggregated centrally	System logs	Medium
F2. System uptime/downtime	Percentage of time in a given period for which the system is operational vs. non-operational.	High	Aggregated centrally		Medium
F3. System latency	Round trip time from request to a response.	High	Aggregated centrally		Medium
F4. System load per period	Number of transactions over a certain period.	High	Aggregated centrally		Medium

Outcomes and Outputs: A positive user experience is critical for the adoption and uptake of any software. When poor system performance leads to frequent crashes and lags, user engagement is

diminished and reporting and data quality adversely impacted. Consistent poor performance of a system may indicate a need to improve configuration to optimize the system or, at worst case re-evaluate the appropriateness of the system in use. Accordingly, system performance requires attention when evaluating any system. If adopting new software, it is recommended that programs analyze these performance indicators for their legacy system, if applicable, and use them as a baseline for comparison.

Data Collection Methods: An objective way to measure performance indicators is to analyze system logs, which are typically automated files that record the trail of events and transactions that occur in a system. System logs are very rarely analyzed during routine M&E, as their utilization typically requires some assistance from IT staff that are familiar with how log files are structured and can be analyzed. System logs are usually readily available, however, and a representative sample of logs can used in order to calculate System Crash Rate, Latency and Load per Period. Alternatively, crash rate and latency can be collected from user survey methods such as interviews or questionnaires, albeit more subjectively. If utilizing these methods, multiple users in diverse geographies and connection settings should be asked to identify these measures from their experiences.

Frequency: It is recommended that these indicators are routinely collected via system logs to assess system performance and analyzed at minimum twice a year. If using questionnaires, it is also recommended that these data collection/analysis activities are conducted twice a year.

G. Hardware Performance

Overview and Purpose: These indicators measure hardware reliability and inform upgrades to administered hardware. The hardware indicators listed provide insight into the quality of the devices used to access software. Poor hardware quality can increase barriers to reporting and monitoring by making the device inconvenient to use and more frequently unusable. Furthermore, poor devices can impact data quality due to inaccurate GPS coordinates, which are critical data elements in the type of fine-resolution case-based surveillance necessary for malaria elimination. As such, these indicators should be measured for each type of hardware used in the country and used in the decision-making progress for hardware maintenance, procurement and upgrade plans.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
G1. Time to charge	How long it takes the device to fully charge.	High	Aggregated regionally	Question-naires, Sample Test	Medium
G2. Battery life	How long the device's power lasts when fully charged. This can also be measured when specific apps of interest are in use.	High	Aggregated regionally		Medium
G3. Ability to connect	How well the device is able to connect to data or wifi.	Medium	Aggregated regionally		Medium
G4. GPS accuracy	Accuracy radius of collected GPS Points for specific hardware.	Medium	Aggregated regionally	Database Analysis, Sample	Medium

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
				Test	

Outcomes and Outputs: Like system performance, hardware quality directly impacts an end user's ability to engage with a system. Poor hardware quality can make devices inaccessible or unusable for large periods of time, completely blocking system access and adversely impacting the ability to collect or use data. Furthermore, this can diminish a user's motivation to engage with the software and inhibit institutionalization of digital solutions. As a result, the role of hardware should not be underappreciated when evaluating how a surveillance system is functioning. When considering upgrading or adopting new devices, it is recommended to measure these indicators for the existing hardware to serve as a baseline for comparison. Furthermore, consumer reports can be found for different hardware models to aid in research prior to making purchases.

Data Collection Methods: Data for hardware indicators can easily be collected by running a series of tests on a representative sample of devices being used, which will provide an objective measurement of how the hardware behaves. Alternatively, this data can be gathered more subjectively by administering questionnaires to end users. Users should be asked to quantify the typical Time to Charge and Battery Life in hours, and the GPS Accuracy in meters. Ability to connect may be a more subjective and open-ended response, and users' geographical differences in general connectivity should be taken into consideration if using a user-reported method.

