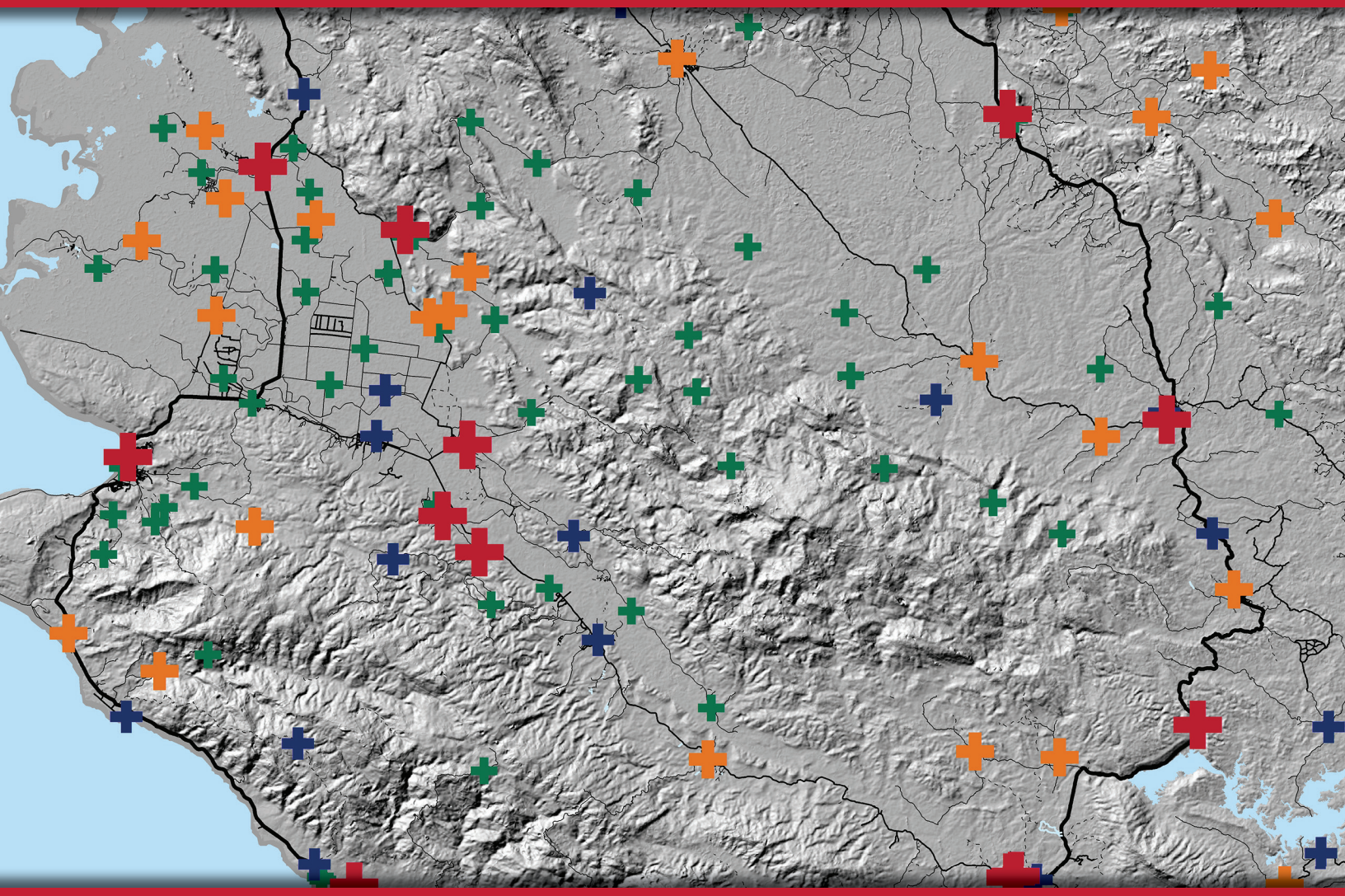
USAID
FROM THE AMERICAN PEOPLE

MASTER FACILITY LIST RESOURCE PACKAGE:

Guidance for countries wanting to strengthen their MFL

Module 6: Geocoding the MFL



May 2017

DRAFT

GEOCODING THE MFL

This module provides guidance on the procedures for assigning geocodes to facilities in the Master Facility List (MFL). The module covers key aspects of geocoding, such as selecting schema, methods for obtaining geocodes, validation of geocodes, processes for maintaining geocodes, and considerations for sharing a geocoded MFL.

