CREATING A MASTER HEALTH FACILITY LIST

World Health Organization
March 2012

ACKNOWLEDGEMENTS

This document was developed through a collaborative process with inputs from a multitude of organizations. We would like to thank the following individuals for their contributions:

WHO Department of Health Statistics and Information Systems
Coordinator: Kathy O'Neill
Contributors: Claire Preaud, Ashley Sheffel, Marina Takane, Laurens Zwakhals, Dag Roll-Hansen

Consultant
Jason Pickering

International Health Facility Assessment Network
Bolaji Fapohunda ,Natasha Kanagat

Kenya Ministry of Medical Services, Kenya Ministry of Public Health and Sanitation
Dr. Bill Martin Osumba

US Agency for International Development
Noah Bartlett, Lisa Maniscalco

US Agency for International Development's MEASURE DHS project (IFC International)
Clara Burgert, Blake Zachary

US Agency for International Development's MEASURE Evaluation project
Andrew Inglis, John Spencer

US Department of State
Nathan Heard

Particular thanks is extended to all those who participated in the workshop on "Building Sustainable National Databases of Health Facilities and Service Capacity" held in Geneva, Switzerland 20-21 September 2010.

World Health Organization be liable for damages arising from its use.

TABLE OF CONTENTS

ACRONYMS

ARV	Anti-retroviral
CIQ	Customer Information Quality
CSO	Central Statistics Office
CSV	Comma Separated Values
FBO	Faith Based Organization
GIS	Geographic Information System
GPS	Global Positioning System
HFA	Health Facility Assessment
HIS	Health Information System
HMIS	Health Management Information System
HTML	HyperText Markup Language
IHFAN	International Health Facility Assessment Network
IT	Information Technology
KML	Keyhole Markup Language
MFL	Master Health Facility List
MoH	Ministry of Health
NGO	Non-Governmental Organization
OGC	Open Geospatial Consortium
PMTCT	Preventing Mother-To-Child-Transmission
SAM	Services Availability Mapping
SARA	Service Availability and Readiness Assessment
SDMX	Statistical Data and Metadata Exchange
SPA	Service Provision Assessment
SQL	Structured Query Language
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
UUIDs	Universally Unique Identifiers
VPN	Virtual Private Network
WHO	World Health Organization
xAL	Extensible Address Language
XML	Extensible Markup Language

EXECUTIVE SUMMARY

Sound information on the supply and quality of health services is essential to provide good services. Without information to guide us, we do not know were to go. The efforts fight major diseases and to reach national goals and the Millennium Development Goals (MDGs) have shown us the need for good data. We neeed to track the progress and performance of health systems. Despite investments in health systems, few countries have accurate and up-to-date information on the state of their health facilities, covering the public, private-for-profit and private-not-for-profit sectors. There is an urgent need to develop efficient and sustainable health infrastructure monitoring mechanisms for effective delivery of health care services. Developing and maintaining a comprehensive Master Health Facility List (MFL) is a corner stone in monitoring the health infrastructure and the services provided to the population. Hence, it should be a natural part of the system that provides information to guide decision making.

A Master Facility List is a complete listing of health facilities in a country (both public and private) and is comprised of a set of identification items for each facility (signature domain) and basic information on the service capacity of each facility (service domain). The set of identifiers in the signature domain serves to uniquely identify each facility in order to prevent duplication or omission of facilities from the list. The service domain contains a basic inventory of available services and facility capacity, providing essential information for health systems planning and management. Consolidating health systems information through the MFL will improve record-keeping and reporting efficiency as well as transparency in the health sector. In addition, a MFL is a prerequisite for the sampling of health facilities to conduct more detailed assessments of service delivery such as the Service Availability and Readiness Assessment. Moreover, linking health facility data and other core health system data (financing, human resources, and infrastructure) through the unique identifiers defined in the MFL will allow better analysis and synthesis of information to improve health systems reporting and planning.

This document describes the steps necessary to create and maintain a Master Facility List, as well as the minimum set of indicators that should be included. The document is divided into three main sections: 1. Establish institutional arrangements, 2. Develop an implementation plan, and 3. Technical aspects of establishing a Master Facility List.

INTRODUCTION

Good information on the supply and the quality of health services is essential for managing health systems. Without it, we do not know ere we are going, we cannot monitor our work and evaluate it to be able to work better. The efforts to scale up the response against major diseases and to achieve national goals and the Millennium Development Goals (MDGs) through global cooperation have drawn attention to the need for data which can accurately track the progress and performance of health systems. There is a growing need to evaluate the way in which health system inputs affect health services as well as health outcomes. Despite this need, few countries have up-to-date information on the availability of health services, both in the public and private sector. Fewer still have the data required to assess and monitor the ability of health facilities to provide quality services, or conduct annual health sector reviews. Most countries face challenges in producing data of sufficient quality to consistently track health system changes or progress, and thus strengthen their health systems.[1]

In recent years, progress has been made at both country and global levels to develop methods of collecting and improving access and quality of data on health services availability and readiness.[2] As a result of the diverse tools being used to collect health services information and the lack of use of common unique identifiers, it is difficult to conduct cross-survey comparisons and synthesis of data. In order to produce sound and timely analysis of data, investment in a sustainable national database of health facilities is necessary. A Master Facility List (MFL) creates a standard mechanism for uniquely identifying health facilities, and allows for information to be compared across time and across data sources for individual facilities.

The development of a comprehensive MFL is an initial step towards strengthening performance monitoring at the facility level and feeds into regional, national, and international monitoring systems. A Master Facility List is a complete listing of health facilities in a country (both public and private) and is comprised of a set of identification items for each facility (signature domain) and basic information on the service capacity of each facility (service domain). The set of identifiers in the signature domain[3] serves to uniquely identify each facility to prevent duplication or omission of facilities from the list. The service domain contains a basic inventory of available services and facility capacity, providing essential information for health systems planning and management. A MFL can be used to integrate disparate data sources across time, minimize duplication, increase efficiency, and enable the linkage of surveys and datasets based on facility-level information. The MFL should be comprehensive, up-to-date, and accurate, with appropriate dissemination to all relevant stakeholders.

Developing and maintaining a MFL provides a multitude of benefits including:

- **Data harmonization** - Comparing and contrasting data across different surveys and across time.
- **Data linkages** - Allowing linkages and collaboration between departments and ministries with related data, in order to optimize the use of all databases.
- **Prerequisite for health facility assessments** – Provides a comprehensive list to be used for sampling facilities to be surveyed.
- **Health information strengthening** - Demonstrating efficiencies, trends, gaps, and the ability to generate facility, regional, and national profiles that combine data from multiple systems and provides the information required for effective planning.

---

[1] WHO (2010). *Monitoring and evaluation of health systems strengthening*. Available at: http://www.who.int/healthinfo/HSS_MandE_framework_Oct_2010.pdf

[2] WHO (2011). *Health statistics and health information systems: Service Availability and Readiness Assessment (SARA).* Available at: http://www.who.int/healthinfo/systems/

[3] Health Facility Assessment Technical Working Group (2007). *The Signature Domain and Geographic Coordinates: A Standardized Approach for Uniquely Identifying a Health Facility* (WP-07-91). Available at: https://www.cpc.unc.edu/measure/publications/pdf/wp-07-91.pdf

- **Resource-saving** - Reducing the financial and human costs by eliminating the current duplication of efforts and reducing the reporting burden.
- **Transparency** – Allows for transparent and efficient access to facility data to the Ministry of Health (MoH), partners, and the public.
- **Provide health information for the public –** Provide up-to-date information on location of health facilities and services offered, which makes an efficient mechanism for finding desired services.

What is a MFL?

A Master Facility List is a complete listing of health facilities in a country (both public and private) and is comprised of a set of identification items for each facility (signature domain) and basic information on the service capacity of each facility (service domain). The set of identifiers in the signature domain serves to uniquely identify each facility in order to prevent duplication or omission of facilities from the list. The service domain contains a basic inventory of available services and facility capacity, providing essential information for health systems planning and management. Consolidating health systems information through the MFL will improve record-keeping and reporting efficiency as well as transparency in the health sector. In addition, a MFL is a prerequisite for the sampling of health facilities to conduct more detailed assessments of service delivery such as the Service Availability and Readiness Assessment. Moreover, linking health facility data and other core health system data (financing, human resources, and infrastructure) through the unique identifiers defined in the MFL will allow better analysis and synthesis of information to improve health systems reporting and planning.

Evidence for benefits

Developing and maintaining a MFL provides a multitude of benefits including:

- **Data harmonization** - Comparing and contrasting data across different surveys and across time.
- **Data linkages** - Allowing linkages and collaboration between departments and ministries with related data, in order to optimize the use of all databases.
- **Prerequisite for health facility assessments** – Provides a comprehensive list to be used for sampling facilities to be surveyed.
- **Health information strengthening** - Demonstrating efficiencies, trends, gaps, and the ability to generate facility, regional, and national profiles that combine data from multiple systems and provides the information required for effective planning.
- **Resource-saving** - Reducing the financial and human costs by eliminating the current duplication of efforts and reducing the reporting burden.
- **Transparency** – Allows for transparent and efficient access to facility data to the Ministry of Health (MoH), partners, and the public.
- **Provide health information for the public –** Provide up-to-date information on location of health facilities and services offered, which makes an efficient mechanism for finding desired services.

Kenya have a publically available and continually updated MFL lists[4]. Haiti is one other example of a country that has started a publically available MFL[5].

[4] http://www.ehealth.or.ke/facilities/

[5] https://sites.google.com/a/netspective.org/haiti-health-facilities/

BACKGROUND

The idea of having a complete listing of all health facilities in a country is not new. The term 'master health facility list' (MFL) was developed several years ago to define this process. This term referred to a list of all health facilities in a country with a set of attributes to uniquely identify each facility. In essence, these attributes constituted the *signature domain*. In the following years the list of attributes was expanded to include a *service domain* containing basic information about the facility's services and capacities.

This document describes the steps necessary to create and maintain a master health facility list, including the minimum set of data elements for the MFL. The document is divided into five sections.

1. **Establish institutional arrangements**
   The necessary institutional arrangements need to be in place prior to the actual creation of a MFL. A coordinating group should be established to secure sufficient institutional buy-in and commitment to develop and to maintain the MFL over the longer term. This section describes the roles and responsibilities of the institutional actors that should be involved in the process as well as the components of the implementation plan that should be developed by the coordinating group. This is a very important phase of the process since a wide acceptance of the project plan and thorough preparation are both key factors for success in creating a master health facility list.

2. **Select and define data elements for the MFL**
   The data elements for the MFL should be identified and defined. This section proposes two domains of data elements to be included in the MFL: the signature domain and the service domain. For each minimum required data element, a definition, data rules, and examples are provided.

3. **Identify data sources and populate the MFL**
   In order populate the MFL, all existing data sources that contain information on the data elements of the MFL should be gathered. These data sources can then be used to initially populate the MFL with available data. This section will describe the processes used to identify data sources and populate the MFL.

4. **Fill data gaps and update the MFL**
   Once an initial list of facilities has been created, it is important to begin filling in the information gaps. When the MFL is initially populated with data, it will very likely not be complete in terms of either having all the facilities in a given country, or having all of the required data elements for each facility. Updating the MFL with data will be an on-going process. This section describes potential processes for filling gaps on signature domain information, facility locational information, and service domain information as well as potential processes for maintenance procedures including adding a new facility, archiving non-existent facilities, and changing information for a facility.

5. **Manage and maintain the MFL**
   Once the MFL is complete, on-going management is necessary to maintain the MFL, update the data, and disseminate the MFL as required. This section will cover what to consider when determining how the MFL will be managed, maintained, and disseminated.

The annexes provide technical information for creating a MFL including database selection, database design, linking the MFL to external data sources, and determination of geographic coordinates.

Step 1: Establish institutional arrangements

The development of a MFL is a long-term commitment requiring support from multiple stakeholders. The implementing agency, typically the Ministry of Health (MoH) or its equivalent, should secure the involvement and commitment of the relevant institutions. The MFL is not simply a listing of health facilities; it is a tool which can help to enforce best practices of information sharing and standardization, which can be used across the health sector. As such, the implementing agency should set up a coordinating group to obtain strong commitment on the part of relevant stakeholders. The coordinating group will be responsible for setting the overall policy framework of the MFL, coordinating the input of information by various stakeholders, and developing the implementation plan for the MFL. The following topics will be covered in this section:

- Establish a coordinating group
- Develop an implementation plan

1.1 Establish a coordinating group

Responsibilities for health information go beyond ministries of health, and include other departments, ministries, and agencies that handle health-related data, including central statistics offices, ministries of education, etc. A strong coordinating body is needed to bring together the various stakeholders and help ensure the development of a comprehensive and integrated plan for health information and statistical system development *(1)*.

The first step in the development of a master health facility list is to establish a coordinating group at the national level to oversee and facilitate the planning, implementation, management, and maintenance of the project. The coordinating group will be responsible for:

- Obtaining agreement and support of key stakeholders, ministries, agencies and other partners;
- Defining strategic user requirements, essential domain elements, and data element definitions;
- Developing and carrying out the implementation plan; and
- Disseminating the MFL to ensure wide-spread use across the health sector.