Frequency: These indicators should be calculated at any point when new hardware options can be considered, such as when rolling out a system to more users or when hardware contracts are set to expire. They should also be measured once a year on a routine basis to understand drivers of system usage and user experience.

H. Governance

Overview and Purpose: These indicators measure whether the digital solutions are aligned with country guidelines and/or influence changes in guidelines. The long term adoption and seamless incorporation of new surveillance system processes and tools demand that they ultimately be aligned with the country's larger guidelines. These indicators should be measured for each software system in the country.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
H1. Digital solutions aligned with country guidelines	Digital solutions configured and implemented in alignment with country surveillance guidelines.	High	Aggregated Centrally	Programmatic analysis	Low
H2. Changes in SOP	SOPs for surveillance have been updated to include digital solutions, have been	High			

	field tested, and are in use.		Aggregated Centrally	Programmatic analysis	Low
H3. Surveillance system assessments	Frequency of formal surveillance systems assessments.	Medium			
H4. Data review	Frequency of data and dashboard review meeting (number per year) at all levels.	High			
H5. Net promoter score with data and dashboards from digital system	User satisfaction with data and dashboards from digital system, as measured by willingness to recommend to others.	Medium			

Outcomes and Outputs: The long term success of a robust surveillance system with appropriate technology requires alignment with country guidelines and SOPs. Accordingly, it is expected that each piece of software that has been incorporated into the surveillance system is also properly and seamlessly incorporated into the workflows and processes established in the guidelines. Baseline indicators are less important for this category of indicators, however it is worth understanding how well a given country has historically kept the guidelines and SOPs current.

Data Collection Methods: These indicators must be calculated by engaging in desk review and programmatic analysis. Evaluators should review both older and newer versions of country guidelines and SOPs and verify if they are aligned to the updated surveillance system processes with their respective digital solutions. Furthermore, evaluators should review whether the updated guidelines and SOPs are properly institutionalized and in use at all levels of the health system hierarchy and in all geographic regions. Evaluators can also assess frequency of data review at a national or district level, as well as general satisifaction score on a scale of 1-10 by high level users.

Frequency: These indicators should be analyzed once per year.

I. Sustainability and Scalability

Overview and Purpose: These metrics gauge the relative ease or difficulty of scaling and maintaining a system over time. They will guide programmatic planning regarding ongoing investments needed to keep the system fully operational. These indicators should be measured for each software system in the country.

Indicator	Description of Indicator	Priority	Level of Analysis	Collection Method	Collection Effort
I1. Feedback processes	Process and timelines for recording and addressing feedback on the surveillance system.	Medium	Aggregated centrally	Question-naires, KII's, Focus Groups	Medium

I2. Program technical capacity	Program's ability to support users and maintain the system without external support.	High			High
I3. Program training capacity	Program's ability to manage and conduct regular trainings (through quality trainers).	Medium			High
I4. Support of ongoing technical costs	Program's ability to support ongoing technical costs, including for: Server maintenance, System updates, Data. Pre and post scale up should be considered, if applicable.	High		Programmatic and Financial Analysis	Medium
I5. Support of ongoing programmatic costs	Program's ability to support ongoing programmatic costs, including for: Initial trainings, Refresher trainings, Supervision visits, Data quality assessments. Pre and post scale up should be considered, if applicable.	High			Medium
I6. Support of ongoing hardware costs	Program's ability to support ongoing hardware costs including for: Tablet/smartphone, Wifi. Pre and post scale up should be considered, if applicable.	High			Medium
I7. Scale up scope	A comparison of the current number of staff/admin units for whom the system has been rolled out to vs. total targeted staff.	High	Aggregated regionally		Medium
I8. One time scale up costs	All costs (technical, programmatic and hardware-related) required for the program's targeted scale up.	High	Aggregated centrally		Medium
I9. Training	Length and frequency of training required to effectively use the system.	High	Per user		Medium
I10. Data integration	Malaria-related data integrated with health information system.	Medium	Aggregated centrally	Programmatic analysis	Low

Outcomes and Outputs: In order to ensure sustainability, programs must ensure that they have the capacity to maintain the system from a technical perspective, support ongoing trainings, ensure a continuous feedback cycle, and fund and manage improvements to the system. Unlike the other indicator categories that are designed to measure system usage, engagement and impact, the above

Sustainability indicators are more designed to support planning processes on addressing gaps and ensuring long-term impact of the digital solutions.

