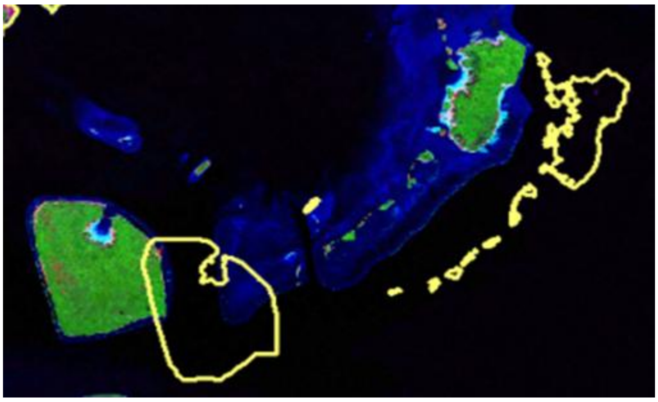
Guidance for the management and use of geospatial data and technologies in health

Part 2 - Implementing the geospatial data management cycle: 2.2 Defining the terminology, data specifications, and the ground reference

Version 1.5 (last update: 22.06.2024)



In collaboration and with the support of:

Revision history

Version	Release date	Comment	By
1.0	September 2016	Document created	Steeve Ebener
1.1	09 June 2018	Relabeling of the document, inclusion of the reference to the new HGL guidance documents	Steeve Ebener
1.2	18 May 2020	Adjustment of the terminology associated to the concepts of geographical features, geographic objects and master lists to align with other volumes of the series	Steeve Ebener
1.3	10 January 2022	Inclusion of the reference to the new HGL guidance document	Izay Pantanilla
1.4	19 February 2024	Update the layout, section 4.2.1 and the URLs across the document	Steeve Ebener
1.5	22 June 2024	Added generic examples of data specifications in Annex 1	Steeve Ebener

Author

Steeve Ebener[1]

1. Health GeoLab, Manila, Philippines

Acknowledgements

Our gratitude goes to the Asian Development Bank (ADB) and the World Health Organization (WHO) for the support provided to the Health GeoLab.

Table of Contents

Purpose and audience

The purpose of the Health GeoLab series of guidance is to inform concerned practitioners about the key elements they need to be aware of when it comes to managing and using geospatial data and technologies in public health and guide them through the processes to be followed in that regard.

The audience for this guidance includes geospatial data managers, technical advisors, and any other practitioners that are directly or indirectly involved in the collection and use of geospatial data and technologies in public health.

Please note that some of the sections in the present guidance require a basic understanding of concepts pertaining to the management and use of geospatial data and technologies.

Abbreviations

ADB	Asian Development Bank
AeHIN	Asia eHealth Information Network
CDC	Centers for Disease Control and Prevention
DOH	Department of Health
GIS	Geographic Information System
GMS	Greater Mekong Subregion
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
HGL	Health GeoLab
IEA	International Epidemiological Association
UNAIDS	United Nations Programme on HIV/AIDS
WHO	World Health Organization
WGS	World Geodetic System

1. Background

The Health GeoLab (HGL) is a regional resource supporting low- and middle-income countries in Asia and the Pacific for them to fully benefit from the power of geography, geospatial data, and technologies to reach the health-related Sustainable Development Goal of healthy lives and well-being for all (SDG 3)[1].

The HGL uses the HIS geo-enabling framework to strengthen in-country capacity. The present document has been developed as part of this approach and with the objective of being used by the largest number of users possible.

This volume is part of a series of guidance started under the umbrella of the AeHIN GIS Lab and now continued by the HGL. The complete series is organized as follows:

- Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle [1]
- Part 2 - Implementing the geospatial data management cycle:
    - 2.1 Documenting the process and defining the data needs [2]
    - 2.2 Defining the terminology, the data specifications, and the ground reference (the present document)
    - 2.3 Compiling existing data and identifying gaps [3]
    - 2.4 Creating geospatial data
        - 2.4.1 Extracting vector format geospatial data from basemaps [4]
        - 2.4.2 Collecting data in the field [5]
    - 2.5 Cleaning, validating, and documenting the data
        - 2.5.1 Documenting the data using a metadata profile [6]
        - 2.5.2 Using advanced Microsoft Excel functions [7]
    - 2.6 Distributing, using, and updating the data
        - 2.6.1 Creating good thematic maps using desktop GIS software [8]
        - 2.6.2 Using thematic maps for decision making [9]
        - 2.6.3 Developing and implementing the appropriate data policy [10]

This guidance is a living document made to evolve based on the inputs received from the users. Please don't hesitate to contact us if you have any suggestions for improvement.