Checklist of things to do before using this module	Module where information is located
☐ Identify the main users of MFL data and documented their requirements	**Key Considerations Module**
☐ Decide the minimum data content of the MFL	**MFL Data Content Module**
☐ Assess content of existing facility lists	**MFL Assessment Module**
☐ Established a Steering Committee for the MFL	**Governance Module**

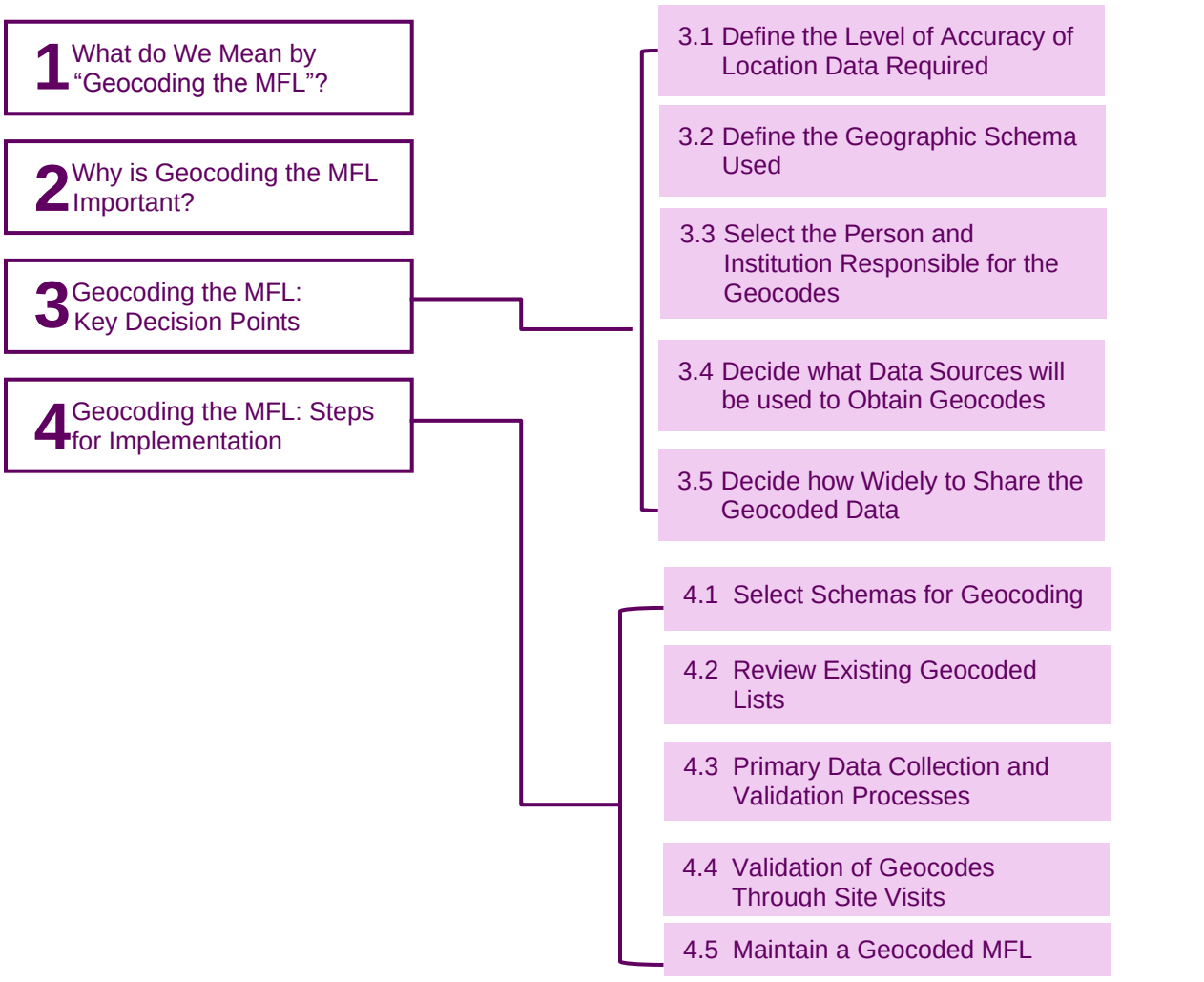
Key audiences for this module:

- Technical staff responsible for collecting or validating geocoded data for the MFL
- Managers who oversee the MFL

Note: words in **bold** are defined in the glossary.

Figure 1: Geocoding the MFL — Module Outline

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



1. WHAT DO WE MEAN BY "GEOCODING THE MFL"?

Geocoding the MFL entails gathering and assigning physical location data (typically using geographic coordinates for latitude and longitude) to each health facility included in the MFL. While an MFL usually includes different types of geographic identifiers, such as the administrative unit (e.g., province or district) in which the facility is found, geocoding provides a more precise location for the facility that can be visualized as a point on a map. Conceptually, geocoding is a simple process but implementation can be complex. There are a number of different approaches to geocoding an MFL, each with costs and benefits attached. The module outlines the most common approaches to geocoding, and discusses key considerations to be reviewed when determining the method(s) to be used. It should be recognized that when geocoding an MFL, it is common practice to use a combination of methods.