The coordinating group should consist of a few individuals from the implementing agency (generally the Ministry of Health[6]) with strong contacts within the Ministry as well as with key partners and stakeholders, who can shepherd the MFL through its development. Key partners and stakeholders are those that are involved in related initiatives and/or would benefit from using the MFL, and generally include the following:

- Regulatory body for licensing of health facilities in the country;
- National mapping agency and the central statistics office (CSO);
- Other governmental agencies which are involved in the health sector, such as the National HIV/AIDS Control Agency or its equivalent;
- Non-governmental organizations (NGOs) and other organizations involved in data collection;
- Universities and other academic institutions involved in research;
- Health-related UN organizations present in the country (i.e., World Health Organization (WHO), United Nations Children's Fund (UNICEF), United Nations Development Programme (UNDP), Joint United Nations Programme on HIV/AIDS (UNAIDS)); and
- International funders active in the country (i.e., The Global Fund to Fight AIDS, Tuberculosis and Malaria, bilateral government agencies for international development).

[6] In the context of this protocol, the Ministry of Health refers to the public institution or institutions which oversee the public health system.

Members of the coordinating group may be drawn from among the stakeholders and partners, as circumstances warrant. The coordinating group should also include a few individuals with the technical background and knowledge required to oversee the technical implementation of the MFL.

1.2 Develop an implementation plan

The implementation plan is the guiding document for the MFL. It describes the main goals and country-specific requirements for the list. It should also include a brief situation analysis of the available resources. The implementation plan should be drafted by the coordinating group and circulated for comments and approval prior to finalization. The plan should lay out the rationale for the MFL, how the project will be executed, how to oversee the project to ensure that it will be completed on time and within budget, and how it will be maintained.

The implementation plan should:

- Clearly define the objectives and scope of the project;
- Specify the institutions to be involved and their roles and responsibilities;
- Define types of health facilities to be included in the MFL;
- Define what types of information to be included and
- Develop a detailed budget and timeline;

A. Clearly define the objectives and scope of the project

The project definition should provide answers to the following questions:

- What is this project aiming to achieve and why is it important;
- What are the desired endpoints;
- What are the intended uses of the MFL by end-users; and
- Is funding available for the MFL? If so, are the objectives and scope aligned with potential funding sources?

B. Specify the institutions to be involved in implementation of the MFL and their roles and responsibilities

Apart from the Ministry of Health, the institutions to be involved in developing and maintaining the MFL will vary by country and by the scope of the project. Some countries may prefer to keep the development and maintenance of the MFL primarily within the MoH, with strategic contacts with other institutions and stakeholders the MoH interacts with on a regular basis. Other countries may want greater involvement from other institutions and agencies in all parts of the project. It is important to clearly define the roles and responsibilities of the parties involved, particularly to ensure sustainability of the project. The following topics should be addressed:

- Roles and responsibilities of national agencies, departments, institutions as well as those of development and implementing partners;
- Who will manage the various implementation processes; and
- Who will maintain the MFL after it is established and how.

C. Define types of health facilities to be included in the MFL

The coordinating group should determine the types of facilities to be included in the MFL. This includes but is not limited to:

- Establishing inclusion and exclusion criteria for facilities or facility types to be included in the MFL; and
- Identifying the appropriate level of integration of the private sector into the MFL.

To aid in accomplishing these essential tasks, a clear definition of "health facility" for the purpose of the MFL is required. In some countries, this might include all formal facilities which operate within the public

sector. In other countries, the concept of a health facility may extend to mobile clinics, pharmacies, laboratories, specialty clinics, and private and faith-based establishments. The clear definition of what comprises a health facility will ensure that the appropriate stakeholders are represented in the coordinating group, the boundaries of the list are clear, and will allow for administrative practices to be more easily established around this class of facilities.

D. Define types of information to be included in the MFL

The goal is to uniquely identify and provide some basic information on the facility. The decision on what to be included ought to be based on what information ideally should be collected, and what is possible to collect with a reasonable effort.

If certain information is considered to be confidential, it can be made inaccessible in the public version of the MFL. This question may be particularly relevant for private health facilities that may have other reasons not to disclose information than public facilities. The question of who can access what kind of information in the list should be addressed in the planning phase.

E. Budget and timeline

The project budget is an estimate of the costs including staff, equipment, and other related expenses. Based on technical implementation of the MFL, the available human resources, and assessment of existing data, the budget should give a detailed list of line-item expenses for the project. Estimates for on-going maintenance costs, including staffing and other non-financial resources, as well as training of district health officers should be included. The budget should also include detailed information on where the funds to cover these expenses will come from. Once the budget has been created, funding should be secured for all MFL activities. If funding is insufficient, a reassessment of the objectives and scope may be necessary.

In addition to the budget, a timeline should be established which identifies the key deliverables of the project with a timeframe for completion. A clear delineation of responsibilities among the parties should be made to ensure accountability. The timeline will vary widely based on the country situation.

Step 2: Select and define data elements for the MFL

The second step in establishing a master health facility list is to define the data elements that will be included in the MFL. Best practices for master health facility lists specify that two domains should be included: the Signature Domain (2) and the Service Domain. These domains serve to uniquely identify and provide some basic information on the facility. Table 1 below lists the minimum set of data elements which should be collected for each health facility, along with a recommendation of whether it is "Preferred" or "Required". "Preferred" elements will add significant information and functionality to the list, but may be more expensive or difficult to collect. In addition to these minimum data elements, any additional data elements that are desired for inclusion in the MFL should be identified and defined.

Table 1: Minimum data elements

DOMAIN	COMPONENTS	STATUS
Signature Domain	- Unique Identifier	Required
	- Facility name	Required
	- Facility type	Required
	- Ownership/managing authority	Required
	- Location / Address	Required
	- Administrative units	Required
	- Geographic coordinates	Required*
	- Operational status	Required*
	- Data year	Preferred
Service Domain	- Services offered	Preferred
	- Human resources	Preferred
	- Infrastructure	Preferred

* If geographic coordinates and operational status are unavailable due to technical or monetary constraints or due to political or national security reasons, they can be kept private or completed at a later date. However, even partial dissemination of geographic health facility information would be very useful in increasing the accessibility of health services, as well as in determining health service capacities. Furthermore, both operational status and geographic coordinates should be collected and included in the list, even if they are not disseminated. Such information is imperative when updating and reconciling the master health facility list.

In order to be able to collect accurate information, it is necessary to develop clear definitions for each data element in the MFL. For each of the above mentioned minimum data elements, a definition, data rules, and examples are provided below.

2.1 Signature Domain

The signature domain is a set of data elements that can be used to establish a fingerprint for a facility, which should not change significantly over time. Much like a person's signature can ensure his or her identity; the elements of the signature domain would ensure a health facility's identity. This domain contains all the information necessary to identify a facility uniquely and therefore the majority of these data elements are **REQUIRED** in the MFL.

A. Unique Identifier

One of the most important, and absolutely mandatory, components of the MFL is the unique identifier. The unique identifier provides a mechanism to authoritatively and unequivocally identify the facility. This unique identifier is used as primary key, allowing the linkage and analysis of different data sources through a common field.Hence, this should be the same identifier used in the health management information system, surveys of service availability and all other reporting from health facilities. Not only should this identifier be completely unique, it should never change for a given facility, even if other data elements of the facility (such as the address, ownership, etc.) change.

Definition: A unique identifier is a unique code that is used to reference a health facility.

There are many types of unique identifiers that can be used, each suitable for different situations. Three options for unique identifiers are presented below: integer codes, facility codes, and universally unique identifiers. Regardless of the unique identifier chosen, manual generation of codes should be avoided as this is prone to error and duplication of codes.

Integer codes

Sequential integer codes are often used for facility identifiers, particularly when the code is generated by a central authority. They are simple, compact, and can be stored in essentially any system. Integer codes are suitable for centralized architectures where the unique identifier is generated by a central authority. Usually, the integer should be generated automatically with no need for human intervention, and therefore be guaranteed to be unique in a centralized MFL. If the code do not contain any information, it does not need to be changed when e.g. administrative boundaries change. Information free numbers, also can be shorter than codes carrying information. A four digit number may cater for up to 10 000 facilities.

Integer codes may be easier to use in decentralized architectures when ranges are established for each distribution node. For instance, District X may be assigned integer codes 1-100 and District Y could be assigned codes 101-200. This rule will ensure that uniqueness is maintained between each node, but is volatile to changes in the node structure.

Facility codes

Facility codes refer to a type of coding system that has been developed expressly for the purpose of uniquely identifying facilities. Typically, these consist of some type of categorization system, consisting of combinations of letters and numbers, and are often meant to encode some type of information in the identifier itself.

Some coding systems use “human friendly” codes that contain some information in the code itself. For instance, the code for "Tambaka" district in Sierra Leone is SL001001000000000. In this coding scheme, the first two letters provide the country code, the second three numbers provide the number of the first administrative district, and the next set of digits provides the district number. The main weakness of this coding system is that some types of information embedded in the health facility code itself may not be time invariant. For instance, if a country introduces changes to their districts, the code may not correspond to the new district to which the facility belongs. In addition, maintenance can be difficult and time consuming as a result of these issues.

If facility codes are going to be used as unique identifiers, a robust algorithm should be developed and implemented, which will be capable of handling these types of situations and potentially others that may exist uniquely in a country. Facility codes may be suitable for decentralized architectures, assuming that there are robust mechanisms established to guarantee that the same code cannot be generated by two separate authorities. If not generated automatically, facility codes are prone to errors.

Universally Unique Identifiers

Universally unique identifiers (UUIDs) refer to an information identifier standard that is used widely in computer applications. The intention of UUIDs is to provide a mechanism to uniquely identify information without the need for any sort of central coordination. UUIDs are essentially guaranteed to always be unique, no matter where or by whom they are generated. In that sense, they serve a very useful purpose in decentralized systems, where communication with a central authority is not possible each time a new identifier is required.

UUIDs are a 16-bit number, which can be generated by any number of programs and database systems according to a standardized algorithm. An example of a UUID in its standard form is 40e74fae-c0ab-11df-b090-0017f2300bf5. One of the drawbacks with the use of UUIDs in a MFL is that they are not amenable to categorization or memorization.

Table 2 below, provides an overview of the strengths and weaknesses of each of the unique identifier types presented.

Table 2: Strengths and Weaknesses of Unique Identifiers

Unique Identifier	Strengths	Weaknesses
Integer codes	- Suitable for centralized architectures where the unique identifier is generated by a central authority - Simple, compact, and can be stored in any system - Guaranteed to be unique in a centralized MFL (generated automatically)	- Not suitable for decentralized architectures, unless ranges are established for each distribution node - Not amenable to categorization or memorization
Facility codes	- Identifier is a categorization system and provides information in the identifier itself	- Some types of information embedded in the health facility code may not be time invariant - Maintenance can be difficult and time consuming - Prone to errors if not generated automatically
Universally Unique Identifiers	- Guaranteed to always be unique (generated automatically) - Particularly suited to decentralized systems, as the identifiers can be generated without the need to contact a central authority for a new code	- Not amenable to categorization or memorization

Data Rules: The unique identifier should in all cases be guaranteed to be unique and time invariant. Several examples are provided below. It is imperative that the unique identifier's structure be the same for all facilities within a country. For example, if facility codes are used, they should be used for all health facilities in the MFL.

Examples:

Integer code: 156

Facility code: SL001002056000

UUID: 40e74fae-c0ab-11df-b090-0017f2300bf5

B. Facility Name

Definition: The facility name is the official name of the health facility.

Ideally, this would be the legal name of the facility issued by the regulatory authority. Additional fields for alternative names and names in local languages could also be added.

Data Rules: The facility name should consist of a single text field to describe the name of the health facility. It is recommended that the name be free of any abbreviations. A facility may have an alternate or several alternate names, especially in local contexts. However, the official name should match the name used for legal purposes, such as licensing of the health facility. It is also recommended to use proper capitalization of the name of the facility, for example, avoiding the use of all capital letters. Excessive abbreviation and the use of backslashes, forward-slashes, apostrophes, and other symbols, may also complicate the storage and retrieval of the name. Where possible, punctuation symbols should also be avoided.

Generally the type of the facility is part of the name itself, such as "Lusaka General Hospital" or "Downtown Urban Health Clinic". Careful attention should also be paid to standardizing and restricting the facility type which is appended to the facility name. A separate field will be used to record the facility type (described below). However, if the type of the facility is appended to the proper name of the facility, care should be taken to standardize these names as much as possible to the actual facility type.

Example: Pearl of Health Hospital

Pearl of Health Hosp. <- would not be acceptable, as it uses an abbreviation for "Hospital".

Pearl of Health<- would not be acceptable as it does not indicate the full facility name

C. Facility type

Definition: The facility type refers to the classification of the facility.

Due to the wide variation in terminology and types from country to country, providing a standardized list is not feasible here. Within a single country however, a standardized list of facility types, each with clear definitions based on characteristics such as the minimum set of services, the number of health workers, and the catchment population size should be developed. The MoH may already have such a list of standard definitions for facility types.

Data Rules: Facility types should be determined by a central authority. Even in a decentralized system, delegates should not be allowed to add new types, but select from a list of existing types. Each type should be clearly defined. Each facility should be assigned a single facility type only.

Example: Some examples include "hospital", "primary health care facility", "mobile health care facility" and "specialist hospital". As mentioned above, the exact facility types will vary from country to country.