Data Collection Methods: Indicators around sustainability require a variety of data collection methods, including programmatic review, financial review, user and stakeholder feedback, and to a less extent, system logs. Programmatic and financial analysis will require extensive desk review of country processes, budgets and system costs. Data requiring user and stakeholder engagement can be collected through questionnaires, interviews, KIIs, or FGDs. Programs should be looking to determine whether there are any gaps in capacity or funds that could compromise the long term sustainability of the surveillance system tools.

Frequency: These indicators should be analyzed once per year.

5. Country Adaptation Approach

To adapt this M&E framework to specific country settings, government malaria control teams and M&E specialists should identify priority indicators to be assessed, log existing baseline data for those priority indicators especially those in the *Outcomes* and *Surveillance* categories, set targets for the indicators where possible, and come to a consensus on a method of verification. Table 2 identifies the minimum information to collect when developing an M&E plan.

Table 2. M&E planning

Indicator	Definition	Unit of Measure	Baseline*	Year of Baseline	Target*	Data source
E.g. Case investigation rate	*Percentage of notified cases that are investigated*	*%*	*28%*	*2018*	*>80%*	*Database analysis on case-based surveillance system, using date-stamped paper forms in facilities as back-up as needed.*

*Applicable to only a subset of indicators and may be estimated where relevant

A. Indicator identification

Indicators can be selected from this M&E framework by countries based on relevance for their specific context. Selection should involve discussions within malaria programs, supported by technical inputs from CHAI and in country partners, and informed by the WHO in order to align with international metrics for certification of malaria elimination. This framework recommends selecting high priority indicators from each category in order to build the foundation of an M&E protocol that can assess not only the outcome and impact of a digital system for surveillance, but also the inputs, process, and outputs required for the digital system to perform as expected.

B. Baseline

Once indicators are identified, a baseline should be assessed based on existing tools for surveillance – whether digital or paper. Whether current tools are paper or digital, a baseline on *Outcomes* and *Surveillance* data should be noted based on current surveillance system metrics. A qualitative

assessment can additionally be conducted on *Sustainability* and *Governance* indicators. Health facility surveys and review of data collected through paper-based forms can be conducted to also identify the baseline measures for metrics on *Data Quality and Reporting*.

If current tools are digital, an additional database assessment can be used to identify baseline metrics for technical indicators in the categories of *Data Quality and Reporting*, *User Profile and Usage, User Engagement and Usability, System Performance,* and *Technical Performance*.

These baseline metrics will be key for measuring whether the digital solutions have incremental positive improvements on surveillance systems over current systems. Additionally, baseline metrics will be vital to inform realistic target setting.

C. Target setting

Targets will need to be set based on country-specific goals, and international standard when available. It will be important to ensure that targets are realistic in order to sustain momentum and motivation to achieve them, while also being aligned with broader malaria program targets around elimination. Currently there are no detailed guidelines to conduct target setting processes, especially within technology-oriented indicators (e.g. *System Performance* and *Hardware Performance*). However, different approaches such as benchmarking based on performance from similar reference geographic areas, regression analysis and forecasting based on historical patterns, or mixed methods approaches that also considers analyses of international data and the advice of experts can be used to recommend initial targets to be then adjusted based on in country discussions[1] .