The terms used in the present guidance are defined in the following glossary of terms maintained by the Health GeoLab: https://bit.ly/3ctoHiS

Please also contact us using the same email address should you use this document as part of your activities and would like to have your institution recognized as one of the document's users.

[1] https://www.un.org/sustainabledevelopment/health/

2. Introduction

Once the objectives, expected outcomes, and related data needs are defined [1, 2], the next steps in the geospatial data management cycle consist in:

1. Ensuring that all the stakeholders involved in the geospatial data management cycle talk the same language using a common glossary of terms,
2. Defining the data set specifications that will be used to ensure compatibility among sources, and
3. Identifying the mosaic of satellite images and establishing the master lists[2] that will serve as ground reference.

The present document's objective is to describe in more details and provide recommendations on the above-mentioned steps. It builds on previous publications [11], guidelines developed for the Department of Health of the Philippines (DOH) in collaboration with the Country Office of the World Health Organization (WHO) in the Philippines [12] as well as some material elaborated for the Asian Development Bank (ADB) in the context of the Region-Capacity Development Technical Assistance (R-CDTA) 8656: Malaria and Dengue Risk Mapping and Response Planning in the Greater Mekong Subregion (GMS).

3. Defining the terminology

The use of common terminology ensures that all stakeholders involved in the collection, maintenance, and use of geospatial data understand each other.

Two types of glossaries can be differentiated when dealing with geospatial data, namely those covering terms related to the management and use of geospatial data and technologies and those covering thematic terms.

While different glossaries do exist for the terms related to the management and use of geospatial data and technologies, the one included in the GIS.com wiki [3] remains the most comprehensive online resource and therefore the one being recommended by the Health GeoLab [4]. Another option is the GIS dictionary maintained by Esri and which is accessible either online [13] or in print.

By thematic terms, we mean all the terms related to the public health issues being addressed using geospatial data and GIS. Among such glossaries, we can mention the:

- The UNAIDS terminology guideline when it comes to HIV/AIDS [14],
- The IEA dictionary of epidemiology [15], and
- CDC's Malaria glossary [16].

This list is far from being comprehensive and is definitively not prescriptive. These are just examples to illustrate the type of resources that can be found on the internet.

These glossaries are living/evolving documents which get updated regularly depending on the evolution of the subject they cover and/or the institutions in charge of overseeing them.

[2] Unique, authoritative, officially curated by the mandated agency, complete, up-to-date and uniquely coded list of all the active (and past active) records for a given type of geographical feature/object (e.g. health facilities, administrative divisions, villages)

[3] http://wiki.gis.com/wiki/index.php/GIS_Glossary

[4] Terms used in the present set of guidance documents that are not included in Esri GIS dictionary were obtained from other sources. The sources in question are indicated in the text.

In addition to that, the definitions contained in these glossaries often have to be contextualized locally in order to account for strategies, plans, practices,... in force in countries.

Finally, language can be an issue, meaning that global or regional glossaries might first have to be translated before being used effectively in countries.

4. Defining the data specifications and the ground reference

Addressing public health issues requires for any data to be of quality across the following 6 dimensions[5] as defined by the Data Management Association International (DAMA) [17]:

1. Completeness[6]
2. Uniqueness[7]
3. Timeliness[8]
4. Validity[9]
5. Accuracy[10]
6. Consistency[11]

Data quality is being assessed and improved based on measurable criteria captured in the data specifications and using remote sensing images and the master lists as ground reference (Table 1).