D. Ownership/managing authority

Definition: Facility ownership refers to type of ownership of a given facility.

As was the case with health facility type, there is no fully standardized list of types of ownerships. Each country would need to define the list of ownership types. The MoH may already have such a list of standard definitions for ownership types.

Data Rules: Ownership/managing authority types should be determined by a central authority. Even in a decentralized system, delegates should not be allowed to add new types, but select from a list of existing types. Each facility should have only one type of ownership. If a facility is owned by several organizations,

the more specific type should be selected. For instance, a "Military" facility could be considered to be owned by both the "Government" and the "Military", but since "Military" is more specific, this option should be selected.

Example: Examples include "Government", "Private", Non-governmental organization", "Military".

E. Location/Address

Definition: The location/address of the health facility should give reference to the physical location of the facility including at least the neighbourhood and the city.

Other address types such as a mailing address (post office box) should not be used, as they do not physically identify the location of the facility.

Data Rules: Given the large variability between countries in regard to addresses, this section of the signature domain will need to be adapted at the country level. At the very least, an indication of the location at the lowest administrative level possible is required (e.g. neighbourhood, town). Ideally, the following specific fields can be defined:

- Street name
- Street number
- City/Town/Neighbourhood
- State/Province
- Postal code

Each separate address attribute should have its own data rules. For instance, the street name should be restricted to a list of standardized street names if they exist for a given location, or at the very least, should be restricted to standardized abbreviations for street types, such as "Ave" for avenue, "St" for street and so forth. Postal codes should be restricted to the format for each country. For instance, for Switzerland, each postal code has 4 integers, followed by a hyphen, followed by the letters "CH".

Example:

Street name: Ave Appia

Street number: 20

City: Geneva

Postal code: 1211-CH

F. Administrative units

Definition: Two core attributes are suggested to describe the administrative area: name and level. The administrative area refers to the district, province, or other administrative level that a facility is in. In order to assure that linkages with other data sources are possible, a standardized listing of administrative units should be used. This listing is generally available from the CSO or a similar agency. If there are alternative administrative units, consider to include all that can be usefull for users iof the MFL. If the geographic coordinates of the facility are available along with sufficiently detailed and accurate administrative boundaries, the administrative unit can be derived from these coordinates.

The MoH may maintain health districts or zones, administrative areas that are specific to the function of the health sector and distinct from the geographic units used in other aspects of a country's governance. Use of administrative units set by agencies such as CSOs promotes standardization across government agencies, which is a key function of the MFL. However, the health district in which a facility is located could

also be added. If a country includes health districts, the rules that govern them as a data element are the same as for administrative units.

- Name: The official name of the administrative unit. If the administrative unit has a number, this should be included as well.
- Level: Refers to the level of an organizational unit within a given hierarchy. For instance, national level might be represented by level “0”, province by level “1”, district by level “2”, ward by level "3", etc.

Data Rules: The facility should be assigned the most specific administration level possible so that the facility can be easily distinguished from other facilities which might have the same name. It is possible that a facility named "Pearl of Health Hospital" would have two branch locations, one located in "Chelstone Ward" and the other in "Central Business District." Correct identification of both the name and level of the administrative unit would help clarify that these are in fact two different facilities.

However, it is important not to include the administrative unit's name or level in the name of the health facility. Indeed, one may be tempted to write "Pearl of Health Hospital - Chelstone Ward" in the section reserved for the facility name. However, this is not the exact name of the facility. Such naming could cause problems in the future, such as when checking for duplicate facilities or linking to external databases. It is therefore imperative to only include the exact name of the facility in the 'Name' section, and be precise with the identification of the administrative units.

Example:

In Sierra Leone, the national administrative level has been defined as level 0, the provincial level has been defined as level 1, and the district level has been defined as level 2.

For a health facility called Fatibra Community Health Clinic located in the district Bo and the province Southern, the following attributes would be recorded:

ID	NAME	Admin 0	Admin 1	Admin 2
123456789	Fatibra Community Health Clinic	Sierra Leone	Southern	Bo

G. Geographic coordinates

Definition: The geographic coordinates refer to the physical location of the facility, typically represented as the latitude and longitude.

Data Rules: The MFL coordinating group should connect with the national mapping agency to determine the existence of any preferred national coordinate reference system. If an established coordinate reference system is not in place in a country, the geographic coordinates should be represented using latitude and longitude in decimal degrees referencing the WGS84 coordinate system. Further restrictions, such as the maximum and minimum latitudes and longitudes of a given country can be placed on these fields, ensuring that incorrect values are not stored by mistake.

Both the longitude and latitude should be specified in decimal degrees (with positive and negative numbers). For latitude, north is considered positive and south is considered negative. For longitude, east is considered positive and west is considered negative.

Example:

The longitude and latitude in decimal degrees of Lusaka, Zambia are:

Latitude: -15.41667

Longitude: 28.28333

For more detailed explanation of best practices for collecting geographic coordinates please refer to *Annex 2*.

H. Operational status

Definition: The status of the facility will indicate its operational status, such as "Operational", "Closed", or "Under construction".

The exact list of operational statuses will need to be defined for each country. A suggested list is presented below.

- **Operational:** A facility that is open and serving patients.
- **Licensed:** A facility that has been approved and issued a license by the appropriate national regulatory body, but facility is not yet operational.
- **Pending Licensing:** A facility that has been recommended by the district health management team, but is waiting for a license from the national regulatory body.
- **License Suspended:** A facility whose license has been temporarily stopped for reasons including self-request, sickness, disciplinary action, etc.
- **License Cancelled:** A facility whose license has been permanently stopped by the national body.
- **Pending Registration**: A facility that has been approved by the local authorities as an institution and a request for official registration has been submitted and with approval pending.
- **Registered:** A facility that has been approved as an institution and a registration number given.
- **Closed:** A facility that has a valid license, but which has permanently closed.
- **Invalid:** A facility where the attributes of a facility (name, location, etc.) are different than those on the facility's license.
- **Does not exist:** A facility which has been licensed, but has been verified not to physically exist.
- **Duplicate:** The facility exists and is properly licensed, but is an effective duplicate of another facility. This usually occurs when two data sources are merged together, with slightly different names but refer to the same facility.

Data Rules: A facility will have a single operational status at any given time.

Example: "Closed".

I. Data Year

Definition: The data year refers to the year in which the data was collected. This information will be particularly important when merging several lists, or when updating information. In such a case, it will be important to highlight the most recent data.

Data rules: A data year will be specified for each health facility entry. In the case of duplicate entries the latest year will be considered valid. If no data year is available, this field should be left blank.

Example: "2007"

2.2 Service Domain

The service domain includes some basic information about the facility, with indicators that are recommended for inclusion in the MFL but are not mandatory including:

- Services offered
- Human resources
- Infrastructure

A. Services offered

Definition: This section provides information on the core basic services being offered in a facility.

The information should be adapted at country level to the package of services offered throughout the health system. These include, but are not limited to: family planning, antenatal care, delivery services, advanced delivery services, child health immunization, child health preventative and curative care, adolescent health services, HIV counselling and testing, HIV/AIDS care and support services, antiretroviral treatment (ARV) therapy, preventing mother-to-child transmission of HIV (PMTCT), tuberculosis diagnosis/treatment, malaria diagnosis/treatment, chronic disease treatment/management, basic surgery, comprehensive surgery, and blood transfusion services. The core indicators and the specific definitions for each service should be agreed upon by the coordinating group, and used to determine service provision.

In general, the number of services should be kept to a minimum. These attributes should be simple Boolean (yes/no) attributes to indicate whether a specific service is offered or not and should not include detailed information on the service. That type of detailed information should be recorded in the context of a service readiness database, such as SARA, SAM, SPA, or HFA, which may be linked to the MFL through the attributes contained in the signature domain.

Data Rules: Data should be limited to Boolean (yes/no) types. Each attribute should be clearly defined in a separate metadata definition.

Example: Availability of antenatal services: Yes/No

B. Human Resources

Definition: This section should provide information on the number of core medical personnel, that is to say physicians, non-physician clinicians, registered nurses, and registered midwives.

The exact types of human resource that will be detailed, as well as the definitions for each human resource, is specific to the country. This information will be used to determine the availability of human resources for health in a given area.

Data Rules: The data should be limited to positive values. Each cadre should be clearly defined in a separate metadata definition.

Example: Number of midwives: 5

C. Infrastructure

Definition: This section should include inpatient and maternity beds present in the facility.

For the purpose of the MFL, it is suggested that only inpatient and exclusive maternity beds be specified. Other equipment and infrastructure details should be collected through a separate health facility assessment (SAM, SARA, SPA, or HFA), which would collect more detail about the exact types of equipment and infrastructure available in each facility. Extra equipment and infrastructure may be added to the MFL after approval of the definitions by the coordinating group.

Data Rules: Attribute responses should be limited to either Boolean (yes/no) or positive integer values.

Example: Number of inpatient beds: 10

Step 3: Identify data sources and populate the MFL

In order to populate the MFL, all existing data sources that contain information on the data elements of the MFL should be gathered. These data sources can then be used to initially populate the MFL with available data. This section will describe the processes used to identify data sources and populate the MFL.

3.1 Identify existing data sources

The MoH usually maintains information on public health facilities and can serve as a starting point for identifying existing data sources for the MFL. The routine health facility reporting system (HMIS) may include information on public facilities only, or public and private facilities, depending on the country. Other MoH data sources on health facilities may exist under different departments and should be thoroughly investigated. For example, disease-specific programs may have data on facilities offering diagnosis and treatment of a specific disease. Other contacts should be identified to obtain information on private, Faith Based Organization (FBO), NGO, and other facilities. Additionally, other government agencies (e.g. CSO, business registration office, professional medical associations) may have pertinent health facility information that should be obtained (e.g. CSO may maintain health facility mapping data). In many countries, a separate regulatory body is responsible for the issuance of licenses for health facilities. Such organizations, where they exist, should be able to provide a list of health facilities. It is expected that there will be significant overlap of health facility information between the institutions and agencies.

In addition to these primary sources of health facility information, previous health facility assessment surveys (Service Availability and Readiness Assessment (SARA) *(3)*, Service Provision Assessment (SPA) *(4)*, Service Availability Mapping (SAM), or Health Facility Assessment (HFA)) can also provide relevant information. Obtaining a list of health facility assessment surveys that have been implemented and their associated data sets will also help inform the baseline; however, it is also important to consider how recent the survey is and whether the data are still valid for this purpose.

Checklist of potential data sources:

- ☑ Ministry of Health
- ☑ Health Management Information System (HMIS)
- ☑ Other organizations representing private/FBO and NGO communities
- ☑ Government agencies other than the MoH (e.g. CSO)
- ☑ Health regulatory bodies
- ☑ Health facility assessment survey (SARA, SPA, SAM, HFA), etc.

3.2 Populate the MFL with existing data

The MFL can be created in a variety of different formats. The coordinating group should determine the most appropriate format for the MFL. Several format options are listed below:

- Spread sheets
- Relational database

- Other documents, such as a facility listing which is not organized as a database

The skills existing in the organisation is important for the choise of tool to use. More information on technical information for creating a MFL including database selection, database design, and linking the MFL to external data sources can be found in Annex 1.

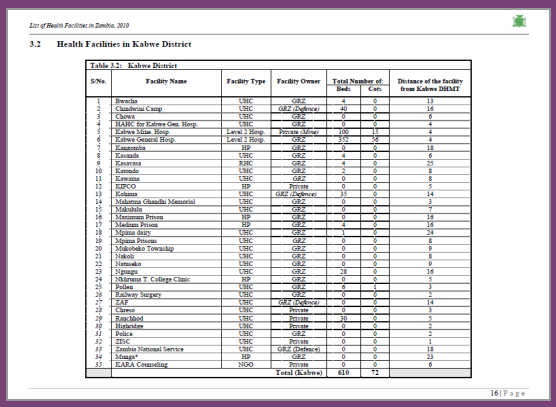
The following steps should be followed when initially populating the MFL with existing data:

1. Identify most authoritative data source and import this information to the MFL
2. Add information from any additional available data sources
3. Match facilities to create one record per facility in the MFL
4. Validate data

1. Identify most authoritative data source and import this information to the MFL

In almost all countries, some type of facility list exists, oftentimes in the Ministry of Health. Sometimes, the only information available will be the facility name, facility type, facility managing authority, and the approximate location of the facility. This list should be used to initially populate the MFL and will serve to create a simple and limited facility list which can be revised over time to achieve completion. Each facility should be given a single record in the MFL.

Zambia example:
In Zambia, an authoritative facility list exists, but does not contain all of the data elements necessary for a complete MFL. The *List of Health Facilities in Zambia, 2010*, contains information on facility name, facility type, facility managing authority, and total number of beds for public and private facilities according to district and region in Zambia.