2. WHY IS GEOCODING THE MFL IMPORTANT?

An MFL with geocoded facility data has various advantages.

- Geocoded facility data helps to better manage health programs.

  Knowing where health facilities are located and where specific services are offered is critical to managing national health programs and targeting interventions efficiently.

  - Facility location data can help examine questions related to access, equity and gaps in service provision. Accurate location information about health allows health planners to target interventions, review and assess the impact of programs, and plan future activities.

  - With geocoded data it becomes possible to easily visualize and query the MFL using a mapping program such as Google Earth or by using a Geographic Information System (GIS). Mapping the locations of facilities helps in visualizing the health facility landscape. The geographic information enables coordination and management of services by identifying areas of high or low concentration of activities and then making adjustments to service locations to improve service availability.

  - Location information can help with budgeting and planning for activities that involve the transport of goods and human resources, such as supervisory visits or the delivery of commodities.

- Geocoding the MFL makes it possible to link the MFL to other datasets using geography.

  Using geocodes is one approach to integrating data from different sources. Linking the MFL data to other datasets allows for greater insights into health programs and their interaction with factors that can influence their effectiveness. From a geographic perspective, it can be of value to understand the location of facilities and services relative to factors such as population distribution (overall population or key populations), transportation networks, markets, climate, or agricultural patterns. The key to this process is having other geocoded datasets—also referred to as "geo-enabled data" (data that can be mapped)—that can be linked to the geocoded MFL.

- Geocoding assists with management of the MFL

  Including geocodes in an MFL facilitates the management of the list. Knowing the exact location of facilities can help identify duplicate listings in the MFL (if for example a facility had been entered twice with a different name). If district lines are redrawn, geographic

coordinates can also help correctly reassign facilities in the MFL to their new administrative unit.

To be most effective, the MFL should include the location of every health facility in the system, however even partially complete geocoded data in an MFL can be beneficial. It is better to have some locations in an MFL than none at all because of the benefits these data provide.

3. GEOCODING THE MFL: KEY DECISION POINTS

A number of key decisions must be made before initiating the process of geocoding an MFL. Some of these decisions will need to be reviewed and adjusted as the geocoded MFL evolves over time. Most decisions need to be made in conjunction with other MFL key decisions, such as those relating to MFL maintenance procedures or decisions related to sharing the MFL data. In this section we highlight the key decisions and in section 4, we provide additional technical detail related to the implementation of these decision points.

The six main decision areas are:

1. Define the level of accuracy required
2. Define the geographic schema used
3. Select the person and institution responsible for obtaining and maintaining the geocodes
4. Decide what data sources will be used to obtain the geocodes
5. Establish the frequency of update and review
6. Decide how widely to share the geocodes

3.1. Define the Level of Accuracy Required

The required level of accuracy of the geocodes depends on how the MFL will be used.[1] If the MFL is primarily used to map locations of facilities for planning purposes, then having the location of the village or town where the facility is located is likely sufficient. If the MFL must be used for navigating to facilities for supervisory visits or for the provision of commodities, then highly accurate location for each facility is needed. Some types of analytics may require more precise locations, for example if one wants to study how distance to facilities affects uptake of services and health outcomes.

The level of accuracy desired affects the level of effort required to obtain the geocodes. Precise locations require visiting the facility to collect the exact geocodes, whereas less accurate locations (such as situating the facility within a town) may be done remotely using software such as google maps.

[1] For more on understanding user requirements for the MFL see the *Key Considerations Module* and the *Establishing a Facility Registry Service Module.*

The decision on accuracy should be reviewed at least every few years to maintain viable geocodes that continue to meet the needs of the MFL users. If the MFL **user requirements** shift, or the cost of obtaining geocodes changes, then the required level of accuracy can be altered. It is important to note that such changes will likely have consequences on the methods used to collect and validate geocodes.

3.2. Define the Geographic Schema to be Used

The *schema* of the geocode is the primary data format used to store the geocodes. Section 4.1 describes the various types of schema formats that can be used. Each format has its own nuances of structure; for example, under geographic coordinates there are the following variations: (1) degrees, minutes, and seconds, (2) degrees and decimal minutes, and (3) decimal degrees. Therefore, the schema needs to be clearly defined and documented so that users of the MFL can access this information.[2] Note that as long as the schema is known, it is easy to convert between different schemas.

Schema Recommendation

The recommended starting point for the schema is to use latitude and longitude in decimal degrees, with WGS84 as the datum. The reasons for this are: (1) latitude and longitude are widely understood, (2) decimal degrees are easy to review and to identify data issues, and (3) WGS84 is a common datum. Together they make it easy to integrate the MFL data with other geographical data.

When deciding which schema to use, consider the following:

- the primary use of the geocodes
- the needs of the MFL administrators
- the needs of other stakeholders with whom the list will be shared

If the geocodes are to be shared widely, then use of a well-known schema is recommended. Generally, it is recommended to match the schema with the primary use of the geocodes, then document its characteristics (data type, format and datum) and share these details with users of the MFL.

