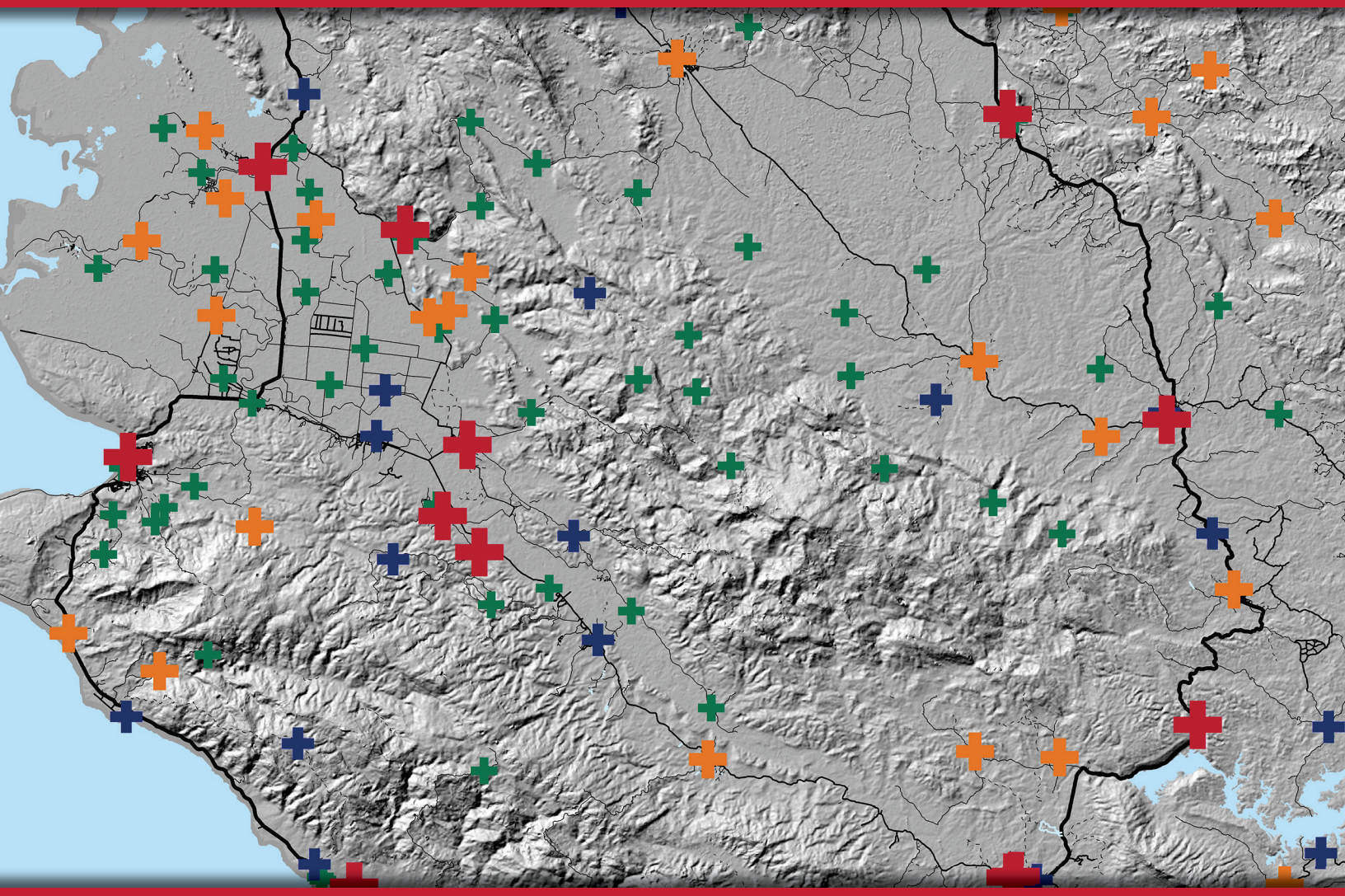
USAID
FROM THE AMERICAN PEOPLE

MASTER FACILITY LIST RESOURCE PACKAGE:

Guidance for countries wanting to strengthen their MFL

Module 7: Establishing an MFL Dataset



May 2017

DRAFT

ESTABLISHING AN MFL DATASET

This module describes the process of establishing an MFL dataset. Some sections of the module apply to countries where no MFL exists, others are applicable to countries that already have an MFL but need to fill gaps in the data or to validate the data content. Using the flow chart in Figure 1, you can quickly determine which sections of the module are most appropriate for your situation.