List of Health Facilities in Zambia, 2010

3.2 **Health Facilities in Kabwe District**

Table 3.2: Kabwe District

S/No.	Facility Name	Facility Type	Facility Owner	Total Number of:		Distance of the facility from Kabwe DHMT
				Beds	Cots	
1	Bwacha	UHC	GRZ	4	0	13
2	Chaibwisa Camp	UHC	GRZ (Defence)	40	0	16
3	Chowa	UHC	GRZ	0	0	6
4	HAHC for Kabwe Gen. Hosp.	UHC	GRZ	0	0	4
5	Kabwe Mine. Hosp.	Level 2 Hosp.	Private (Mine)	100	15	4
6	Kabwe General Hosp.	Level 2 Hosp.	GRZ	252	58	4
7	Kamfinsha	HP	GRZ	0	0	18
8	Kasanda	UHC	GRZ	4	0	6
9	Kawama	RHC	GRZ	4	0	25
10	Katondo	UHC	GRZ	2	0	8
11	Kawama	UHC	GRZ	0	0	8
12	KIPCO	HP	Private	0	0	5
13	Kohima	UHC	GRZ (Defence)	35	0	14
14	Mahatma Ghandhi Memorial	UHC	GRZ	0	0	3
15	Makululu	UHC	GRZ	0	0	7
16	Maximum Prison	HP	GRZ	0	0	16
17	Medium Prison	HP	GRZ	4	0	16
18	Mpima dairy	UHC	GRZ	1	0	24
19	Mpima Prison	UHC	GRZ	0	0	8
20	Mukobeko Township	UHC	GRZ	0	0	9
21	Nakoli	UHC	GRZ	0	0	8
22	Natwasko	UHC	GRZ	0	0	9
23	Ngungu	UHC	GRZ	28	0	16
24	Nkrumu T. College Clinic	HP	GRZ	0	0	5
25	Pollen	UHC	GRZ	6	1	3
26	Railway Surgery	UHC	GRZ	0	0	2
27	ZAF	UHC	GRZ (Defence)	0	0	14
28	Chowo	UHC	Private	0	0	3
29	Ranchhod	UHC	Private	30	0	5
30	Hirlerides	UHC	Private	0	0	2
31	Police	UHC	GRZ	0	0	2
32	ZESC	UHC	Private	0	0	1
33	Zambia National Service	UHC	GRZ (Defence)	0	0	18
34	Munga*	HP	GRZ	0	0	23
35	EARA Counseling	NGO	Private	0	0	6
			Total (Kabwe)	610	72	

16 | Page

This authoritative list would be used to populate the initial MFL.

2. Add information from any additional available data sources

In many cases, multiple data sources exist, each containing different types of information on health facilities in a country. This information must all be added to the MFL. In order to add the information from other data sources to the MFL, create a new record in the MFL for each health facility listed in the additional data sources.

3. Match facilities to create one record per facility

Now that all of the existing data has been entered into the MFL, it must be reconciled to produce a single record per health facility. When multiple data sources are combined, differences arise due to the intrinsic differences in the data sources. Some of the challenges include:

- Facilities don't have unique codes, thus it may be difficult to match information from disparate data sources;
- Discordance between data sources often arises due to differences in the sampling methodology and time periods in which the data was collected; and
- Data collected several years in the past could be inaccurate, as facilities are constantly being opened and closed.

The following steps can be taken to reconcile the MFL:

1. Organize facilities by lowest appropriate administrative unit, often by district.
2. Run an electronic match per district based on facility name. A portion of facilities will be matched through this process. For each match, combine the information from the matched records so that the facility has only one record.
3. For unmatched facilities, matching will have to be a manual process. The following data elements should be the same in both records for the records to be considered a match.
    - Facility name. If there is a slightly different spelling of the facility name, abbreviations are used, or differing use of capitalization, an electronic match process will not recognize these facilities as a match.
    - Facility type
    - Managing authority
    - GPS coordinates- GPS coordinates should be mapped to determine if they are a match. Most likely, two sets of GPS coordinates for a facility will differ slightly due to the fact that the coordinates will most likely not be taken in the exact same place on multiple occasions. Mapping the coordinates will enable you to determine if the coordinates are very close and thus most likely for the same facility.
4. If some but not all of the above criteria are met, leave both records in the MFL and this can be reconciled during the validation process. If all of the criteria are met, treat the records as a match and combine the information from the matched records so that the facility has only one record.

D. Assign unique identifiers

At this point a unique identifier should be assigned to each facility record in the MFL. The type of unique identifier used should have been determined when defining the data elements for the MFL (section 2.1). Regardless of the type of unique identifier used, automatic generation of the unique identifier is preferred as it helps to ensure each facility receives a truly unique identifier.

E. Validate data

The MFL is now ready to be validated. The accuracy of the signature domain component of the MFL is of critical importance. It is recommended that where possible, people familiar with the health facilities in their own localities, such as district health information officers, be responsible for the initial verification of the data. The MFL should be sorted by district and then sent to district health information officers (or others with equivalent knowledge) for verification and further reconciliation. Verification of each data element (name, ownership, type, etc.) should be carried out through either normal supervisory visits or by dedicated visits specifically to determine the accuracy of the information provided by the MFL. Telephone or email contact, where possible, with the managing authority of the facility may also be employed.

Due to the special equipment required to validate the geographic coordinates of the facility, it may only be feasible to verify this information virtually through satellite images available in services such as Google

Earth or Google maps, depending on the available resources. These approaches are described in more detail in *Annex 2*. If Global Positioning System (GPS) receivers are available, verification of these data could be conducted through normal supervisory visits to facilities on an ad-hoc basis.

Step 4: Fill data gaps and update the MFL

Once an initial list of facilities has been created, it is important to begin filling in the information gaps. When the MFL is initially populated with data, it will very likely not be complete in terms of either having all the facilities in a given country, or having all of the required data elements for each facility. Updating the MFL with data will be an on-going process. This section describes potential processes for filling gaps on signature domain information, facility locational information, and service domain information as well as potential processes for maintenance procedures including adding a new facility, archiving non-existent facilities, and changing information for a facility.

4.1 Updating and filling gaps on signature domain information

There are several possible mechanisms that can be used to obtain information for the signature domain data elements of the MFL:

1. In countries where an appropriate HMIS system is in place, information on each facility can be entered electronically at the district level, and then transmitted through the normal HMIS data flow procedure to be compiled into the MFL. However, it is important to note that this may not capture information on private facilities, which may not be included in the routine reporting system.
2. Primary data collection in the form of a health facility census whereby facilities on the initial MFL are visited plus canvassing for additional facilities that are not on the list. A health facility assessment survey, such as the SARA, SPA, or HFA can be administered as a census of all facilities to gather the information necessary for the MFL. The International Health Facility Assessment Network (IHFAN) provides information on many available HFA tools and methodologies available for country use *(5)*.
3. A simplified questionnaire, specifically designed to complete the MFL, can be sent to each responsible district health information officer. Each district health information officer would be asked to fill in the information for all of the facilities in their area, both public and private.
4. Review of other governmental databases, such as a Human Resource database, that may contain relevant information.

The approach taken should be determined by the existing resources and health information systems in a country, and if feasible a combination of all available methods should be applied.

4.2 Updating and filling gaps on facility locational information

Collection of latitude and longitude coordinates for all facilities can be challenging. There are several approaches that can be used to obtain the geographical position of a health facility including:

- Direct survey of the facility with a GPS receiver or mobile phone with GPS capacity
- Calculation from proximity to a geo-referenced school, village, or market
- Calculation from 1:50,000 scale topographic maps
- Calculation from scanned, hand drawn maps
- Calculation from the centre of the lowest administrative level
- Determined from satellite imagery (e.g. using Google Maps or Google Earth)
- Searching in another geo-referenced database, perhaps compiled by another ministry or a NGO (its accuracy would then have to be verified)

The preferred method would be to obtain latitude and longitude from a direct GPS survey of the facilities or authoritative confirmation of a facility using imagery using a resource such as Google Earth or Google Maps, but when this strategy is not possible or resources will not allow for it, any of the other above methods can be used to fill in the data gaps. The goal, over time, is to replace any estimated geographic coordinates with directly measured GPS latitude and longitude as the information becomes available. *Annex 2* provides detailed technical information on how to collect geographic coordinates through the various mechanisms listed above.

4.3 Updating and filling gaps on service domain information

Service domain information will need to be updated on a regular basis as new facilities are added to the MFL, or as information is determined to be out of date or non-existent. There are two primary mechanisms to update this data: a primary survey exercise where each facility is visited and a questionnaire is administered, or regular updates through the existing HMIS. Primary surveys have the advantage of involving a trained surveyor who can also verify the information provided by the facility. These surveys, however, as standalone exercises, can be costly both in terms of financial and human resources. One possibility is to combine such a survey with other initiatives, e.g. with supervisory visits from the district office, or with a service availability and readiness assessment.

The use of existing data channels, such as the HMIS, can be applied in situations where such a system exists, and this system has been linked to the MFL. Service domain information can be considered to be a yearly dataset, which can be reported on through existing reporting systems. However, for private facilities and other facilities that may not be part of the main health information system, a primary survey is likely the only feasible mechanism to update and fill data gaps on service domain information.

Another mechanism to fill service domain information would be through ad-hoc supervisory visits. Typically, district health information officers make routine visits to facilities as a course of their normal duties. During the course of these supervisory visits, they could validate the information contained in the MFL, and add new information to fill data gaps. This updated information could then be added to the MFL directly by the information officer or channelled to the responsible authority as appropriate.

4.4 Add new facilities

The specific process used to add new facilities to the MFL will vary by country. When developing a process to support the addition of a new facility, several factors should be taken into consideration and are detailed below. The overriding principles when adding a new facility are to ensure that duplicates are not entered, to ensure that each facility receives a unique identifier, and to ensure that all facilities which are added have accurate and up-to-date information.

- ***How will notification be received that a new facility has been opened? Is there a regulatory body responsible for issuance of licenses that can inform the MFL managers of new facilities?***

  Ideally, notification of new facility openings would occur in cooperation with regulatory authorities that are responsible for the issuance of licenses or other official documents when a facility is opened. However, in some instances, this type of regulation does not exist. This is most often the case for facilities that do not fall within the public sector, such as private, NGO, and FBO facilities that operate outside the direct control of the national Ministry of Health. In these situations, other methods such as a periodic primary facility census may be necessary to capture these newly opened facilities.

- ***Who is responsible for adding a new facility to the MFL? Is this centrally administered or de-centralized?***

  In a centrally administered system, the MFL manager would be responsible for adding a new facility to the MFL. This addition would typically occur after the facility has received regulatory clearance to

operate or, in some countries, it may also be appropriate to add facilities that are under construction and are not yet formally open. If a regulatory body is responsible for issuing licenses that allow a facility to operate, mechanisms should be established for the regulatory body to send information on new facilities to the MFL manager.

In a de-centralized system, typically the district health office or its equivalent would be responsible for adding a new facility to the MFL once the facility has obtained regulatory body clearance and is operational. Internal administrative mechanisms would need to be established in order to ensure that the person(s) in charge of maintaining the MFL for a given de-centralized region would be informed when a new facility is established.

- ***How to cope with mobile facilities in the database?***

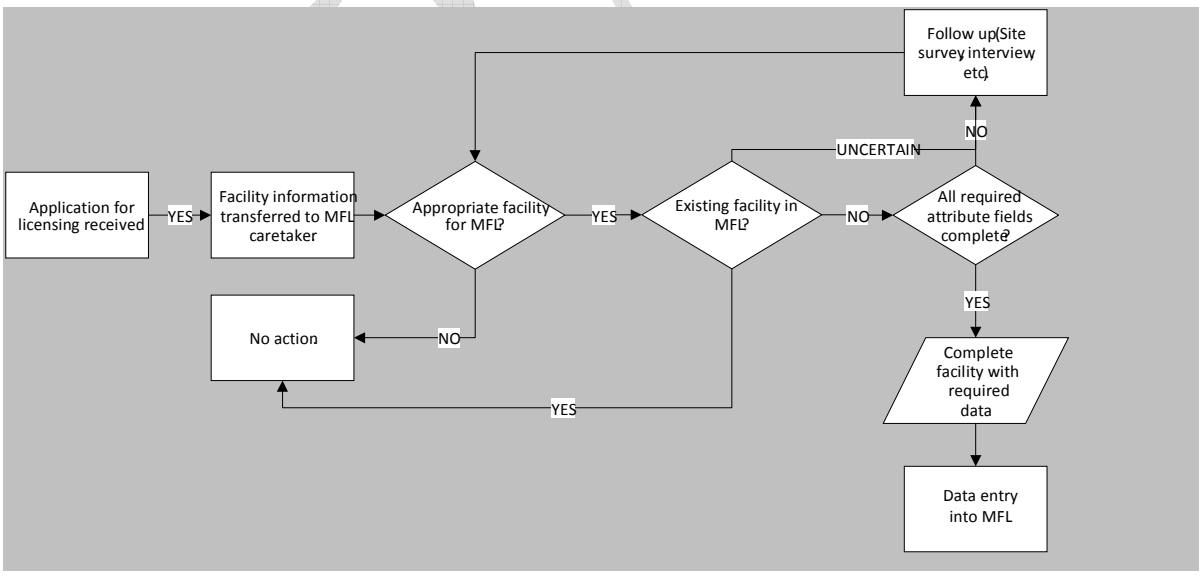
  Mobile facilities are quite common in some countries. Such a facility will visit several locations periodically. It is recommended to add all locations where the mobile facility provides services as a unique record with a unique id in the MFL. It should also be distinguished by different geo-identifiers such as address, town name, and/or coordinates. Extra information about the frequency and opening hours of the mobile facility should be added to the MFL for every unique location.

- ***Who will enter and verify the data?***

  Regardless of how the MFL is administered, data entry should only take place once the facility has been verified to actually exist. Subsequent verification of the newly entered data would occur through normal supervisory visits or through email/telephone verification where possible.