3.3. Select the Person and Institution Responsible for the Geocodes

When deciding who to charge with maintaining the geocodes it is important to understand the roles and responsibilities this job entails, and how they will interact with the other roles and responsibilities associated with the MFL.[3] The responsibilities of the team charged with managing the MFL geocodes are described below:

[2] The *MFL Data Content Module* has information on defining and documenting data specifications for the MFL.

[3] See the *Maintaining the MFL Module* for a description of the various responsibilities associated with the long-term maintaining and management of the MFL.

1. Establishing and implementing the procedures for geocode collection
2. Identifying processes for validation of geocodes
3. Assessing and validating geocodes
4. Maintaining and sharing the MFL geocodes
5. Responding to inquiries regarding the geocoded data
6. Documenting the procedures for geocoding the MFL and making them accessible to users

These responsibilities can be divided up or assigned to one person, depending on the size of the MFL. They can be managed centrally or regionally as long as there are standardized procedures. These tasks can be contracted out as long as there is close oversight and coordination to ensure alignment with other MFL maintenance processes. Typically, the person(s) tasked with managing the geocodes should be within the same institution and unit as others managing the overall MFL to facilitate management and coordination. It is recommended that more than one MFL team member be involved in the geocoding process so that with the departure of key staff, geocoding skills are not lost.[4]

3.4. Decide what Data Sources will be Used to Obtain the Geocodes

A key decision with regard to geocodes is whether you will collect new data or use existing geocodes from other facility lists. The decision will depend on what data are available, how trustworthy they are, and whether they meet the accuracy criteria you have previously established. The availability of financial and human resources for new data collection also needs to be considered.

Many existing facility lists already have geocodes, which can be used as a foundation for generating a geocoded MFL.[5] A logical first step is to review these lists and determine what information can be pulled from them. Section 4.2 provides additional details on how to review, assess and validate existing geocoded data. Keep in mind that the pulling data from existing lists requires matching facilities across lists which can sometimes be a time-consuming and cumbersome process.[6]

If new data collection is required for all or a subset of facilities, it is important to develop specific plans and procedures for the collection of missing geocodes and for validating them within the MFL. Visiting all facility locations solely for the purpose of collecting geographic coordinates is a time- and resource-intensive activity that has limited value, and thus is probably not a viable option. Another means of data collection is to link the collection of health facility geocodes with other scheduled visits to health facilities such as supervisory visits, facility surveys, and commodity deliveries. Data collection for geocodes should be coordinated

[4] See *Maintaining the MFL Module* for more information on MFL roles and responsibilities.
[5] For more information on review of existing facility lists, see the *MFL Assessment Module*.
[6] For more information on harmonizing lists, see the *Establishing an MFL Dataset Module*.

with other MFL data collection or data validation activities.[7] Section 4.3 describes best practices for collecting geocodes.

3.5. Decide How Widely to Share the Geocoded Data

The decision on whether and how to share the geocodes for the MFL should be done in accordance with the policies developed around governance of the MFL. The *Sharing the MFL Module* discusses the importance of a MFL sharing policy and describes various factors to consider in developing such a policy. That information is also relevant to the sharing of geocodes.

Among the factors to consider regarding the sharing of geocodes is the trade-off in value between the utility of the data to MFL stakeholders and the sensitivities associated with the data. For instance, the location of military clinics could be considered sensitive. The MFL Steering Committee and MFL Managers need to carefully consider what data to make available and to whom, and to develop written policies around this.

Data Access

When sharing a geocoded MFL, the sensitivity of the location data needs to be taken into account. It may be necessary for some data to be restricted. The following will need to be carefully defined prior to sharing the data:

1. Who has access to the geocodes
2. What procedures are required to receive access to the geocodes
3. Are separate steps necessary to access the geocodes (compared to other MFL data)
4. Do specific types of sites require separate access policies (e.g., health facilities located at military establishments)

Data Formats and Metadata

In addition to sharing the geocodes it is important to also share the metadata, or information about:

- how the geocodes were collected
- When were the geocodes collected
- what exact schema used
- the level of accuracy used

It is necessary for data users to have this information (1) to manipulate the geographic data elements in the MFL, and (2) to integrate MFL data with other geographic data.

[7] For more information on data collection approaches for the MFL, see the *Establishing an MFL Dataset Module*.

Feedback on Geocoded Data

Receiving and incorporating feedback from MFL data users on the accuracy and utility of the geographic data within the MFL is important. It is likely that MFL data users will eventually become reviewers of the quality of the lists of geocodes and can help in identifying errors and improving data quality. Establishing a mechanism whereby errors can be flagged and suggested changes submitted for review is important. For more in this please refer to the *Maintaining the MFL Module*.