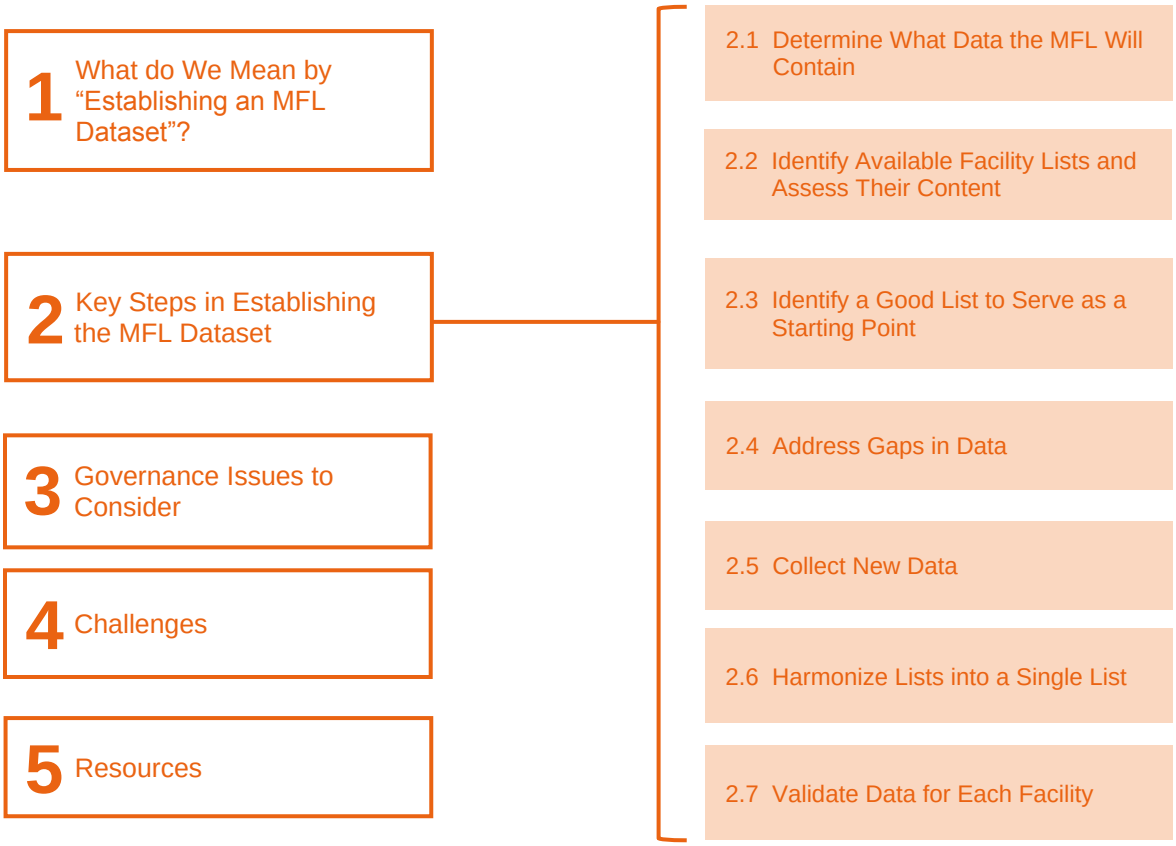
Checklist of things to do before using this module	Module where information is located
☐ Consult stakeholders to understand how the MFL data will be used	**MFL Governance Module** **Key Considerations Module**
☐ Determine what data should be included in the MFL	**MFL Data Content Module**
☐ Conduct an assessment of the MFL and available facility data in your country	**MFL Assessment Module**
☐ Set up a steering committee to lead the process of strengthening the MFL	**MFL Governance Module**

Key audiences for this module:

- Steering committee for MFL strategic planning
- Managers of the MFL
- Implementing organizations who will assist in establishing the MFL dataset