Figure 1 below presents an example workflow which defines when a new facility would be added to the MFL.

Figure 1: Example process for adding a new facility to the MFL



In this scenario, a license application would be received by the regulatory body responsible for issuing health facility licenses. Once the application is received, it would be transmitted to the person responsible for the updating the MFL, who would decide whether the facility is appropriate for inclusion in the MFL. If the facility does not already exist in the MFL and all of the required data elements are complete, the record would be entered into the MFL. If the information on the required data elements is not complete, subsequent follow up in the form of a facility visit or remote interview to determine the missing data

elements would need to be conducted. Once all the information is complete, the new record could be entered into the MFL.

4.5 Archive non-existent facilities

Facilities that are determined not to exist should not be permanently deleted from the MFL but should have their operational status declared as “Invalid” or “Does not exist”. It is recommended to have a separate field to record how the facility was declared as being non-existent, such as direct verification. Note that facilities which did exist, but are closed and are no longer operational, should not be declared as being non-existent, but should rather have their operational status declared as “Closed”. The non-existent operational status should be limited to facilities that can be conclusively verified as having never existed.

4.6 Change information for a facility.

The information in the signature domain and the service domain will vary to some extent over time (e.g. a facility may change its name) and will require updating. The appropriate changes should be made in the MFL and a tracking history for changes made to the MFL, either on paper or electronically, should be maintained to ensure that information is not permanently lost.

Step 5: Manage, maintain, and disseminate the MFL

Once the MFL is complete, on-going management is necessary to maintain the MFL, update the data, and disseminate the MFL as required. This section will cover what to consider when determining how the MFL will be managed, maintained, and disseminated.

5.1 Manage and maintain the MFL

Once the MFL is complete, on-going management is necessary to maintain the MFL, update the data, and disseminate the MFL as required. A set of procedures should be developed outlining how the implementing agency (usually the MoH) will handle the on-going maintenance of the MFL. Several questions will need to be answered, including:

- How often will the data set be updated, and what are the mechanisms for updating the content, e.g. census of facilities every 5 years?
- Who will be responsible for the technical maintenance and on-going development of the system?
- Who will handle technical inquiries related to the list, e.g. what format does the list exist in, how can I integrate my own database with data from the MFL, does the MFL offer a web-service?
- Who will handle policy questions related to the dataset, e.g. request for access to the list and requests to contribute to the list?
- Which unit within the implementing agency carries ultimate responsibility for maintenance of the list?

The fundamental principles of the master health facility list are that:

- Information that is not likely to change should be should be updated as part of the administrative procedures surrounding opening and closure of health facilities. Othe types of information should be gathered at a minimum every 2 years, and if possible, annually.
- Information should be available to everyone who has permission to access the data.
- The system should make working and reporting easier for everyone, and not add to the reporting burden.

These principles should be kept in mind when determining how the MFL will be managed and maintained.

In addition to the maintenance of the facility list itself, reviews of the data element definitions should be conducted on a regular basis. In many countries, the classifications of health facilities may change from time to time, and will need to be updated in the guidelines which define the data elements in the MFL. It is also important to ensure the data element definitions remain consistent over time. For example, if the definition for facility type “Public Health Centre” already exists, a new facility type “PHC” (abbreviation) or “Public Health Centre” (alternative spelling) should not be added to the data element definitions.

5.2 Disseminate the MFL

Providing access to the information to the widest possible audience should be viewed as a strategic endpoint for the MFL. The effectiveness of the MFL will be determined in large part by whether it is adopted as a national standard, not only by the implementing agency, but by as many actors as possible, such as NGOs, international organizations and other bodies which are involved in implementation and monitoring of a countries health sector. In order to provide this information in a timely and accessible manner, the dissemination strategy must be implemented in an informatics platform. Typically, this platform would consist of multiple formats, which would ensure the broadest possible use of the MFL. In order to implement the MFL dissemination strategy, a team of technical experts may be required.

The MFL is of limited use if only one unit in the Ministry of Health has access to it. The data dissemination strategy should address the ways in which the MFL will be promoted and made accessible to other departments, ministries, institutions, agencies, and partners. Different types of users of the MFL should have different types of access to the information, with a corresponding level of security and functionality. In designing the dissemination strategy several questions will need to be answered including:

- Who should have access to the MFL?
- Which parts of the MFL should be available to whom?
- What are the appropriate mediums for MFL dissemination (e.g. Excel, XML, HTML, web services, etc.)?

The following are some examples of user groups and their level of access, which should be adapted to the country context and needs:

Public access

Public access to the MFL is important to ensure that many actors have access to the information for their own use. In addition to the general public, academic institutions, survey agencies, and other partners may want to access the content of the list for their own purposes, without necessarily contributing to the content. If there is sensitive information contained in the MFL, such as personally identifiable information, a subset of the data can be released publically, while all sensitive information can be withheld for privacy or security reasons.

Trusted access

Trusted access should be granted to individuals who are authorized to make changes to the content of the MFL, such as district health information officers responsible for the maintenance of the list of facilities in their district. If a database system is used, authorized users would connect to the database though an extra layer of security such as a Virtual Private Network (VPN) connection or through a password-protected web-based interface, and enter information on new facilities or alter information on existing facilities. Personally identifiable information, such as a facility manager's name and phone number, would be available to this group.

Administrative access

A second layer of security would need to be implemented for the small set of individuals who would have complete access to the database. As super-users, they would be allowed to delete or alter facility information, override information which has been entered by others, as well as alter metadata which typically should not be altered by normal trusted users.

Deciding how the MFL should be disseminated depends on the overall strategic goals. Basic information on health facilities should be widely available, while more detailed information can be provided to key stakeholders of the MFL. The data in the MFL may need to be produced in several different formats to cater to different stakeholders with different data requirements. For instance, a PDF booklet could be produced and posted on a website for general usage, and a corresponding database file (such as an Access database) could be released on the same website for users who wish to integrate the MFL into their own data sources. In addition to the list of health facilities, all metadata and data rules should also be made as widely available as possible through similar formats. Some examples of dissemination platforms include:

- A printed report
- A spread sheet which could be used by external organizations for linking and standardization of their own data
- An interactive web portal

- Database dumps in specific formats which could be made available on a request basis
- An interactive map on the web (an Open Geospatial Consortium (OGC) compliant geographical data server)
- XML Web services which could be implemented as part of a broader enterprise architecture approach to the Health Information System (HIS)

CASE STUDY: KENYA MASTER FACILITY LIST

The Kenya Ministry of Medical Services and the Ministry of Public Health and Sanitation, Division of Health Information System (HIS) designed and implemented the Kenya Master Facility List (MFL). The MFL is a single, centrally maintained database of health facilities with a unique code for each facility and includes information on hours of operation, contact person, owner, facility classification, number of beds, types of services provided (e.g. EMOC, HIV testing), and geographic location. The MFL currently contains nearly 7,000 facilities each with a unique identifying code. The data are publically available at http://www.ehealth.or.ke.

Prior to the creation of the MFL, there were over 70 lists of facilities in Kenya each serving separate and distinct purposes. There was no common facility code meaning that data could not be linked between systems. Additionally, facility attributes like type, name, and managing authority were not standardized causing great confusion and duplication.

In 2008, the MFL Working Group, which included members from the Ministry of Health, United States Agency for International Development (USAID), the Regional Center for Mapping of Resources for Development (RCMRD), the Kenya Medical Research Institute (KEMRI)-Wellcome Trust, and the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) project partners examined all the existing facility lists, selected the most comprehensive and up-to-date, and combined the lists where possible to create the starting point for the MFL. The 'baseline MFL' was then built and verified between 2008 and 2010 through support from two USAID funded projects: Health Systems 20/20 (Abt Associates) and MEASURE DHS (ICF International). The verification process was incorporated into the routine work of the District Health Management Teams (DHMT) simplifying reporting requirements and encouraging local ownership. As the MFL evolved, paper sheets were upgraded to electronic spreadsheets and finally to a web-based data entry application. All DHMTs received training on this application, the collection of geocodes, and the standardized categorization definitions.

Since August 2010, the MFL has been maintained at the district level by the District Health Records Information Officer (DHRIO). Using the web application the DHRIO adds new facilities or updates facility information, including a geocode via global positioning system (GPS) receivers, during their routine inspections. The response has been positive both among district staff, within HIS and throughout international development organizations. For example, DHRIOs have mentioned that the new system helps them reduce their already heavy workload. Also, two recent facility surveys in Kenya used the MFL as the authoritative list for sample design. The MEASURE DHS 2010 Kenya Service Provision Assessment (SPA) survey and the World Health Organization 2010 Cold Chain Inventory used the MFL and its facility codes in data collection.

The Kenya MFL is a great example of a functioning, available health facility list. The MFL code has become the standard identifier for health facilities in Kenya simplifying data linkage. Integration with existing HIS reporting structures mean that the MFL is updated often. Public dissemination of the MFL ensures that the data are highly utilized.

FINAL REMARKS

A Master Facility List (MFL) is crucial to produce good statistics, with results that can be compared over time.

In this paper, an outline for the design of a Master Facility List has been presented. The MFL should consist of two components: a signature domain which uniquely identifies each health facility unequivocally through the use of specific attributes, and a service domain which should be populated with a minimum number of attributes which provide critical information on the types of services which are offered in the facility. The strategic direction of the MFL should be determined and implemented by a steering committee.

A data dissemination strategy and platform will ensure the wide-spread adoption of the MFL. The successful implementation of the MFL will ensure that long-term longitudinal comparison of health facility level performance can be executed, and that multiple information sources can be more easily integrated. This will in turn provide a more robust and accurate assessment of a countries' health system.

REFERENCES

(1) WHO (2008). *Measuring Health Systems Strengthening and Trends: A Toolkit for Countries*. Available at: http://www.who.int/healthinfo/statistics/toolkit_hss/EN_PDF_Toolkit_HSS_Introduction.pdf

(2) Health Facility Assessment Technical Working Group (2007). *The Signature Domain and Geographic Coordinates: A Standardized Approach for Uniquely Identifying a Health Facility (WP-07-91)*. Available at: https://www.cpc.unc.edu/measure/publications/pdf/wp-07-91.pdf

(3) WHO (2011). *Health statistics and health information systems: Service Availability and Readiness Assessment (SARA).* Available at: http://www.who.int/healthinfo/systems/sara_introduction

(4) MEASURE DHS (2011). *Service Provision Assessment Overview*. Available at: http://www.measuredhs.com/What-We-Do/Survey-Types/SPA.cfm

(5) IHFAN (2011). *International Health Facility Assessment Network.* Available at: http://www.ihfan.org/home/

(6) UNAIDS (2006). *Data exchange with the Country Response Information System and UN Agency Software.* Available at: http://data.unaids.org/pub/BaseDocument/2007/cris_de_web_final_en.pdf

(7) UNAIDS (2007). *IXF Version 3.0*. Available at: http://data.unaids.org/pub/Manual/2008/sdmx_ixf3_en.pdf

(8) DHIS (2011). *District Health Information Software - "DXF".* Available at: http://bazaar.launchpad.net/~dhis2-documenters/dhis2/dhis2-docbook-docs/files/head%3A/src/schemas/dxf_v1_schema/

(9) SDMX-HD (2011). *Statistical Data and Metadata Exchange - Health Domain*. Available at: http://www.sdmx-hd.org/

(10) DHIS (2011). *District Health Information System Documentation.* Available at: http://dhis2.org/documentation/

(11) WHO (2011*). Extensible Address Language (xAL) Standard Description Document for W3C DTD/Schema (version 2.0)*. Available at: www.immagic.com/eLibrary/TECH/OASIS/XAL_V2.PDF

Annex 1: Using a database for the MFL

The MFL can exist in a variety of formats; perhaps it is an Excel spread sheet, an Access database, or a relational database. Whether or not the MFL will be stored in a relational database will be determined in the planning stages, based on the country-specific objectives for the MFL and availability of resources. For example, if the primary purpose of the list is to serve a sampling frame for health facility assessments, then a listing of facilities in a spread sheet would suffice. However, such a list would be of very limited use, and would not tap into the full potential of having a MFL. In order to be able to link various data sources to enable better synthesis and analysis of data, the MFL should be stored in a relational database. This section is aimed at IT professionals and covers the following technical aspects of relational databases and their application to the MFL:

- Database selection
- Database design
- Link external sources to the MFL database
- Disseminate the MFL

1.1 Database selection

A relational database is a computer program that stores information about different objects in different tables as collections of rows and columns. Examples include Microsoft Access, MySQL, Postgresql, and Oracle. Almost all major database systems use SQL to access data stored within the system. For the purpose of this document, the term “database” refers to a relational database management system.

Information stored in Microsoft Excel spread sheets and Microsoft Word documents will not be considered databases. Although they are sometimes referred to as databases, they lack the necessary rigor imposed by a relational database. Information stored on such systems will have to be migrated to an acceptable format as previously mentioned.

Choosing which database system to use will depend on the desired objectives of the MFL including overall resources available, and the management and maintenance structure of the MFL. Single-user databases are often easier for non-specialists to construct, but limit the access of data to a single person at a time. Modern multi-user databases often have user friendly graphical user interfaces which can be used to design and interface with the database, and offer the advantage of allowing multiple persons and services to have simultaneous access to the database.