4. GEOCODING THE MFL: STEPS FOR IMPLEMENTATION

This section provides additional technical information and recommendations for establishing a geocoded MFL dataset.

1. Selecting schemas
2. Reviewing existing geocoded lists
3. Primary data collection and validation processes
4. Maintaining the geocoded MFL

4.1. Selecting Schemas for Geocoding

The schema of the geocode is simply the data format in which the geocodes are stored. It is possible to have more than one schema, though it will be necessary to define one as the primary schema. Selecting the schema(s) will require consideration of how the geocodes will be used and the needs of the users of the MFL. For instance, if a particular schema is commonly employed in other databases accessed by MFL users, then it may be helpful to select that schema. However, it is important to note that if a schema is known, it is generally easy to convert between different schemas. Common data schema types are described below.

Address

A formal physical address can be a schema. Usually it includes a block number, street name, and city. While a physical address can be part of an MFL, the formal physical address infrastructure is not always complete or well-known and can be subject to change. Therefore, the physical address is not generally recommended as the primary geocode, although it is useful information to be maintained in the MFL.

Advantages:

- Provide information that is useful to navigate to a facility location without needing GPS or other navigation devices

Disadvantages:

- Requires as a prerequisite that a formal address structure is in place and electronically available that can be used for geo-referencing the address so the location can easily be mapped

Resources needed:

- List of facilities with the physical addresses

Coordinate Systems

A coordinate system is a reference system for pinpointing locations in relation to one another. They have three main elements, reference point (0,0), units of measure (meters, degrees, etc.) and mathematical algorithm (datum) representing the curvature of the earth. Based on a coordinate system you can use GIS software to generate maps of different locations. Below are two common coordinate systems and considerations if a local coordinate system exists.

Latitude/Longitude

This is perhaps the most well-known geographic coordinate schema, and is commonly used in computer systems. For instance, if you put latitude and longitude into Google, a map will be displayed of the location. The unit of measure is degrees measured in latitude and longitude.

- **Latitude** measures north/south location**.** North of equator values range from 0-90 degrees with 0 being the equator. South of the equator values range from 0 to -90 degrees;
- **Longitude** measures east/west location. East of the prime meridian (which runs through Greenwich UK) values range from 0 to 180 degrees, and west of the prime meridian, values range from 0 to -180 degrees.

A degree is approximately 111 km at the equator, which reduces as you move further from the equator. This is the default schema used in many geographically enabled devices such as smartphones and tablets. As stated earlier in Section 2.1, this schema is recommended for the MFL because it is commonly used.

Advantages:

- Commonly understood by people and computers
- Immediately available for mapping and navigation to location
- Default schema for many data collection tools (e.g., GPS units, mobile phone data collection applications)

Disadvantages:

- Possibility of data collection errors in the field. Recording latitude and longitude requires recording up to 13 digits, and small errors can lead to the location being incorrectly recorded.
- Potential errors if data format is not clearly defined and followed. If the data format, whether Decimal Degrees (DD), Degrees, Decimal Minutes (DDM), or Degree, Minutes, Seconds (DMS), is not clearly stated or known, the ability to determine the true location is limited.
- Only useful for navigation; unable to determine distance between locations without conversion.

Resources needed:

- GPS-enabled device (either GPS unit or device with a GPS chip such as a smartphone or tablet)

UTM Universal Transverse Mercator

This coordinate schema is based on a system that divides the world into 60 zones, each one being 6 degrees of longitude in width. Within each zone, location is defined as *easting* and *northing* in meters from the origin point for each zone. Easting refers to the eastward-measured distance (or the *x*-coordinate), while northing refers to the northward-measured distance (or the *y*-coordinate). This schema is often used for measuring *length* (e.g., length of a river) and *area* (e.g., size of a forest).

Advantages:

- It is a commonly used schema
- Units are easily understood (meters)
- Able to determine distance between locations without conversion

Disadvantages:

- Possible data collection errors in the field. Recording Easting and Northing measurements requires up to 13 digits, which increases the chance of small errors that lead to the incorrect location being recorded.
- Need to know which of the 60 UTM zones the coordinates correspond to

- If the area of interest extends across more than one zone (e.g., Tanzania extends across 3 zones), the GPS device needs to be programmed to the proper UTM zone before taking a location recording. Failing to do so can lead to distortion in actual location. Additionally, areas that cross the equator can introduce complications around use of UTM zones.

Resources needed:

- GPS-enabled device (either GPS unit or device with a GPS chip)

Local Coordinate System

Some countries or regions of the world have their own coordinate system, defined (i.e., designed) to better match the curvature of the earth across a smaller area such as a country, as opposed to the entire planet. This schema is usually based on historical mapping units.

Advantages:

- If the local coordinate system is widely used at the local level, integrating the MFL data with local GIS datasets will not require conversion.