Figure 1: Establishing an MFL Dataset—Module Outline

(Press Control and click on any of the boxes to be taken directly to that section)



1. WHAT DO WE MEAN BY "ESTABLISHING AN MFL DATASET"?

Establishing an MFL dataset involves compiling the facility data you need from different sources and validating the information you have obtained. There are various approaches to establishing an MFL dataset. The best approach depends on which data are already available, the quality of those data, and how well they align with the pre-established MFL requirements.[1] Depending on the country, you may need to build a new MFL dataset from scratch, harmonize existing facility lists into a MFL, collect and add additional data to complete a MFL, or simply **validate** an existing MFL.

The module outlines the various steps and decisions that must be made in the process of establishing an MFL dataset. Some steps may be skipped for countries that are further along.

[1] For additional information see the *Key Considerations Module* and the *MFL Data Content Module*.

2. KEY STEPS IN ESTABLISHING THE MFL DATASET

2.1 Determine What Data the MFL Will Contain

Before you begin the process of establishing an MFL, you need to have determined what facility data you want included in your MFL and what format the data should be in. See the *MFL Data Content Module* for detailed information on these issues.

2.2 Identify Available Facility Lists and Assess Their Content

You will also need to have identified and assessed the existing facility lists in the country, including the MFL, if there is one. The *MFL Assessment Module* contains detailed guidance for assessing facility lists. The assessment will help you identify what facility lists exist, how complete and up to date the data are, and whether the data contained in these lists can help build or supplement your MFL.

After assessing individual lists, consider the group of lists as a whole, to determine:

- How complete are the data available?
  - Do you have data for all or most data elements of the MFL?
  - Do you have data for all or most health facilities?
- What data are missing?
- Will you need to collect new data to fill gaps?
- Are there important discrepancies across lists (for example, facility names or addresses are not the same)? Is it possible to determine where the error lies?
- Are the data in the lists defined according to the data specifications and standards you have set for the MFL?[2]

2.3 Identify a Good List to Serve as a Starting Point

If an MFL already exists in a country, and you have determined that the data are reasonably complete and of good quality—even if there are some gaps—you can skip this section.

If a country does not already have an MFL, the information you have gathered about existing lists will help determine if there is a facility list of sufficient completeness and quality that it can be used as the basis for establishing your MFL. To serve as a solid foundation for the MFL, this list must meet most of the MFL criteria for data content and quality. The list does not need to be complete, but it should:

- Be credible—you trust the data sources and methods used to update and validate the list;
- Have enough information about facilities to populate a good proportion of needed MFL data;
- Contain data that was collected or verified within the past five years.

[2] See the *MFL Data Content Module*

If some data elements or facility types are missing, those can be collected separately. The more important aspect of this step in the process of establishing an MFL is determining whether you have confidence in the validity of the data in the list you have selected. If there are too many errors, incomplete information, or other data quality concerns, it may be best to start from scratch with an entire census of health facilities. If you are unsure about data quality, you may want to do spot checks and data validation for some data elements. See Section 2.7 below, "Validate Data for Each Facility," for more information.

If none of the lists meet sufficient assessment criteria—for example, the lists are too outdated or the data are fragmented and incomplete—then you will need to consider starting from scratch and collect new facility data. For more information, see Section 2.6 below, "Collect New Data."

CASE STUDIES: STRATEGIES USED FOR ESTABLISHING THEIR MFL DATASET

Kenya: New data collection was used to create the MFL because no existing list had sufficient information to serve as a starting point. The MOH conducted the new data collection by sending each district a template that outlined the requested data fields. The districts collected the information and reported it to the MOH, which then compiled the data and used it to create the initial MFL.

Nigeria: Each state had a different process for registering and identifying facilities. The Federal Ministry of Health harmonized data from multiple lists, assigned new unique identifiers, and began a process of to remove duplicates.

Tanzania: The MFL was created by implementing both new data collection and harmonization of existing lists. The existing lists came from various sources such as government ministries, multi-national organizations, and other agencies.

Haiti: After the 2010 earthquake, Haiti realized the importance of having a list of all the health care facilities in the country. A preliminary MFL was developed by merging information from various lists and consulting knowledgeable persons. This preliminary MFL was then posted to a public website where anyone could recommend corrections or additions to the list.