There are a multitude of open-source and commercial database product offerings. Deciding which product to use involves an analysis of the technical expertise of the staff that will design, implement, and maintain the system. Microsoft Access and OpenOffice Base are good options for single-user databases. Postgresql and MySQL are robust, open-source, multi-user database systems. There are many multi-user commercial systems such as Oracle, DB2, and Microsoft SQL server as well.

The choice of a database management system may be dependent on the IT standards that are already in use. Therefore, it is recommended to consult with in-country IT experts to determine if there is existing expertise in a particular database management system. If no in-house IT expertise is available, input from an external consultant may be useful in determining the most appropriate database management system for the project.

1.2 Database design

A flexible and comprehensive database design is a necessary component to the development of a MFL. *Figure 3* below provides a suggested database schema that can be used for the MFL in most major

relational database management systems. This schema has been developed based primarily on the World Health Organization's GIS unit health facility database as well as the minimum data requirements for a MFL given in Part 2 of this document. The schema presented here is meant to allow for the creation of a MFL with certain compulsory elements that should be present in all databases, while allowing the adaptation of the schema to country specific needs. This schema presents the absolute minimum number of elements that are required in order to effectively ensure that the facility can be uniquely identified in a database and enable exchanges with external systems.

Given the specific system requirements of the MFL, the database system, dissemination strategy, and other factors, the final database design will likely differ significantly from the basic outline presented in *Figure 3.* A database developer may be required to modify the recommended schema in *Figure 3* and build the appropriate MFL database.

Other tables related to presentation of the data, security, and linkage with other systems fall outside the scope of the technical discussion in *Figure 3*. Other system development such as graphical user interfaces and web services which would involve the use of procedural language would also need to be implemented, but these also exceed the scope of this paper.

Multilingual considerations

Should the MFL need to support multiple languages, some special considerations should be taken into account. It is recommended that in all cases the UTF-8 encoding system will be used, especially for non-Latin languages. Other encoding systems such as the WIN-1252 encoding system often used to represent Cyrillic languages present many challenges, especially as it relates to different systems storing and retrieving data from the database system. For many Asian languages, the use of UTF-8 is essential to avoid unexpected results when querying or storing data represented in complex scripts. For other languages that only use Latin characters, the use of UTF-8 may not be required, as long as no non-ASCII characters are used anywhere in the database. In general however, the use of a UTF-8 encoded database, along with proper encoding of the information that is stored in the database, is recommended.

Database field types

Depending on the rules that have been established, an appropriately restrictive field type in the database should be chosen. For instance, if an integer code has been determined to be used as the unique identifier, when designing the database an “integer” type should be chosen, instead of “character varying 10”. Although the “character varying 10” type is capable of storing an integer, it is also capable of storing non-numeric characters, which will lead to unexpected results when retrieving the code. As a rule, the most restrictive field type should be used for each attribute. If a postal code is known to only have 5 characters, then the field type that should be chosen is “character varying 5”. Other business rules may need to be established that further restrict the nature of the information that is stored in the field (e.g. the facility name should not be written with all capital letters).

Database schema

The development of a formal database schema will depend on the detailed user requirements of the MFL. An exceedingly simple version of a basic MFL is presented in Figure 3 below as an entity-relationship diagram. Note that “PK” refers to a primary key constraint and “FK” to a foreign key constraint. Explicit field types are not provided in this diagram, and would need to be developed for actual implementations.

The model presented here is based to a large extent on the DHIS2 data model *(10)*, with some simplifications for clarity. The signature domain consists of a main “organisationalunit” table which contains all information related to the signature domain of an organization unit. Organizational units could be districts, facilities, or another entity (such as a ward within a facility) which is contained within the

organizational hierarchy of a given country. A self-referencing foreign key reference within the table establishes parent-child relationships between organizational units. As an illustration, a district would serve as the parent to a facility. This self-referencing structure allows for a more compact and standardized representation of a hierarchy, rather than establishing separate tables for “district” and “facility.

Lower order normalization may be desirable in an actual implementation, e.g. by separation of the “organisationalunittype” structure into separate physical tables to represent different organizational unit groups, such as “Ownership”, “Type”, etc.

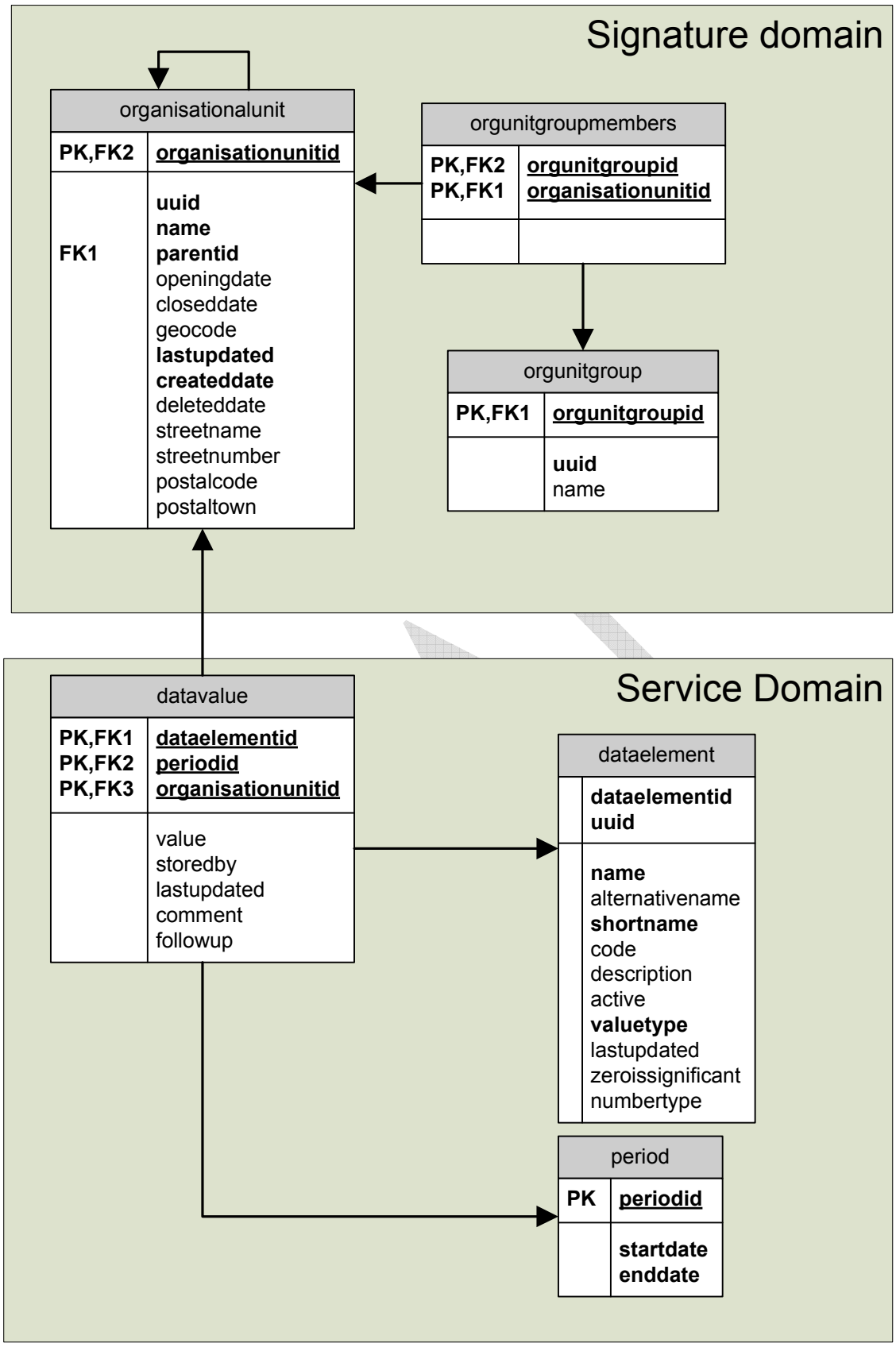
The service domain consists of a central fact table where values of specific data elements can be stored, referenced to an organizational unit, period and data element. A separate table of data elements, which would contain the detailed metadata on each service domain attribute, is shown.

Note that the service domain has an implied parent-child relationship with regards to organizational units. For instance, a district contains multiple facilities. The parent-id attribute will implicitly define which parent district a particular facility belongs to. For the purposes of this schema, a district is simply a different “type” or organizational unit. A key difference with the signature domain is the implied association of the data in the service domain with a particular time period. The signature domain is regarded to be essentially immutable. Should there be a need to record historical information, such as when a facilities name is changed, a separate audit structure would be recommended to deal with this historical data. However, for the service domain, the data should always reference a particular time period (for instance the year 2011). This structure will allow for analysis of changes in the service profile of a given facility over time.

Address considerations

The elements presented in the formal database schema in this Annex have been adapted from the OASIS Customer Information Quality (CIQ) family of specifications and specifically the extensible address language (xAL) *(11)*. This set of standards is an open framework for specifying addresses in different countries. The use of this international standard would be recommended in cases where no existing health facility database exists, and especially where there is no accepted national standard for addresses. In cases involving existing national systems, these standards should be considered for use if they are more appropriate than the standards already used by the national system. Mapping of concepts between national systems and the internationally recommended standard (xAL) should be conducted to ensure that data exchange between these proprietary systems and standardized systems could occur when needed.

Figure 3: Database schema of a basic MFL



1.3 Link external sources to the MFL

The Master Facility List has many uses, many of which are far broader than simply listing a country's health facilities. In fact, the value of the MFL is fully realized when it is used to link different data sets. The implementing agency needs to promote, support, and enforce the use of the MFL both internally and externally by demonstrating how it can be used as a tool to link disparate datasets together. The dissemination of the data contained in the MFL, both to the public and to other agencies and ministries, is crucial to ensure the effectiveness and maximum utility of the information.

In order to link the MFL to external sources, a combination of technical and policy practices must be instituted. The implementation of technical solutions, which may enable two data sources to communicate with each other, on its own, is not sufficient. Management practices must be put in place to ensure when information is updated or changed in the MFL, subsequent updates occur in any database that depends on or links to the MFL. This section will outline both managerial as well as technical aspects of how the MFL can best be put into use.

In order to successfully ensure that data from the MFL can be utilized by external systems, several conditions must be met including the following:

- A common understanding of metadata exists between systems.
- Connections for data exchange, such as web services or databases, have been made available.
- Methods for exchange, such as standardized data exchange formats, have been agreed upon.
- Management practices have been put into place to ensure that the data exchange actually occurs.

The implementation plan should specify how and to what extent linking to external sources is supported. Direct database linkage with a common unique identifier is the easiest way to link two different datasets. However, in reality this is not always possible. Other options need to be provided. Some possibilities include crosswalk tables or data exchange with XML, Comma Separated Values (CSV) or other formats.

Whatever options will be chosen, metadata standards are crucial. Metadata standards encompass both technological and semantic standards of concepts between systems. Technological standards ensure that information can actually be exchanged between two information systems, whereas standardized semantics allow for similar concepts in two separate systems to be matched to each other.

A. Metadata standards

When connecting data from multiple sources, common definitions, called metadata, are essential to understanding the characteristics of each piece of information which is collected. One database may refer to the name of the facility as “Facility name” while another may refer to the same concept as “Name of facility”. Although from a semantic standpoint these are essentially the same, from an informatics standpoint, they are distinct. In order to harmonize data from multiple sources, it is crucial that a common understanding of metadata concepts is established. Metadata may even differ between different institutions within the same country, and unambiguous definitions for the metadata elements are essential when merging disparate data sources. Metadata standards encompass both technological and semantic standards of concepts between systems. Technological standards ensure that information can actually be exchanged between two information systems, whereas standardized semantics allow for similar concepts in two separate systems to be matched to each other. The coordinating group will be responsible for the final authorization of these standards. For more information and guidance about metadata concepts and applying them to the MFL, it may be useful to engage a consultant or include in the coordinating group individuals who have had experience with metadata concepts. This may include researchers from academia, IT staff from key government ministries, or private industry.

Semantic standards

In order to allow for similar concepts in separate systems to be matched, two elements are necessary: a concept dictionary and a thesaurus.

A concept (or metadata) dictionary provides a detailed description of all metadata concepts and elements. For example, the metadata concept “Managing authority” would be defined in the concept dictionary. Each metadata concept contains several metadata elements which are also clearly defined in the concept dictionary. For example, the metadata concepts “Private ownership”, "Public facility", and "Faith-based organization" would all be defined in the concept dictionary.

A concept dictionary is essential for building a semantic understanding of the concepts that have been defined in the MFL and to allow for the mapping of similarly named concepts to one another. Databases have limited ability to interpret variations in concepts. For instance, a human can easily recognize that “Type of ownership” and “Ownership type” are the same concept; however a database would not recognize that these two slightly different phrases refer to the same idea. In order to allow for the automated exchange of data between distributed systems, concepts must be defined in a standardized manner in the form of a dictionary. Within each concept, each of the metadata elements, such as “Private ownership”, “Faith-based organization”, or “Public facility” must also be clearly defined.

A thesaurus lists words grouped together according to similarity of meaning, or synonyms, and allows for the mapping of similar concepts. A metadata thesaurus is helpful for mapping similar concepts and metadata elements to each other. For example, the attribute “Private ownership” is a synonym of “Privately operated”. The thesaurus would specify this so that the database can recognize these two terms as the same. If there are multiple terms in use within a given country, a metadata thesaurus can be used to map these similar terms to each other.