Targets for indicators in the *Outcome*s and *Surveillance* categories should align with targets set for existing surveillance processes in country, or be established based on guidelines set by global standards. For technical indicators around *System Performance* and *Hardware Performance*, benchmarks can be initially set based on industry targets and impact on users – for example, proposing a target of less than 0.25% crashes for critical flows (e.g. “Login”) based on aggregated data on crash rate for top rated iOS and Android apps[2]. Other targets for indicators on *User Profile and Usage* or *User Engagement and Usability*, and *Data Quality and Reporting* may be set incrementally each quarter based on previous performance in order to assess impact of software iterations. Results of piloting can be used to set baselines for usability with the goal of improving usage, usability, and data quality as digital solutions improve with each release.

D. Data Sources

Most sources of data should come directly from the analytics built into the digital solutions – especially for *System Performance*, *User Profile and Usage* or *User Engagement and Usability* indicators. *Data Quality and Reporting* information should be extracted from multiple quantitative and qualitative sources – including directly from the digital tools and where relevant, from legacy surveillance system databases.

---

[1] Arur A, Mohammed-Roberts R, Bos E. Setting Targets in Health, Nutrition, and Population Projects. Washington, DC: World Bank. http://hdl.handle.net/10986/13587 (2011).
[2]Kwok, R. 7 Practices for Optimizing Mobile App Experience. San Francisco: Apteligent. https://www.apteligent.com/wp-content/uploads/2016/06/00-Whitepaper-7-Best-Practices-V1-1-1.pdf (2016).

Measurement of program data can be conducted through varied sources. For example, *Outcomes* and *Surveillance* data will be available both in raw format from national surveillance systems or automatically analyzed through the digital tool databases and dashboards. They can also be cross-validated through date-stamped paper registers in health facilities where relevant. Data on *Governance* and *Sustainability and Scalability* will need to be collected through analysis of programmatic documents or key informant interviews with government officials. Several sources of data may be available for program data, and when relevant cross-validation should be conducted to validate accuracy of data coming through digital surveillance systems.

With the availability of data through an M&E system, a process evaluation can be conducted to identify how indicators are used – particularly how surveillance data is used to improve outcomes and impact, or how technical indicators around usage and system functionality are used to improve tools and performance. Key informant interviews, observations, questionnaires, and other qualitative methods to track processes around the surveillane system would provide insight into the value and impact of an enhanced system. In particular, these qualitative data would allow for an in depth analysis on how data from surveillance systems are used to adjust interventions and control strategies in order to accelerate progress towards malaria elimination.

6. Next Steps

An M&E plan will be developed for each pilot country in the DSME grant in partnership with the WHO, which will also conduct a formal evaluation during the course of the grant. These pilot countries will serve as deep dive M&E countries, with more extensive planned M&E indicators, targets, and reporting. Adaptation of this M&E framework to realistic targets and M&E protocols for each pilot country will be conducted prior to the roll out of the digital solution enhancements in Q4 2018. M&E plans will also be developed for the non-pilot countries in the DSME grant, though these are expected to be more limited in indicators and reporting format due to operational and financial limitations. Funding for the finalized M&E protocol will take into consideration existing funding resources in country for surveillance systems.