Philippines: The directory was maintained by the NHFR Team at the Department of Health (DOH), with the sub-national health offices and the licensing bureau both having upload rights. This arrangement led to the creation of duplicate records and, from an initial roster of 17,000 facilities, the list ballooned to 40,000. Additionally, input into the directory was not efficient, resulting in the facilities list being both overly large and incomplete–i.e., duplicate facilities and missing and incorrect data. The NHFR Team spent a substantial amount of time cleaning the directory and flagging potential duplicate facilities, which was a challenging and time consuming process. Duplicates were matched by health facility name and location (administrative unit, geographic coordinates). The team additionally devised a set of rules for identifying potential duplicates. For example, they flagged all districts that had more than two hospitals.

2.4 Address Gaps in Data

If you have identified a facility list that will serve as the basis of the MFL, you need to clearly document any gaps that exist in this list. Gaps can be: (1) missing data (for example, there are no geocodes), (2) incomplete data (for example, a sub-set of facilities such as military hospitals are missing), or (3) data that appear to be erroneous. You will likely already have noted these gaps during the course your MFL assessment. However, at this point it is important to examine the gaps in data carefully to know exactly what additional information needs to be gathered to supplement the existing data in the facility list, and thereby create a comprehensive MFL.

Next you need to determine where to get the data needed to complete the facility list. Generally, there are two options for doing this. You can either pull data from another existing facility list, or you can collect data directly from the facilities. In some cases you will need to do both. For example, you may find a list that has the necessary data for facilities in one region of the country, but need to collect that data for other regions.

When choosing whether to pull from other lists or to collect new data, give careful consideration to the level of effort required and the associated costs. Pulling from different lists can be more complicated than expected if facility names and unique identifiers don't match. If you think the gaps in your list can be addressed using data from other available lists, see Section 2.6 "Harmonize Lists into a Single List."

If the missing data are not available in current lists, or if harmonization is not a practical solution, you will have to fill the gaps by collecting new data. See Section 2.5 "Collecting New Data" for information on how to do this.

2.5 Collect New Data

During the process of establishing an MFL, you may need to conduct new data collection. This is the case if you do not have all the information for your MFL **minimum data content**, and you cannot fill the gaps using information available in other lists. Be clear about the types of new data you need to collect; specify the following:

- Types of facilities you are including[3]
- Geographic areas where data collection will occur
- Data elements you need to gather, along with clear definitions for each
- Data collection tools or measurement approaches you will use
- Data sources you will use

Approaches to new data collection include the following:

Health facility assessment surveys—can provide new data for the MFL. Examples of such surveys are the World Health Organization's Service Availability and Readiness Assessment (SARA) and the DHS Program's Service Provision Assessment (SPA). These surveys can be administered as a census, or in select regions, or targeted to a specific type of facility, depending on the information that needs to be collected. You can coordinate with them to make sure they are collecting the data needed for the MFL per the data specifications you have decided upon. For planning purposes, you will need an estimate of the number of facilities in each geographic area and their approximate locations. A facility assessment survey will provide information

[3] See the *Key Considerations Module* for more information on determining what facilities to include in the MFL.

beyond what is needed for the MFL and may be beneficial for other purposes. However, this is an expensive option to carry out.

Targeted facility census—An alternative to a full facility assessment survey, is to send teams into the field to collect the data directly from facilities only on the data elements needed for the MFL. Sending trained data collectors improves the quality of the data you will get.

District Health Information Officers—can be enlisted to collect information about facilities in their districts. They tend to be familiar with the various facilities and can coordinate with local MOH staff to help gather specific data as needed. These efforts can either be coordinated with scheduled supervisory visits to facilities or be done separately. A simple questionnaire specifically designed for the MFL is helpful to aid the data collection process.

Crowd sourcing—soliciting contributions from a large group of people in a community or from data consumers can generate new data. In remote areas particularly, the community itself is often a necessary source of data. However, this method is not ideal because the persons collecting and sending in the information have not been trained to collect the data, and may not understand the exact definitions and measurement approaches specified for the MFL.

Collecting data on facility locations requires a more meticulous approach. For more information, see the *Geocoding the MFL Module*.