Technical standards

From the managerial standpoint, practices would need to be instituted to ensure that there is a technical mechanism to allow for the exchange of metadata information, including updates. Additionally, changes may need to be made in the recipient database in order to comply with the metadata standards of the MFL, especially if bi-directional transfer of data is to occur.

Standards for data exchange

In order to bring about the actual exchange and update of (meta)data between the MFL and external sources, the metadata would be transmitted in some type of message such as XML, CSV or other format which has been mutually agreed upon. A number of XML schemas such as Statistical Data and Metadata Exchange (SDMX), Dublin Core, and other XML schemas offer technical mechanisms to effect data exchange of statistical data and associated metadata. Use of one of these standard schemas is recommended, unless there are compelling technical reasons not to use them. Development of a custom schema is also a possibility, but will limit the number of systems that the MFL may be able to provide data to without extra development effort. If there is any existing in-house expertise with a specific data exchange standard, it is recommended to engage those experts in the process of selecting and implementing the data exchange standard. If no in-house expertise is available, an external consultant may be required to select and implement the most appropriate data exchange standard.

Data exchange can occur through several different mechanisms, but it is typically accomplished through either direct database linkage, or exchange of an XML or CSV message that has been encoded in such a way that the system importing the message can recognize, transform, and import the data into its own structure. Alternatively, data may be manually inputted from one system into the other, but this method should be regarded as inefficient and prone to error. Examples of these data transfer mechanisms are provided in the following sections.

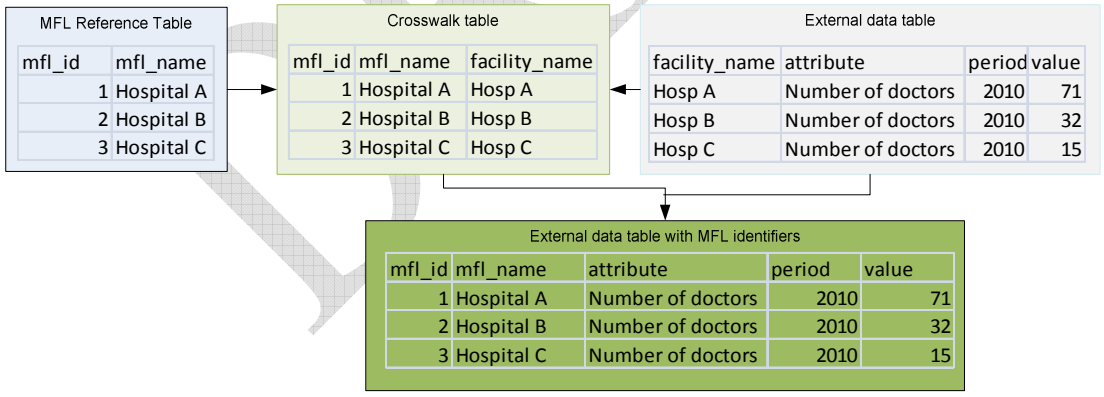
Direct database linkage

Direct database linkage involves the use of common key fields that allow two separate database tables to be related through common fields. The use of the unique identifier, discussed previously in the signature domain section, provides the most direct mechanism to link two separate tables together. Consider a health facility survey that has been conducted to determine service availability and readiness: if the survey has made use of the unique identifier, and this identifier is contained in each row of the database table, the survey can be linked to the MFL fairly easily through the unique identifier field.

If the two database tables do not use the same unique identifier, a cross-walk table will typically need to be developed. A crosswalk table is a table composed of two columns which establish a one-to-one relationship between the primary keys of a table of the source database with a target table of the target database. Establishing a crosswalk table can be tedious, especially if there is not a common field. As an example, the MFL may contain an entry "Pearl of Health Hospital" while the target database to be linked to the MFL may list the facility as "PEARL OF HEALTH HOSP". In this example, it is fairly easy for a human to recognize that these two entries refer to the same facility; however this task is not straightforward for a computer. If there are a large number of facilities which have to be matched, manual matching of the source and target databases may be required. Once the crosswalk table has been established, however, the two databases are simple to link through a normal database table join on the common field.

A more detailed example is provided below in Figure 2: Linking databases and developing a crosswalk table. The "MFL Reference Table" on the left of the diagram represents a database, such as the MFL, to which an "External data table" should be linked to. In this example, the external data table provides the number of doctors in each facility; however, there is no standardization of names between the source and target tables, which is a fairly common scenario. The crosswalk table provides a linkage between these two data sources and contains at least two fields. In this example the names of the facilities in MFL reference table may be "Hospital A", "Hospital B", etc. In the external data table, the same facilities may have slightly different names such as "Hosp A", Hosp B", etc. The crosswalk table establishes a one-to-one correspondence between the source and target tables, and allows the two tables to be linked through a normal database join. This enables the production of an "External data table with MFL identifiers".

Figure 2: Linking databases and developing a crosswalk table



In the ideal situation, the names present in the "External data table" would be modified to correspond to the official names which would be obtained from the master facility list. Should this not be possible, a crosswalk table can serve as a mechanism to impart the MFL official names and identifiers to an external data source, thereby making this external data source more easily interchangeable among all systems which utilize the MFL identification scheme.

Data exchange with XML

In many distributed information systems, XML is the language of choice to exchange data between disconnected nodes and different information systems. The use of a standard XML schema between different parties participating in the exchange of data ensures that multiple specialized systems can coexist within a data ecosystem. As an example, information systems that manage human resources, logistics, and laboratories have been developed to address these particular business models. The strategic health management information system collects the most important data from each specialized system that is most relevant for decision making. For instance, the human resource database may be responsible for the detailed procedures related to hiring and termination of employees, but for the purposes of the strategic HMIS system, the only relevant information may be the number of doctors and nurses present at each facility. In order to ensure that data can be exchanged between the Human Resources (HR) system and the HMIS there are several technical hurdles that must be overcome:

- A common data structure, such as an XML schema
- A common understanding of concepts (e.g. data elements and indicators)
- At least some common field to allow the data to be linked

A metadata dictionary should provide a standardized and comprehensive list of common concepts such as what the data element "Number of doctors hired this month" actually means. A common field, such as a facility code, allows the two systems to make a linkage. An XML schema is analogous to a common script that enables the data exchange to take place, but does not necessarily guarantee that the two systems will understand each other.

While an exact schema is difficult to provide given the wide variability of systems, there are several initiatives which have produced XML schemas specifically for data transfer between aggregate data nodes and systems including IXF2 *(6)*, IXF3 *(7)*, DXF *(8)*, and SDMX-HD *(9)*. SDMX-HD is regarded by some to be the state of the art in regards to exchange of statistical metadata exchange.

B. Harmonization of external data sources

Once the MFL has been developed, it can be used as a tool to link multiple data sources together. This coordination can be accomplished because of the invariant nature of the signature domain, which should provide a mechanism to link various data source through the use of common fields.

Identify data from different sources (from HFAs, surveys, HMIS)

In many countries, facilities surveys are conducted every few years. Linking routine information sources with surveys can provide important information regarding the quality of the routine information system. A difference in the same indicator, which has been determined through a direct survey as opposed to the routine information source, provides important insight into the reliability of the HMIS.

Prepare external sources for linkage to a MFL and maintain linkages across databases

It is important that any facility survey conducted after MFL has been established uses the unique identification code of the facility from the MFL and embeds them within the survey itself. However, if a facility survey must be linked with the HMIS ex post facto, the approach that was outlined in the direct database linkage section would need to be performed. Namely, the corresponding metadata elements (such as the name of the facility) would need to be matched to the MFL. Usually, the name of the facility which is provided in the facility survey will not match exactly the name present in the MFL. In this case, a crosswalk table would need to be produced, which would allow the two data sources to be linked through a standard database join procedure.

Similar to the facility code, other data elements such as the facility type, ownership, etc. should also be matched if possible. Usually this matching process is manual, but some level of automation may be possible by joining the names from the MFL to the external data source, and then manually resolving the names that

do not match. Those responsible for database management and maintenance will also be responsible for preparing the external data sources for linkage to the MFL, as well as for maintaining these linkages.

CASE STUDY: KENYA Linking Data to a Master Facility List

Key decision/action issue: The Chief Nursing Officer (CNO) is responsible for nursing and workforce management and s/he needs good information to manage effectively. The NWIS (Nursing Workforce Information System) provides data on nurses in over 6,000 facilities. However, the Division is not able to analyze or report nursing workforce data based on new districts or constituencies. The nursing data cannot presently be mapped and does not contain number of beds. The Chief Nursing Officer has asked for information about facilities and the nurses who work there:

- Where are the facilities?
- How many beds are in each facility?
- What is the ratio of beds to nurses for each facility?

Type of data

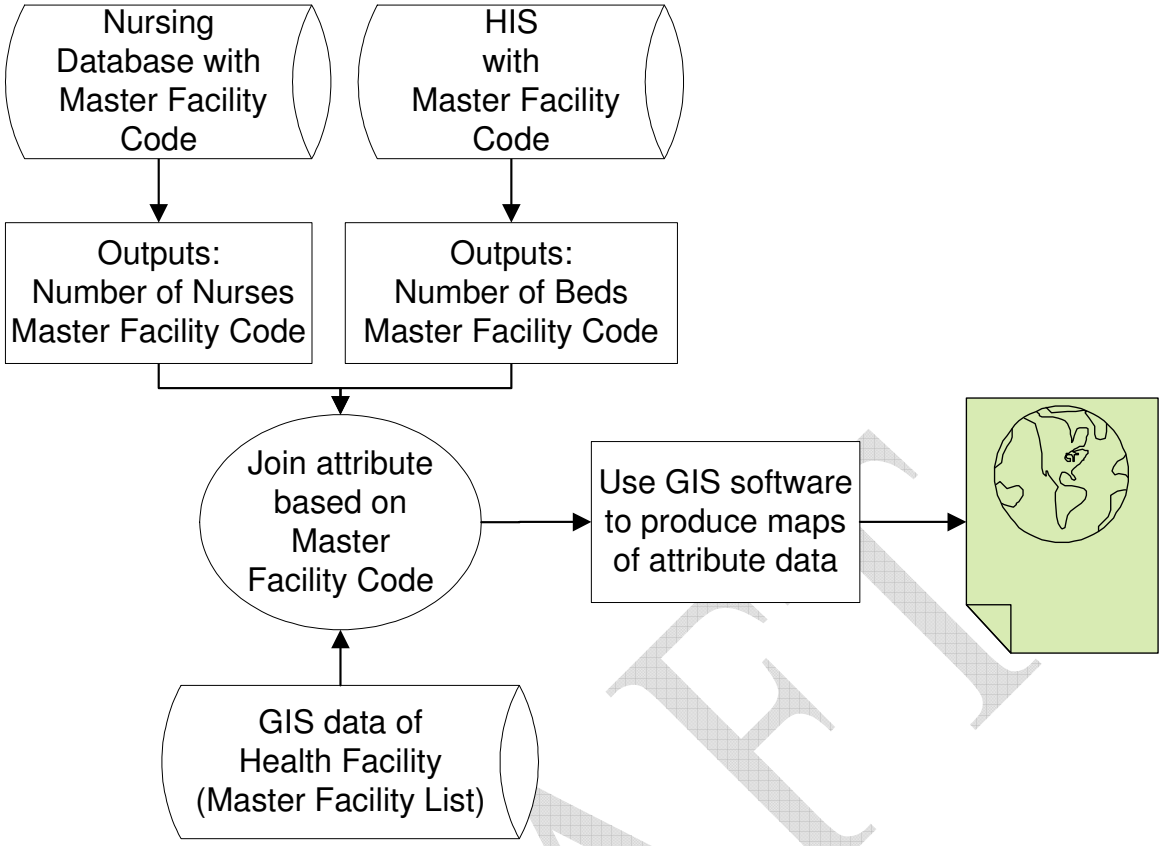
The data needed includes the number of facilities, their location, the number of nurses at each facility, and the number of beds at each facility.

Data Source

The data from the nursing workforce database and data from the master facility list need to be linked in order to answer these questions. What does "linking" to the MFL Mean? The process is simple but time consuming. Someone needs to lookup each facility in the nursing workforce database, match this to a facility in the master facility list, and then enter the MFL code.