Disadvantages:

- Conversion to other common data formats and data sources outside the country can be complex. For example, gridded population, road network (open street maps) are in global common coordinate systems and for them to be used with a local coordinate system requires conversion. In such a case, a global common coordinate system is preferred over a local coordinate system.

Resources needed:

- Parameters of the local coordinate system and the ability to be added into the GPS device
- GPS device (either GPS unit or device with a GPS chip)
- Conversion algorithm to and from common coordinate system(s) so that MFL can be mapped over geographic data.

New and Innovative Geographic Identifiers

In addition to traditional methods of collecting geographic location, the advent of mobile technology has initiated new and innovative ways of collecting and storing geographic location. One of the new methods being used is *What3words*.

What3words is a new schema, a location system that divides the world into squares three meters on a side (3mx3m); it then assigns a unique 3-word identifier to each square. This algorithm

provides a way to identify which square your facility is in. Using a specially designed app, you can type in the name of the square, and it will map it for you. It thus becomes a communicable address like any zip code or street address, but accurate to 3mx3m. This type of schema may reduce errors in coordinate reporting by using words rather than long series of digits to record a location. The *What3words* schema requires use of an app and is best suited for locations where smartphones are available and used. The system is being used in over 170 different countries by the World Bank, the United Nations and others, including national governments and postal services. Generally, it is not the main geocoding scheme used for MFL, but it can augment a more traditional scheme.

Advantages:

- Using a smartphone that collects, stores, and records 3 words is more reliable and less likely to have data collection errors than trying to record 13 digits.
- Can be incorporated with other data collection apps on smartphones using SDK or API
- Can be used for offline navigation with smartphone
- Is available in multiple apps and across GIS platforms including ArcGIS and QGIS
- Free batch conversion or use of the API to convert to and from latitude/longitude coordinates for use in other maps or datasets (some users prefer converted data)
- Limited technical skills required to use the system
- Can easily be written, spoken, or sent digitally

Disadvantages:

- Primarily designed for use with smartphones
- Needs to be converted to be used in certain maps
- Still new and requires some technological skill to obtain the full benefit of *What3words*

Resources needed:

- Smartphone or tablet with a GPS chip.
- Internet access if manually converting to latitude and longitude, but not for any tool or application with the SDK inside

4.2. Reviewing Existing Geocoded Lists

Once the decision has made on the schema to be used for the MFL, the next step is to review existing facility lists containing geocodes to determine which existing geocodes can be used and which sites will require a geocode be obtained. This activity should be included in the overall MFL assessment. The outcome of the assessment should place the geocodes for each facility in one of three categories: validated (ready to use), provisional (needs to be validated), or to be collected (missing geocode). Facilities that have a geocode that meets the required accuracy from two independent sources can be classified as *validated*; facilities that have a geocode that meets required accuracy from only one source can be classified as *provisional*; facilities that do not have geocodes that meet the required accuracy should be classified as *to be collected*. This classification helps determine which existing data can be used and determines the level and type of action required to complete the MFL with corresponding geocodes for each health facility.

Assessment of Completeness, Accuracy and Quality of Existing Geocodes

It is important to look at the overall completeness of the geocodes from the available list(s), the level of accuracy, the date of the last update, and the method of data collection. This information provides the foundation for determining whether the data can be used and whether facilities need geocodes to be collected.

Completeness: you will need to determine how many of the facilities in the list(s) have geocodes, and how many are missing.

Accuracy: you will need to determine if the accuracy of the existing list(s) meets the accuracy requirements defined for the new MFL. For example, if the data are used for navigation, aiming to arrive within 10 meters of the site, and the schema is geographic (using decimal degrees), then the accuracy of the geocodes would need to be to the fourth decimal place to meet the stated accuracy requirements.

Quality: you will need to determine whether the location data are correct. To do this, you can map the location to see the geocodes fall within the recorded administrative area, and whether they are plausible (e.g., they don't fall in the middle of a body of water). You should also verify that the geocodes correspond to the right facility. When reviewing existing geocode data, it is important to examine how they were collected to determine how reliable the data are.

Once the useable geocodes have been identified in existing lists, the next step is to compare geocodes from different lists with each other. The aim is to use the existing lists to determine the validity of each geocode. Codes that match lend assurance to their validity. Matches do not need to be exact, but the locations should be within a reasonable range of each other, for example within 100 meters (the exact range you set will depend on the level of accuracy you

have determined needing in the MFL). Where large discrepancies exist, further verification will be needed. The outcome of the review process will be a list of facilities, categorized according to the status of their geocodes (validated, provisional, and to be collected). The results make it possible to determine the level of effort needed to completely geocode the MFL. Following this process, you still need to further validate the geocodes, as described in section 4.4 below.