New data collection can be costly, depending on the approach that is used. Therefore, it is important to consider carefully the available budget before planning for new data collection. If the budget is limited, data collection can be done in stages, such as targeting one sub-group of districts at a time. It is important to prioritize which data are needed most urgently. Usually, data pertaining to the signature domain is critical to a MFL and should take precedence over data describing services offered at the facility.[4]

[4] See the *Key Considerations Module* for more information.

Once the data have been collected and entered into your list, you will need to begin the process of validating the data for each facility. See Section 2.7, "Validate Data for Each Facility."

CASE STUDY: NEW DATA COLLECTION

Tanzania: In addition to the harmonization of lists, a data collection tool was used to collect additional information and to validate or verify the information in the existing lists. The tool was about 4 pages and collected information on a wide range of information including: name, location, administration level, unique facility identifier, contact information, facility type (which included better categorization of facilities than was used in the existing lists), ownership, infrastructure (number of beds, rooms, transport, waste, etc.), services offered, physical location and service area population.

2.6 Harmonize Lists into a Single List

When the plan calls for harmonizing lists, keep in mind the costs and resources required to carry out this procedure. The time and effort required for matching facilities across lists and for data cleaning is often underestimated, which has the potential to cause significant delays in establishing an MFL. Ideally, you establish a technical working group or committee of experts that oversees the harmonization process.

Harmonization can serve two purposes. It can be carried out to add additional facilities and their complete records to the MFL. Alternatively, harmonization can be carried out include new data about facilities already in the MFL.

To harmonize the lists, you will first need to:

- Identify which data will be kept from each list
- Compare data element definitions (or **data specifications**) across lists that you will use to populate the MFL to make sure they are consistent
- Estimate what proportion of facilities can be matched electronically across lists. To match electronically, the facilities will need to have identical data in one or more key data elements (e.g., same unique identifiers or geocodes, or the same name plus administrative units). The inconsistent use of accents or abbreviations, and misspellings will impede electronic matching. If the proportion of non-matching facilities is high (over 25%) determine how long it will take to verify and individually cross-reference each facility between lists.

You will then:

- Match facility records to create one record per facility in the MFL.
  - Organize the facility data—you will want to organize the facilities by lowest appropriate administrative level (often, this will be district level). It is also helpful to sort them alphabetically if the matching will be done by hand.
  - Identify facility records that correspond to the same facility
    - Electronic matching is best when there is a large number of facilities. The simplest approach is to try to match facility data per district based on facility name. However, small differences in spelling can fail to electronically match a same facility across lists. Geocodes and unique identifiers are usually not consistent across lists and are not recommended to be used for matching.
    - For matched facilities, do a quick check of the other data fields for the **signature domain** to verify that the facility is in fact the same across both lists.
    - Manually—for facilities that were not matched electronically, you will have to match them on an individual basis. For facilities to be considered a match, the following data should be the same for both records
      - Facility name—a slight difference in spelling, naming, or abbreviations may have prevented the electronic match from recognizing these facilities as a match
      - Facility type (hospital, clinic, rural health center, etc.)
      - Managing authority (public, private)
      - Location within administrative units
  - Combine the information in matching records to yield one record per facility. If entries cannot be matched, flag the multiple entries for reconciliation during the validation process.
- You may also want to import data for new facilities (not included in your MFL). In this case, you do not need to merge records, but you do want to verify that the facility is not already listed under another name.

A common challenge that arises during data harmonization is discrepancies in facility-level data–spelling differences, address or location differences, and differences in the naming of facilities. If you are not sure whether some facilities are in fact matches, it will be necessary to consult district or facility personnel.

Once you have harmonized the data into a single list, you will create the unique facility identifier codes for facilities that do not already have them.[5]

When harmonization is complete, you should reassess the new list for gaps.

[5] See the *MFL Data Content Module* for more information on facility codes.

CASE STUDIES: PROCESS OF HARMONIZATION

***Tanzania*:** Harmonizing the existing lists included matching for geographic administration level, facility name, geocodes, multiple IDs, ownership, and facility type. In addition, a data collection tool was used to obtain additional information about the facilities, and to validate and verify the information in the existing lists.