Data quality dimension	Data specifications	Ground reference	
		Remote sensing images	Master lists
Completeness	X	X	X
Uniqueness	X		X
Timeliness	X	X	X
Validity	X		X
Accuracy	X	X	X
Consistency	X	X	X

Table 1 - Data quality dimensions covered by the data specifications and/or ground reference

The next sections provide more details regarding the content and use of the data specifications and ground reference to assess and improve the quality of data as part of the implementation of the geospatial data management cycle.

All this information is then captured in the geospatial data management guidelines used by the Ministry of Health to ensure the quality of the data its uses to support planning and decision making[12].

[5] A Data Quality (DQ) Dimension is a recognised term used by data management professionals to describe a characteristic of data that can be measured or assessed against defined standards in order to determine the quality of data.
[6] State of having all the necessary or appropriate parts (no data gaps).
[7] Quality of being the only one of its kind (no duplicates).
[8] Degree to which data represent reality from the required point in time
[9] Conformity to the syntax (format, type, range) of its definition.
[10] Degree to which data correctly describes the "real world" feature or event being described.
[11] Absence of apparent contradictions.
[12] https://www.healthgeolab.net/KNOW_REP/MMR_MOHS_Geo-Guidelines.pdf

4.1 Data specifications

Data specifications are measurable criteria defined to ensure the same minimum level of quality for the data used by an organization [1]. These criteria are defined across all the 6 intrinsic dimensions of data quality and this separately for georeferenced master lists, geospatial data and statistical data (generic example in Annex 1).

When it comes to geospatial data for example, these criteria should at least cover the following (example for Myanmar in Annex 2):

1. Validity:
   a. Geographic coordinate system and map projection.
   b. Geographic extent of the area being covered.
   c. Language(s) included in the data.
   d. File format(s) for sharing data.
   e. Metadata standard used to document the data.
2. Accuracy:
   a. Scale (vector/raster layers).
   b. Spatial resolution (raster layers).
   c. Positional accuracy (vector/raster layers).
   d. Positional accuracy (GPS readings).
   e. Precision level.
   f. Indication of the remote sensing images used as ground reference.
3. Timeliness:
   a. Period for which the data is being considered as relevant.
4. Consistency
   a. Indication of the classification table and/or associated master list with which the data should be consistent

When it comes to completeness and uniqueness:

1. For vector format geospatial data associated to a master list
   a. Completeness:
      i. Records: When applicable, the layer contains all the active records included in the corresponding master list.
      ii. Data elements: When applicable, a value is available for all the data elements as included in the corresponding master list.
   b. Uniqueness:
      i. No duplicate records using the corresponding master list as ground reference.
2. For other geospatial data
   a. Completeness: All the features existing in the reality are included in the dataset (e.g. road network) or area completely covered (E.g. DEM)
   b. Uniqueness: No duplicated geographic objects

Among the above-mentioned criteria, it is important to provide more information regarding the concepts of geographic coordinate system, projections, scale, accuracy, resolution, and precision. This is done in the following sections.

The concept of metadata is itself being covered in volume 2.5 of the Health GeoLab guidance document series [6]. It is crucial to decide on the metadata standard as well as develop the metadata profile to be specified in the data specifications and to be in the position to collect the necessary information to fill the profile during the implementation of the next steps in the geospatial data management cycle.

4.1.1 Geographic coordinate system and map projection

The choice of the geographic coordinate system and, when it applies, map projection depends on the initial intended use of the geospatial data and resulting maps.

The geographic coordinate system is a system in which geospatial data is defined by a 3-D surface and measured in latitude and longitude.[13] In other words, such system is a model which tries to be as close as possible to the shape of the earth. This model is principally defined by two elements, namely:

1. The spheroid: A three-dimensional shape obtained by rotating an ellipse about its minor axis, with dimensions that either approximate the earth as a whole, or with a part that approximates the corresponding portion of the geoid [13]
2. The datum: The reference specifications of a measurement system, usually a system of coordinate positions on a surface (a horizontal datum) or heights above or below a surface (a vertical datum).[12] In other words, the datum defines the position of the spheroid relative to the center of the earth.

The most widely used geographic coordinate system nowadays is the World Geodetic System 1984 (WGS 84). The one designed by the U.S. Department of Defense is the one reported in Annexes 1 and 2.