4.3. Primary Data Collection and Validation Processes

Assigning a geocode to a location in the MFL is a simple concept. The complication comes from ensuring that all geocodes within the MFL can be proven to be reliable. If even a few geocodes are found to be incorrect, the whole list of geocodes can come into question and trust may be lost in the reliability of the MFL. Establishing sound procedures for collection and validation, documentation of processes, and documentation of the source for every geocode can help build trust in the MFL geocodes. Providing transparency through documentation enables data users to know the quality of the geocodes within the MFL, and thus trust the geocodes.

The process of collecting and validating geocodes must necessarily be coordinated with overall MFL management. Three steps are recommended: (1) primary data collection of the geocode, (2) data quality check, and (3) data validation check by revisiting the site. The third step is optional because while revisiting the site is ideal for validation, it may not be practical in some instances. How these three steps are implemented should be well documented to demonstrate the reliability of the geocodes within the MFL.

Action Plan to Fill in Gaps and Validate Existing Data

Once the quality and percentage of geocodes for all facilities is known, the next step is to develop strategies to deal with any gaps in the data. This action plan will depend on a variety of factors such as (1) type, size, and distribution of data gaps, (2) resources available for data collection (people, equipment and funds), (3) level of effort required to carry out the data collection method(s), and (4) importance of geocoding the MFL. The plan should be documented to assist with securing funds and for collaboration with other activities.

There are two main strategies for collecting geocodes using a GPS receiver or a GPS-enabled device:

1. Targeted visits to facilities solely to collect geocodes
2. Adding GPS data collection to routine or planned visits to facilities

The primary difference between the two approaches is in the level of control associated with the data collection process, timing of data collection, need for coordination with other activities, and overall cost.

Targeted visits to facilities to collect geocodes are a focused effort to collect site coordinates and can provide high quality data in a timely fashion. However, cost can be a factor in using targeted visits so this approach is rarely used.

Including GPS data collection with other activities that bring people to facilities (e.g., supervisory visits, commodity delivery) is one option for collecting new data or for validating existing data. It is important to keep in mind that "pairing" with another activity requires coordination and can result in additional time being needed to complete the geocoding process. Added effort and oversight may be needed to ensure that proper data collection methods are used.

The action plan for addressing data gaps can be a combination of approaches: targeted visits and opportunistic collaboration with other routine visits. As the MFL becomes more established, the recommendation is that an increasing proportion of GPS data collection be done through collaboration with other routine data collection. It is also recommended that when finalizing the plan for filling in the gaps in geocodes, a phased approach to long-term processes for data collection (for new sites) should be specified.

Regardless of the approach, it is imperative to have clear written protocols and training materials on how GPS data are collected in the field, stored, and transmitted.

Finally, for both primary data collection and validation of geocodes, implementation of a feedback mechanism to identify and report issues with geocodes should be considered.

Primary Data Collection of Geocodes

When visiting a health facility, a geocode can be collected using a *GPS receiver* or a *GPS-enabled device* (e.g., smartphone or tablet). Each device has advantages and disadvantages, although there are a number of similarities in the process. Below are key points for weighing the benefits of the two devices.

GPS Receivers

A stand-alone GPS receiver is primarily designed for personal navigation but it can be a useful tool for capturing geocodes. Many GPS receivers are ruggedized so they can function well in remote locations. Each receiver has its own process for capturing a geocode but in general the receiver will display the geocode location and store it on the GPS receiver for download later. Because they vary, it is important to refer to the receiver's instruction manual for specific information about how to capture and download. The primary advantage of GPS receivers is that they are specifically designed to handle geocodes. At the same time, this characteristic is also a limitation because the GPS receiver is a single-task device that cannot easily collect any other attribute information.

Smartphones or Tablets

There has been a rapid growth in the use of smartphones for data collection. Most tablets, smartphones, and even some basic phones come with a built-in GPS chip and therefore can capture location information. The quality of GPS chips has improved to the point where these devices can have the same accuracy as a GPS receiver. A key advantage of using smartphones or tablets for collection of facility geocodes is the ability to capture other health facility information as well through specially designed data collection forms. This makes it possible to pair the geocode data with other data, such as facility name and services offered, ensuring that the geocode is correctly linked to the right facility. It is also possible to sync the directly with the MFL database if connectivity and the facility registry solution permits. Typically, data collection on these devices requires an app specially designed to capture data and geocodes. These apps can be found in the device's relevant app store or custom apps can be created using tools such as ODK or iFormbuilder.

When collecting geocodes at facility sites, it is important to also record the method used (GPS devices, smartphone, map) and the date and time of the data collection. This information is useful for documenting the source of the geocode and for checking data quality. With the advent of electronic data collection and near real-time data transfer, there is opportunity for rapid data transfer and feedback. The value is that the electronic data collection can allow feedback while data collectors are still in the field. However, expectations and understanding of roles and responsibilities need to be well established to ensure that the feedback mechanism is used effectively.