***Nigeria*:** To establish Nigeria's MFL, the Federal Ministry of Health (FMOH) had to harmonize data from multiple facility lists. The goal of the process was to allocate new unique identifiers and eliminate duplication of facilities. An intelligent unique identification system was used to create new unique identifiers. Following this allocation, matching of independent identifiers across different information systems (those previously deployed in the country) was attempted. A manual matching process was employed—any facility records that were a 100% match were considered similar records and the data in the other system was used to improve the information from the primary MFL that the FMOH had compiled. Any facility records that were a partial match were reviewed further by the FMOH. The FMOH was responsible for verifying whether the data were associated with one or more than one facility and entering the verified facility information into the MFL.

2.7 Validate Data for Each Facility

Once the necessary data have been input into the MFL, you need to **validate** the data for each facility. Validating the list involves determining if you have data quality issues that need to be resolved. This is an ongoing process that should continue throughout the life of the MFL. All data in the MFL need to be validated, but accuracy of the signature domain component of the MFL is of particular importance.

It is recommended that, when possible, individuals familiar with the health facilities in their own localities (for example district health information officers) be responsible for validating the data. Data validation can be carried out through supervisory visits or through dedicated visits to the facility to determine the validity of the MFL data. If visits are not possible, telephone or email contact with the managing authority can suffice (with the exception of validating coordinates).[6]

[6] See the *Maintaining the MFL Module* for more information on validating MFL data.

CASE STUDY: PHILIPPINES VALIDATION WORKSHOPS

The Philippines organized subnational validation workshops during which the MFL team and Department of Health (DOH) representatives worked together to clean and complete the existing MFL. They reviewed the list facility by facility to verify the data and correct them as needed. This process took about one year to cover 16 of the regions in the country. These validation workshops were useful whereas previous attempts to correct and fill gaps in the data by sending lists to subnational representatives were unsuccessful due to low response and poor internet connectivity. At the same time, the workshops provided a forum for (1) training DOH representatives on the process of updating the MFL, and (2) improving their skills in the collection of data on geographic coordinates of facilities using Google Map or GPS devices.

Once you have validated the data for each facility and assigned unique identifiers, you have a validated MFL that is ready to share.

3. GOVERNANCE ISSUES TO CONSIDER

When you have determined the best process to follow in establishing your MFL dataset, you will need to set up a technical working group tasked with implementing the work. It will help with developing a work plan, creating a budget, and establishing a timeline to aid in managing the process.

Stakeholder engagement is critical at every step in the process of establishing an MFL dataset. Stakeholder buy-in and support is especially important if you want to harmonize lists owned by different groups, and reach consensus on a single authoritative list for the country.[7]

[7] For additional information on stakeholder engagement, see the *MFL Governance Module*.

4. CHALLENGES

Establishing MFL Dataset Challenges	
Challenge	**Potential solution**
Too many facility lists exist	• Work with the MFL Steering Committee or Technical Working Group to determine what lists should be assessed and considered for building the MFL • Eliminate lists that are old or contain few relevant data elements or are duplicative of data available elsewhere. • Determine whether harmonizing many lists is more or less cost-effective than new data collection.
Too many data elements included in the MFL	• Work with the MFL Steering Committee or Technical Working Group to determine what the minimum content should be and prioritize the need for additional data • Consider the costs • Fully understand the implications and costs of adding more data elements to the list, what issues might arise, and how those issues will be dealt with • Consider the costs and data collection burden required to keep the list up-to-date in the long-term
Non-standard facilities (i.e., not "brick and mortar") or mobile facilities	• Determine what sources of data will be used for these types of facilities • Determine what value their inclusion in the MFL will bring against added cost of doing so • Location data will need to be general (for example, just naming the district where they operate) • Anticipate needing to validate data for these types pf facilities more frequently as they tend to change more often
Addressing discrepancies between MFL and externally managed facility lists (donors or civil society)	• Work with the MFL Steering Committee, Technical Working Group and list owners to determine how data discrepancies will be addressed • Validate data in the MFL to make sure it is accurate

5. RESOURCES

- Development of a Master Health Facility List in Nigeria
- Development of a Master Health Facility List: Haiti's Experience
- Tanzania MFL data collection form
- Rwanda MFL data collection form

ACKNOWLEDGEMENTS

The MFL Resource Package was developed with extensive input from a team of persons who have been involved in various capacities in the development or management of MFLs in different countries. The content builds off of previous MFL guidance developed by the World Health Organization, MEASURE Evaluation and Open HIE. This MFL Resource Package seeks to expand and update the guidance and make it accessible to a wide audience. Development of this Resource Package included a literature review, a series of in-depth interviews with key informants, a three-day meeting attended by various experts in this area to discuss in detail the content and structure of the guidance document, and a thorough review process.