A map projection is a method by which the curved surface of the earth is portrayed on a flat surface.[12]

Map projections can be classified according to their basic types or projection techniques.[14]

Four main types of projection exist, each of them having a particular purpose as it preserves a particular relationship or characteristic on the map. These types are as follows:

- Equal-Area: Conserves the size of a geographical feature[15],
- Conformal: Conserves the shape of geographical features,
- Equidistant: Conserves the distance between two geographical features, and
- True Direction: Conserves the direction between two geographical features.

It is important to note that a map cannot be both equal-area or conformal – it can only be one or the other, or neither.

The projection technique describes how an imaginary piece of paper (which will become the map) is laid on the Earth to obtain locations. These techniques are
(Figure 1):

- Cylindrical: the imaginary 'piece of paper' is rolled into a cylinder, this is usually used over Equatorial areas or for World Maps;
- Conical: the imaginary 'piece of paper' is rolled into a cone, this is usually used in mid-latitude areas (approximately 20° – 60° North and South); and
- Azimuthal: the imaginary 'piece of paper' is flat, this is usually used over Polar areas.

[13] http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/What_are_geographic_coordinate_systems/003r00000006000000/
[14] http://www.icsm.gov.au/mapping/about_projections.html#types
[15] Naturally and artificially-created features on the earth.

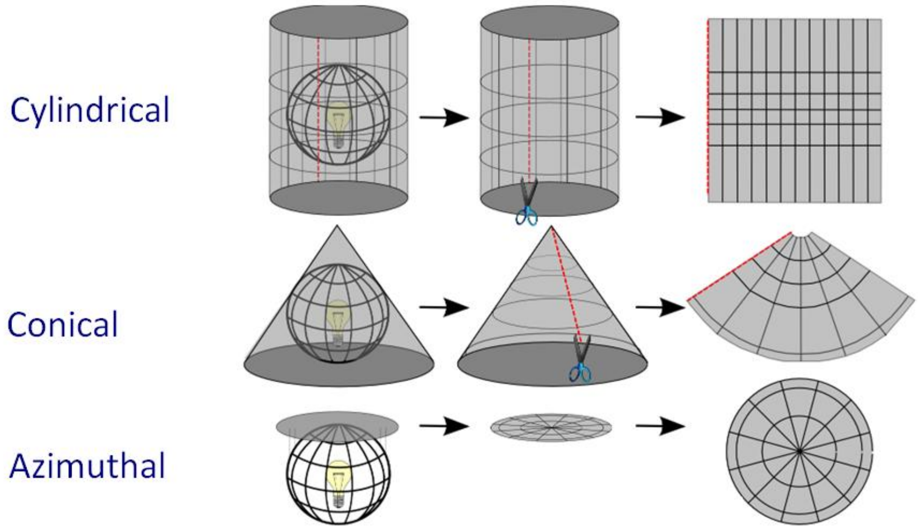


Figure 1 - Basic projection techniques

Each map projection can be described as a combination of these two classifications (Examples in Table 2).

Map Projection	Technique	Type
Equirectangular	Cylindrical	Equidistant
Lambert cylindrical equal-area	Cylindrical	Equal-area
Universal Transverse Mercator (UTM)	Cylindrical	Conformal
Robinson	Pseudocylindrical	Compromise (neither equal-area nor conformal). Generally used to create global maps

Table 2 - Example of map projections with their corresponding type and property

It is important to mention here that un-projected geospatial data, meaning that they are stored only using a geographic coordinate system (example in Annexes 1 and 2), are actually projected on the fly when presented in a GIS software. In this case, a 'pseudo-Plate Carree' projection is being used to treat the coordinate values as if they are linear and therefore just displayed like on a scatter plot.

4.1.2 Scale, resolution, positional accuracy, and precision

The following definitions are considered in the context of the present guidance:

- Scale: The ratio or relationship between a distance or area on a map and the corresponding distance or area on the ground, commonly expressed as a fraction or ratio. A map scale of 1/100,000 or 1:100,000 means that one unit of measure on the map equals 100,000 of the same unit on the earth.
- Resolution: The dimensions of the cells or pixels in a raster format layer.
- Positional accuracy: Quantifiable value that represents the positional difference between a geospatial layer and reality.