Appendix A: Summary of Indicators

	Indicator	Indicator Description	
	Outcome		
A1	Case burden	Number of confirmed cases	High
A2	Incidence (per 1,000)	Percentage of population at-risk that are confirmed cases.	High
A3	Annual blood examination rate (ABER)	Proportion of the at-risk population receiving a test	High
A3	Test positivity rate	Number of cases tested that are confirmed for malaria.	Medium
A4	Malaria free areas	Number of operational units with zero local incidence (or zero incidence) for 1 year.	High
	Surveillance		
B1	Point of care completeness	Proportion of points of care reporting cases out of those expected to report	High
B2	Case notification rate	Percentage of total cases – as reported in the aggregate system – notified through case notification solution.	Medium
B3	Notification time lag	Time difference between date of case detection/diagnosis and case notification.	High
B4	Case notification timeliness	Percentage of cases notified within country guidelines timelines.	High
B5	Case investigation rate	Percentage of notified cases that are investigated.	High
B6	Case investigation time lag	Time between case notification and investigation.	High
B7	Case investigation timeliness rate	Percentage of cases investigated within country guidelines for timelines.	High
B8	Case classification rate	Percentage of notified cases that are classified.	High
B9	Foci investigation rate	Proportion of foci investigated out of all identified foci.	High
B10	Intervention coverage rate	Percentage of targeted household or individuals covered by the interventions.	High
B11	Supervision intervention rate	Percentage supervision visits out of those targeted based on surveillance information.	Medium
	Data Quality and Reporting		
C1	Reporting lag	Time between date of the initial event and date of entry/syncing with the system database.	High
C2	Data sync frequency	Proportion of expected cases that are reported in case-based reporting.	High
C3	Data completeness	Proportion of required fields filled out (including comparison between legacy data or paper records and digitally entered data).	High
C4	Concordance rate	Proportion of malaria cases confirmed in aggregate reporting that are reported in case-based reporting systems.	High
C5	Correct entry	Accuracy of data entry against a source of truth (in the absence of this, can assess accuracy of format, i.e. – number of digits in a phone number).	Medium
C6	Case geolocation rate	Proportion of cases in case-based system that are linked to a geolocated health facility, locality or household.	Low
C7	GPS organizational unit accuracy	Percent of GPS points that fall within their expected administrative unit.	High
C8	Usable data	Percentage of records removed from analysis due to illogical and unusable data.	Medium
	User Profile and Usage		
D1	Monthly or daily active user	Number of users on a given day or month compared to expected number of users (based on case load).	Medium
D2	Device metrics	Devices used to access the system.	Low
D3	Geographic distribution	Geography of where the system is used.	Low
	User Engagement and Usability		
E1	System Usability Scale rating	SUS score using on standard questionnaire.	High
E2	User competence	How well the user is able to navigate and complete tasks.	High

E3	User perception of performance	Feedback from the user on how well they perceive the system to be performing (e.g. speed, crash rates, errors, etc.) – which may affect user adoption.	Medium
E4	User pain points	Feedback from the user on frustrations, pain points, and commonly encountered issues.	High
E5	User identified deficiencies	Feedback from the user on items/information they need that is absent from the system.	High
E6	User satisfaction	Feedback from the user on how satisfied they are with using the system for their work.	High
E7	Data entry time	Average length of time to complete data entry from start to completion of form.	High
E8	Object hits	Proportion of users on a given page that interact with specific objects of interest. Programs should break down this indicator for each individual object they want to measure.	Medium
E9	Error rates	Frequency of users encountering system errors (can be indicative of usability or training).	Medium
E10	Session length	Time between opening and closing the system.	Low
E11	Session interval	Length of time between users' first session and their next one.	Low
E12	Retention rate	Percentage of users who return to the system based on the date of their first visit.	Low
E13	Page hits	Proportion of users who have accessed a page, for each page in the system. Programs should break down this indicator for each individual page they want to measure.	Low
E14	Time spent per page	Amount of time spent by users on each page of the system (generally in minutes).	Low
	System Performance		
F1	System Crash Rate	How often the system experiences an unexpected exit.	High
F2	System uptime/downtime	Percentage of time in a given period for which the system is operational vs. non-operational.	High
F3	System latency	Round trip time from request to a response.	High
F4	System load per period	Number of transactions over a certain period.	High
	Hardware Performance		
G1	Time to charge	How long it takes the device to fully charge.	High
G2	Battery life	How long the device's power lasts when fully charged. This can also be measured when specific apps of interest are in use.	High
G3	Ability to connect	How well the device is able to connect to data or wifi.	Medium
G4	GPS accuracy	Accuracy radius of collected GPS Points for specific hardware.	Medium
	Governance		
H1	Digital solutions aligned with country guidelines	Digital solutions configured and implemented in alignment with country surveillance guidelines.	High
H2	Changes in SOP	SOPs for surveillance have been updated to include digital solutions, have been field tested, and are in use.	High
H3	Surveillance system assessments	Frequency of formal surveillance systems assessments.	Medium
H4	Data review	Frequency of data and dashboard review meeting (number per year) at all levels.	High
H5	Net promoter score for digital system	User satisfaction with data and dashboards from digital system, as measured by willingness to recommend to others.	Medium
	Sustainability and Scalability		
I1	Feedback processes	Process and timelines for recording and addressing feedback on the surveillance system.	Medium
I2	Program technical capacity	Program's ability to support users and maintain the system without external support.	High
I3	Program training capacity	Program's ability to manage and conduct regular trainings (through quality trainers).	Medium
I4	Support of ongoing technical costs	Program's ability to support ongoing technical costs, including for: Server maintenance, System updates, Data. Pre and post scale up should be considered, if applicable.	High
I5	Support of ongoing programmatic costs	Program's ability to support ongoing programmatic costs, including for: Initial trainings, Refresher trainings, Supervision visits, Data quality assessments. Pre and post scale up should be considered, if applicable.	High
I6	Support of ongoing hardware costs	Program's ability to support ongoing hardware costs including for: Tablet/smartphone, Wifi. Pre and post scale up should be considered, if applicable.	High