Cristina de la Torre and Clara Burgert from ICF led the development and drafting of this guidance document. Lwendo Moonzwe, and Kirsten Zalisk (from ICF) and Aubrey Casey (formerly from ICF) helped to draft the MFL Resource Package, organize resources, and document discussions during the three-day meeting. Andrew Inglis (formerly from MEASURE Evaluation/JSI) and Scott Teesdale (from InSTEDD) helped draft sections of the MFL Resource Package.

Lynne Franco led a team at EnCompass to conduct a series of in-depth interviews to inform the content of the Resource Package, and subsequently helped facilitate the three-day meeting to review the guidance proposed for the MFL Resource Package.

The following tables list persons who contributed to the MFL Resource Package by attending a three-day meeting, participating in in-depth interviews, contributing resources, reviewing drafts or providing information for the case studies.

Table 1: Persons who participated in the three-day meeting to review the content and structure of the Resource Package.

Meeting Participants	Affiliation
Tariq Azim	MEASURE Evaluation/JSI
Noah Bartlett	USAID, Bureau for Global Health
Clara Burgert	The DHS Program/ICF
Aubrey Casey	The DHS Program/ICF
Niamh Darcy	RTI
Anita Datar	Health Policy Project/Futures Group
Cristina de la Torre	The DHS Program/ICF
Mark DeZalia	PEPFAR/CDC
Lynne Franco	The DHS Program/EnCompass
Erick Gaju	MOH Rwanda
Nate Heard	US Department of State

Meeting Participants	Affiliation
Andrew Inglis	Deliver Project/JSI
Denise Johnson	MEASURE Evaluation/ICF
James Kariuki	PEPFAR/CDC
Esther Kathini	MOH Kenya
Carl Leitner	iHRIS/Capacity Plus/IntraHealth
Lwendo Moonzwe	The DHS Program/ICF
Annah Ngaruro	MEASURE Evaluation/ICF
Kola Oyediran	MEASURE Evaluation/JSI
Jason Pickering	Consultant/DHIS2
John Spencer	MEASURE Evaluation/UNC
Charity Tan	MOH Philippines
Scott Teesdale	Open HIE/InSTEDD
Kavitha Viswanathan	WHO
Sam Wambugu	MEASURE Evaluation/ICF
Kirsten Zalisk	The DHS Program/ICF

Table 2: Persons who contributed through interviews or review of the MFL Resource Package Modules.

Name	Affiliation at time of participation
Ian Wanyeki	Health Policy Project/Futures Group
Elaine Baker	Health Policy Project/Futures Group
Bernard Mitto	Health Policy Project/Futures Group
Vanessa Brown	PEPFAR/Department of State
Robert Colombo	WHO
Steeve Ebener	Gaia Geo Systems
Mike Gehron	PEPFAR/Department of State
Karin Gichuhi	Office of HIV/AIDS/USAID
Marty Gross	Bill & Melinda Gates Foundation
Jason Knueppel	BAO Systems
Rachel Lucas	USAID
Andrew Muhire	Rwanda MOH
Martin Osumba	AFYAinfo, Kenya
Alyson Rose-Wood	Office of Global Affairs/HHS
Dykki Settle	iHRIS/IntraHealth
Jim Setzer	Abt Associates, Inc
Ashely Sheffel	Consultant/WHO
Brian Taliesin	Digital Health Solutions/PATH
Ola Titlestad	DHIS2/University of Oslo

The MFL Resource Package was undertaken with support from the United States Agency for International Development (USAID) and the President's Emergency Plan for AIDS Relief (PEPFAR) through The DHS Program.