Cartographically Determining Location

When there are limited opportunities to physically visit a site, then locating the site using cartographic methods is a temporary option (until the location can be visited). The main disadvantage of using cartographical methods is lack of confirmation that the right geocode has been assigned to a particular facility. If this method is employed, using multiple sources such as Google Earth, topographical maps, and people with local knowledge, can minimize the likelihood of the geocode being incorrect. When a cartographic method is used, it should be documented in the MFL and viewed as an interim source of information until a geocode from the location can be obtained.

Potential Data Issues during Data Collection

In the process of data collection of geocodes, potential issues regarding the list of facilities may come to light. For example, a common type of data issue is duplication – the same location has two names. For example, Alpha District Hospital may also be known as St. Paul Hospital. In this case, the duplication needs to be documented and eventually resolved with the MFL maintenance team. Therefore, as part of in the data collection process, there needs to be a

procedure in place for documenting discrepancies in the facility list and resolving them according to the MFL structure.

Documenting Data Collection Procedures

It is important to have well documented procedures for the collection of geocodes to ensure that they are obtained in a standardized manner. In addition, the process for physical data collection and recording requires training materials to provide the data collectors with clear instructions on how to collect geocodes. The storage of data and the process of transferring the data collected will also need to be agreed upon and documented.

Data Quality Checks

After collection of new geocodes, the data must be reviewed to ensure data quality of the geocodes before they are added to the MFL. There are multiple ways of checking whether the geocode is representing the correct facility at the correct location. However, the best way to validate a geocode is for someone to physically visit the site and confirm that the geocode there matches the recorded geocode.[8] Below are ways of checking the validity of geocodes that do not require going into the field. The shared principle behind these methods is assessing the validity of geocodes by comparing them with other known geographic information.

1. Do the coordinates conform to the MFL schema?
2. When mapping geocodes, compare them against other known locations or landmarks, e.g., other health facilities in the MFL
3. Do the geocodes appear in or near the border of their associated administrative unit? If a location is near or on the border of its administrative area then it may be correct, but other data are needed to confirm it is in the right location.
4. If it is a large health facility, is the location on or near a road? Is it in the center of a town? Larger health facilities are always located on or near a road and usually in a town center. If it is not, then other data are needed to confirm it is in the right place.
5. When mapped over imagery, is the location an improbable site, e.g., in a river, in a dense forest, or in the middle of open plains?

After the geocode is reviewed and identified as a viable location by checking the geocode against other known geographic data, then it becomes available for provisional use. The geocode will still need to be validated but it may be some time before there is an opportunity for validation by revisiting the site.

[8] It is considered a match when both coordinates are within 10 meters.

4.4. Validation of Geocodes through Site Visits

Validation of geocodes is required regardless of the data source (a pre-existing list or new data collection). Validation serves to verify that geocodes are correct and that they have been assigned to appropriate facility within the MFL. Validation of location data can be done when a site is being visited for other purposes (e.g., commodity delivery or supervisory visits). The process entails re-collecting location data and comparing it to the data in the MFL. A pre-defined margin of error (i.e., acceptable deviance between the two location readings) should be pre-determined. If there are discrepancies, it is important to review the methods used to collect the geocodes and to re-verify until two separate readings provide the same information.

The opportunity for validation through revisiting a site usually requires both a willingness to coordinate with all parties involved, and that all parties involved see the value of validating geocodes during the site visit. It also requires careful training of the teams going to the field as well as written instructions on how to collect, store and transmit the data.

As with other changes and updates to the MFL, geocoded data should include the date when they were collected and date of verification, to show that the geocoded MFL is being well-maintained.

4.5. Maintaining a Geocoded MFL

As with all information within the MFL, geocoded data need to be maintained to ensure that reliability and trust in the MFL are retained. Effective maintenance recognizes that the MFL will have to accommodate regular changes in the list, and resources will need to be set aside specifically to maintain geocodes.

Together, the increased use of electronic data and the integration of GPS sensors into smartphones provide an opportunity to decrease the burden of physical data collection and management of geographical data through the greater use of IT innovations. Regardless of the technology used, four main factors are suggested to ensure sound maintenance of geocoding:

- Create processes to identify changes in the MFL and whether they trigger a likely change in the geocode. If the change in the MFL does trigger a possible change in the geocode, then the site is added to a tracking sheet of geocodes to be checked and either replaced or validated.
- Setup and document the standard methods for checking the geographic data.
- Update documentation of the geographic data elements, generally known as “metadata.” The metadata is descriptive documentation that contains information on the schema, datum, method of data collection, and format of the data elements in a geographic dataset. The metadata will need to meet national geospatial metadata standards.

- Schedule regular mapping and review for tracking changes in the geocodes. This should be done in coordination with review of the rest of the MFL. The aim is to review how well the MLF has been maintained and to improve the maintenance process to increased reliability and reduce maintenance costs.

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.