I7	Scale up scope	A comparison of the current number of staff/admin units for whom the system has been rolled out to vs. total targeted staff.	High
I8	One time scale up costs	All costs (technical, programmatic and hardware-related) required for the program's targeted scale up.	High
I9	Training	Length and frequency of training required to effectively use the system.	High
I10	Data integration	Malaria-related data integrated with health information system.	Medium

Appendix B: Focus Group Discussion (Sample)

Adapted from WHO M&E Toolkit for Digital Health Interventions:
http://apps.who.int/iris/bitstream/10665/252183/1/9789241511766-eng.pdf

Overview: Focus group discussions are used to solicit the opinions or perspectives of a group of individuals. The purpose of a focus group discussion is not to achieve consensus but to gather a wide range of opinions and views

Considerations:.The questions below should be used as a guide rather than a set list. Encourage participants to expand on responses, and observe participants' non-verbal cues (body language, emotion during response) during the discussion.

Focus Group Members: Deployed workers who use digital solutions for surveillance

---

1. What are common challenges that you face in conducting malaria surveillance activities in the community?
2. What features of the digital solutions do you like in the mobile interface? In the way data is entered? In the way data appears on the tool? In the way data is presented in dashboards? And any other other component of the tools?
3. Which features do you dislike? In the way data is entered? In the way data appears on the tool? In the way data is presented in dashboards? And any other other component of the tools?
4. What is easy to use with the digital solutions?
5. What are difficulties you experience with the digital solutions?
6. What would you change about the digital solutions?
7. What additional features would you like to see in the digital solutions?
8. How would your ability to conduct surveillance activities change without digital solutions?
9. How do you think this tool compares to the previous system you used (paper or otherwise)?
10. How do you think data availability compares to the previous system you used (paper or otherwise)?
11. How do you think data usability for programmatic action compares to the previous system you used (paper or otherwise)?
12. What else would you like to tell me about your experience with the digital solutions?
13. Is there anything else I might have missed?Is there anything else I might have missed?

Appendix C: Key Informant Interview (Sample)

Key Informant: Program Managers
Indicator Category: Sustainability

Deployment and Training

A.1. How are systems deployed to users at provincial, district, facility, and community levels?

A.2. What are the biggest deployment challenges?

A.3. How could these deployment challenges be addressed?

A.4. Can you describe the methods used to identify whether users are having performance challenges with the digital solutions?

A.5. Can you describe the methods used to ensure users are skilled in using the digital solutions (trainings, supervisions, job aids, etc)?