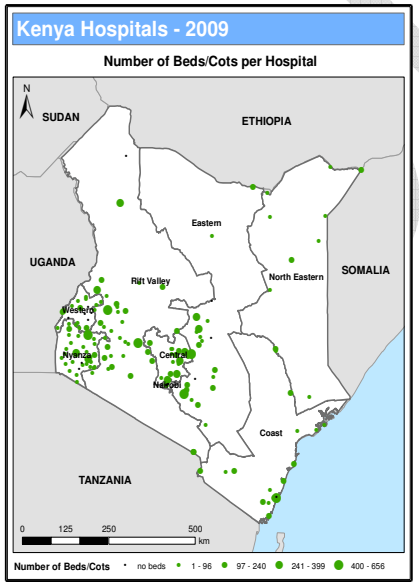
CNO-PROVIN	CNO-OLD_DIST	CNO_CC	CNO_TY	CNO_NAME	CNO_AGENT	Nurses	MFL_CODE
Central	Kiambu	00027	Hospital	Kiambu District	Government Of Kenya (G	196	10539
Central	Kiambu	00071	Hospital	Tigoni Sub-District	Government Of Kenya (G	68	11104
Central	Kirinyaga	00083	Hospital	Kimbimbi S D	Government Of Kenya (G	19	10609
Central	Kirinyaga	0301	Hospital	Kerugoya District	Government Of Kenya (G	152	10520
Central	Maragua	00100	Hospital	MARAGUA DISTRICT	Government Of Kenya (G	39	10686
Central	Murang'a	00008	Hospital	Muriranjas SD	Government Of Kenya (G	41	10782
Central	Murang'a	00067	Hospital	Muranga District	Government Of Kenya (G	172	10777
Central	Nyandarua	00050	Hospital	Nyahururu District	Government Of Kenya (G	138	10890
Central	Nyandarua	00093	Hospital	Ol'Kalou	Government Of Kenya (G	106	10916

Once this process is complete, the databases can be linked together using the unique MFL code. The nursing database now has the MFL data. These facility data are maintained by the national Health Management Information System within the Ministry of Health. This database contains information on new district names, facility type and ownership, constituency, division, location, sublocation, contact information, number of beds/cots, and geocodes.

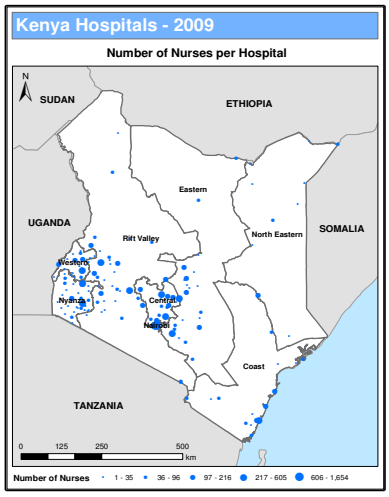


Data analysis

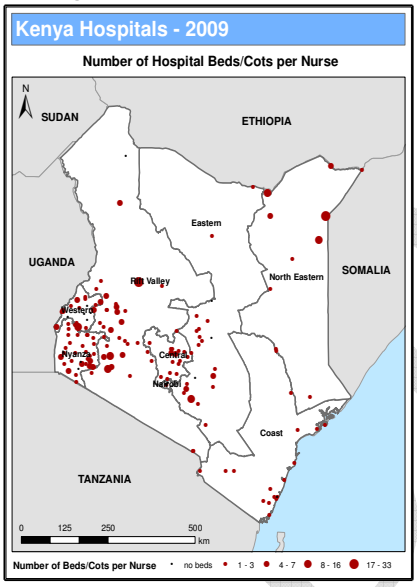
The nursing database can now be analyzed and visualized in new, useful ways like with maps. In addition, the nursing workforce data can be combined with any other dataset that uses the MFL codes. You begin by mapping the data from each database. From the master facility list, the number of beds/cots per hospital can be mapped by province.



Using the geocodes from the master facilitiy list and the number of nurses from the nursing workforce database, we can map the number of nurses per hospital by province.



Combining this information, we can now map the number of beds/cots per nurse per facility by province. These figures were previously not possible because neither database contained all this information. However, by linking the two databases with the MFL code, we can extract new, useful data. Quickly looking at the map, we notice areas in North Eastern Province and Rift Valley Province where there are high numbers of beds per nurse. You can extract more information from the MFL about these facilities. In your report you list the facility name, district and consitutency of these facilities and suggest targeting them for nursing recruitment efforts.



Annex 2: Determination of geographic coordinates

Collection of the geographic coordinates of a health facility should be conducted where possible. Several approaches can be used to obtain or update the geographic position of a health facility. The decision of which method to use should be guided by the precision that is necessary to answer program-relevant questions as well as the available human and technological resources. The use of the geographic coordinates should be enumerated as part of the initial strategic user requirements for the MFL. Examples may include use of coordinates to optimize supply chain routes or estimation of facility catchment areas. For these purposes GPS or satellite imagery can provide an appropriate level of precision. For rough cartographic representation or as a placeholder until more accurate coordinates are obtained, some combination of the other methods may be adequate.

The market price of GPS devices has dropped significantly over the past few years, and a dedicated device can be obtained for less than 200 USD. In recent years, many mobile telephones and other mobile computing devices have integrated GPS devices. Geospatial tools such as Google Earth have made high resolution satellite imagery more accessible. Although the collection of the coordinates is relatively straightforward, there are some important points to keep in mind, which will be outlined in this section. This section will provide some tips and best practices that should be followed.

The following methods are listed in descending order of the precision they provide (from the highest precision method to the lowest):

- Direct survey of the facility with a quality GPS device.
- Use of satellite imagery or aerial photography, including platforms such as Google Earth.
- Determination of the location with the use of existing national topographical maps.
- Geocoding of other attributes attached to the facility, such as the name of the town in which it is located, the centroid of an administrative unit in which it is located, or a street address.
- Indirect determination of the coordinates through a known source for which coordinates exist, such as proximity to a school, water source, or marketplace.
- Use of scanned and georeferenced hand-drawn maps, which are often available through local leaders or district health authorities.

Direct survey with a GPS device

Regardless of the particular device that is used, all GPS receivers should be used outdoors, in a large, open area that has a clear view of the sky. Attempting to collect GPS coordinates under thick vegetation or in a dense urban area may be challenging. Where a large open area is not possible, more time may be required for the GPS receiver to establish a suitable fix on the satellites.

Latitude and longitude

Latitude is the north/south value measured from the equator. Longitude is the east/west value measured from the Prime Meridian that runs through West Africa and Western Europe. A latitude and longitude identifies an exact location on the earth's surface.

Note that some GPS receivers will display the letters N (north) or S (south) in front of the latitude, and W (west) or E (east) in front of the longitude. For example, the point represented by N 45 degrees latitude, W 45 degrees longitude is equivalent to +45 degrees latitude, -45 degrees longitude. Thus, based on the location of the health facility positive (+) or negative (-) coordinates should be properly reported. Typically however, most countries are located such that longitudes and latitudes are always either positive or negative.

Often the elevation will be displayed on the same screen and might be useful information to record as well. For instance this is useful in malaria endemic countries, to determine which health facilities are located in malarial and non-malarial regions. Hence, it may be usefull to recor the elevation, even though it is the least reliable reading on a GPS receiver, and mauy be subject to large fluctuations.

Configuring the GPS receiver

- Before starting data collection, the GPS receiver's setting should be configured properly.
- Make sure that if the receiver had been previously used to collect data, that the data is downloaded to a computer. The GPS receiver's internal memory should then be cleared of all previously collected data.
- To ensure optimal compatibility with other geographic information, it is advisable to record the latitude and longitude of a facility in decimal degrees. The GPS receiver should be set-up as follows:

Navigation units	Metric/Km
Map datum	WGS 84
Coordinate reference system	Lat/Lon (Hddd.dddd)
North reference	Magnetic

GPS receivers have the ability to collect data in many other coordinate reference systems, and this may be appropriate in some countries where there is a well-established geospatial infrastructure, with established guidelines on which particular coordinate reference system should be used. The MFL coordinating group should connect with the national mapping agency to determine the existence of any preferred national coordinate reference system.

Collecting geographic coordinates with GPS

Once you have configured the settings on the GPS receiver, you can use it to record the geographic coordinates of a facility. GPS receivers can store data internally but it is recommended to record important information such as the latitude and longitude and the name of the waypoint on paper and/or other data collection being used. To record the geographic coordinates of a facility:

- Stand in an open area with clear view of the sky at the health facility entrance so that its antenna is able to receive signals. If it is not possible to obtain a clear view of the sky due to buildings or other obstructions, collect a point in an area where there is a clear view of the sky and record the offset parallel to the ground. Some receivers are less effective when held parallel to the ground and need to be held perpendicular to the ground. Follow the rules listed below depending on the type of facility.
- Do not collect a coordinate until the receiver indicates it has acquired a signal from enough satellites to produce an accurate reading. This message usually reads "Ready to navigate" but may vary from device to device. Many GPS receivers will indicate an estimated accuracy. Wait until the receiver unit has stabilized and the accuracy is less than 10 m. If the receiver does not quickly have sufficient accuracy or the message "Ready to navigate" does not appear, wait at the same place for about five minutes until the readings stabilize.
- When the signal is sufficient and the accuracy is at the recommended level, the geographic coordinates can be recorded. The coordinates will be stored on the GPS device, but should also be recorded using a paper or electronic data collection form as a back-up to the GPS unit. The geographic coordinates may change slightly every few seconds; these fluctuations are usually very minor and can be safely ignored.
- In order to conserve battery life, switch off the GPS receiver once the geographic coordinates are recorded.

The rules for GPS data collection for different facility types are as follows:

1. **Single facility in a building**

- The geographic coordinates should be recorded in front of the main sign attached to the building of the facility.
- If there is no sign attached to the building then the geographic coordinates should be recorded in front of the main door or reception area of the facility.

2. Multiple facilities in a single building

- The geographic coordinates should be recorded in front of the sign(s) that lists what facilities are located in that building (if the sign is outdoors and attached to the building).
- If there is no sign listing what is in the building (or if the sign is indoors), the geographic coordinates should be recorded in front of the main entrance door or reception area of the building.

3. Single facility in multiple buildings

- The geographic coordinates should be recorded in front of the door or main entrance to the reception area of the facility (preferable where there is the main sign for the facility). If there is no reception area the geographic coordinates should be recorded in front of the door to the administrative offices of the facility.

GPS Check-list

Before collecting data with the GPS receiver, the following points:

- Check battery level and verify that the GPS receiver works properly. You can verify this by turning on the GPS receiver, and waiting approximately 10 minutes until a fixed location is established. After the fix has been obtained, walk away from your initial position and see if the values change. Return to your initial location, and see if the values return to their initial values. This should provide an indication of whether the GPS receiver is working properly.
- Check the settings of your GPS receiver (Hddd.dddd, WGS 84, etc.). Be sure that the latitude and longitude are displayed in decimal degrees and that the WGS84 datum has been set. If another coordinate reference system will be used, ensure that the GPS receiver is set to that.
- Have the questionnaire that will be used for data collection ready (paper form, PDA form, Excel form).
- Stand in an open area with a clear view of the sky.
- Record geographic coordinates noting positive and negative values.
- Make sure that all information relative to the facility has been recorded.

In many cases, geographic coordinates of a facility are recorded as part of a larger data collection exercise, such as a health facility assessment (SAM, SPA, SARA, or HFA). Because primary data collection is expensive and time-consuming, it would not be efficient to visit all facilities in a country to collect only the geographic coordinates. In many countries, remote and hard-to-reach rural facilities may take several hours to reach. Since a health facility assessment would usually take an hour or two to conduct, it would make more sense to collect the geographic coordinates as part of the broader assessment of the facility. It is also possible to equip district health information officers with GPS receivers, and have them record the facility location as part of a normal health facility visit. The geographic coordinates could then be entered into the MFL database at a later point in time.

Use of satellite imagery and aerial photography

If high resolution satellite imagery or aerial photography for a given region or country is available, it is possible, coupled with very good local knowledge, to precisely locate a given health facility. Google maps and other online sources of data have made available high resolution satellite imagery for many parts of the world.

As an example, using Google maps, one can locate the approximate location of Lusaka, Zambia. Knowing the approximate location of a facility, one can continue to zoom into the location until the desired feature

is found by examining known landmarks and street names where available. For example, University Teaching Hospital located in Lusaka was determined to have a coordinate of 15.431805 S, 28.314899 E. This information was obtained fairly quickly and without the need for a physical visit to the facility. Placemarks can be created in Google Maps or Google Earth for facilities and these placemarks can be exported to a KML file or manually entered into a spread sheet. The exact steps necessary to do this is beyond the scope of this document, but possible resources that can provide assistance include staff at national mapping agencies, faculty at national universities in departments of geography, surveying, public health, and/or Google Earth/Google Maps documentation.

Local knowledge of landmarks and a good sense of geography are required to perform this type of geocoding, but it is certainly possible to utilize these data sources for the precise determination of a facility's location.

Geocoding of facilities

In many cases, the approximate location of the facility may be known, such as the name of a town, ward, or neighbourhood in which the facility exists. Typically, geocoded databases of towns are usually readily available, including the free Geonames service (http://www.geonames.org/) which offers an extensive database of towns and administrative district locations throughout the world. Geocoding involves the matching of place names (such as town or ward) attached to a health facility, to the names in an established database. By searching for the name "Livingstone" in the country Zambia, through the Geonames web service, a geographical coordinate of -17.84221, 25.86336 is returned.

One major limitation with this type of geocoding is the level of precision. If multiple facilities exist in a single location, all the facilities will be assigned the same geographical coordinates. Thus multiple facilities will appear to be in the same place. Also, the coordinates may not be precise enough. Geocodes with fewer decimal places are not as precise. In the example above, geographic coordinates of -17.8, 25.8 for Livingstone, Zambia are not as accurate as -17.84221, 25.86336. In fact, these two locations are over 8 km apart.

Additionally, not all place names are unique. There may be multiple places with the same name. Care should be used when assigning geographic coordinates based on name alone. Subsequent follow-up by other means, such as the use of satellite imagery, to determine the exact location, if a physical visit to the facility is not feasible, will be required.