A.6. Can you describe the routine trainings provided to new users of the digital solutions (mobile tools, DHIS2, dashboards, etc)? How many routine trainings are provided, who attends, who facilitates, and what is taught during these trainings?

A.7. Can you describe the routine trainings provided to existing users of the digital solutions (mobile tools, DHIS2, dashboards, etc)? How many routine trainings are provided, who attends, who facilitates, and what is taught during these trainings?

A.8. Can you describe ad hoc trainings provided to users of the digital solution system (mobile tools, DHIS2, dashboards, etc)? Who attends, who facilitates, and what is taught during these trainings? How often are these provided? How do you decide an ad hoc training is needed?

Maintenance and Support

A.9. How many people are involved in system maintenance, data quality assurance, and reporting? Who are the key players in ensuring the digital solutions are functional?

A.10. Are there other technical partners supporting one or some of the systems? Please describe.

A.11. If a system breaks or needs an enhancement, do you do fix it or get a technical partner to fix it?

A.12. What are the biggest challenges in system maintenance? How could these maintenance challenges be addressed by the way the digital systems are designed?

A.13. Can you describe key gaps in supporting the digital solutions used? What capacity improvements (skillsets, new hires, etc) would be needed to address these challenges?

Feedback Processes

A.14. How do you determine whether the digital solutions are being used successfully for surveillance?

A.15. How do you collect feedback on what is going well or not well with the use of the digital solutions (for community workers, program managers, IT staff, supervisors, etc)?

A.16. What is the process for addressing feedback on the use of digital solutions?

A.17. Can you describe key gaps in functionality of digital solutions currently used? What improvements have been made to address these gaps? What improvements have not been made, and what are reasons preventing them from being made?

Appendix D: System Usability Scale Questionnaire (Sample)

Source: SUS – A Quick and Dirty Usability Scale (Brooke 1986)

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. I think that I would like to use this system frequently.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
2. I found the system unnecessarily complex.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
3. I thought the system was easy to use.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
4. I think that I would need the support of a technical person to be able to use this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
5. I found the various functions in this system were well integrated.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
6. I thought there was too much inconsistency in this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
7. I would imagine that most people would learn to use this system very quickly.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
8. I found the system very cumbersome to use.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
9. I felt very confident using the system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
10. I needed to learn a lot of things before I could get going with this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Calculating Usability Score:

- For each of the odd number questions, subtract 1 from the score
- For each of the even numbered questions, subtract their value from 6
- Add these new values and multiply the sum by 2.5
- This new score is now out of 100. This is not a percentage but a clear way of seeing the score.

Score Overview:

- 80.3 or higher is an A. People love your tool.
- 68 or thereabouts gets you a C. You're doing OK but could improve
- 51 or under gets you an F. Make usability your priority now and fix this fast.

Appendix E: Mobile Tool for Malaria Questionnaire (Sample)

1. How easy is it for you to log into the application on your mobile phone?

☐	☐	☐	☐	☐
1	2	3	4	5
Very Easy				Very Difficult

2. Do you think the application is easy to use?

☐	☐	☐	☐	☐
1	2	3	4	5
Very Easy				Very Difficult

3. How do you think you did in completing the forms in the application?

☐	☐	☐	☐	☐
1	2	3	4	5
Very successful				Not at all successful

4. How useful is the tool for your work?

☐	☐	☐	☐	☐
1	2	3	4	5
Very Useful				Not useful at all

5. Which process do you find challenging in using the application? (circle all that apply)

A. Logging in
B. Case registration
C. Completing the case investigation form
D. Syncing data
E. Updating the application
F. Updating a record
G. Navigating from one form to another
H. Other (please list)

6. Are there forms in the application that were difficult to complete? Please list.

7. Are the questions on the application clear, concise, and easy to understand?
Yes No

8. Please list any suggestions for how to rephrase the questions so they are easier to understand.