- Precision: The number of significant digits used to store numbers, particularly coordinate values. Precision measures exactness.

Accuracy and resolution are both a function of the scale at which geospatial data or a map was created.

The expected positional accuracy is defined by cartographic practices such as the mapping standards of the United States Geological Survey[16, 17] (Table 3).

For example, if one were to take a picture of the entire Earth from space (Small scale map) and trace a line along the Amazon River as it appeared in the picture, the line would generally follow the path of the river but will most probably fall far away from its real course. In contrast, if the same person zoomed into South America before taking the picture and then traced the line along the Amazon River, the line would most likely more precisely follow the path of the river and be more accurate. This line would include less horizontal error and therefore be more accurate than on the first map.

Classification	Map examples	Range examples	Expected positional accuracy (m)
Large scale	Village, town or sub national level map	1:1 - 1:10,000	0 - 8
		1:50,000 - 1:100,000	26 - 52
		1:250,000 - 1:500,000	130 - 259
Medium scale	Country map	1:750,000 - 1:1,000,000	389 - 518
		1:1,500,000 - 1:2,000,000	777 - 1,036
Small scale	World map	1:5,000,000 - 1:10,000,000	2,591 - 5,182
		1:25,000,000 -1:50,000,000	12,954 - 25,908

Table 3 - Relation between scale and expected positional accuracy

From a geospatial data generation or collection perspective, the use of Table 3, for example, means that a project conducted at the scale of 1:100,000 (1 cm on the map is equivalent to 1 km on the ground) should be based on geospatial data presenting a positional accuracy, or a maximum positional error, of 52 meters.

While the cost associated to the generation of polygon or line type geospatial data increases drastically when passing from small to large scale maps, the high accuracy offered by today's Global Navigation Satellite Systems (GNSS) is such that it is possible to collect geographic coordinates presenting a positional accuracy close to the meter and therefore generate point type geospatial data that can be used across the whole range of scales presented in Table 3 [5]. This is the reason why the positional accuracy for GPS readings reported in Annexes 1 and 2 is lower than the one for vector layers.

The relation between scale and the expected resolution of a raster layer has been defined by Waldo Tobler in 1987 [18] through the following rule: divide the denominator of the map scale by 1,000 to get the detectable size in meters. The resolution is one half of this amount. Table 4 presents the application of this rule for the ranges of scales reported in Table 3.

[16] https://pubs.usgs.gov/fs/1999/0171/report.pdf

[17] 90 per cent of all measurable points must be within 1/30th of an inch for maps at a scale of 1:20,000 or larger, and 1/50th of an inch for maps at scales smaller than 1:20,000

The values reported in Table 4 can also be used to define the minimum resolution for the imagery to be used as ground reference or to generate geospatial data for a particular scale of work.

As we can see in Table 3 and 4, the expected positional accuracy and the raster resolution are very close to each other for a given scale, confirming that the size of the cells in a raster layer should not extend beyond the maximum error (accuracy) you would expect for that particular scale.

Due to its specific nature, it is important to remember that accuracy and resolution will have to be considered together when looking at the potential shift existing between the reality and its representation in a raster format layer.

Scale Range	**Raster resolution (m)**
1:1 - 1:10,000	0.0005- 5
1:50,000 - 1:100,000	25- 50
1:250,000 - 1:500,000	125- 250
1:750,000 - 1:1,000,000	375- 500
1:1,500,000 - 1:2,000,000	750- 1,000
1:5,000,000 - 1:10,000,000	2,500- 5,000
1:25,000,000 -1:50,000,000	12,500- 25,000

Table 4 - Relation between scale and the corresponding raster resolution

Last but not least, the data currently available might not necessarily present the appropriate resolution and generating such resolution might be too costly. This is the reason why, in Annexes 1 and 2 the spatial resolution for raster layers has been fixed to 90 m while it should normally have been of 50 m at 1:100,000 scale (Table 4).

Precision, as defined here, directly depends on the number of digits being captured by geographic coordinates. This relation is illustrated in Table 5 when considering a geographic coordinate taken at the level of the equator using the WGS 84 Geographic coordinate system[18]. At that level, the circumference of the Earth is equal to about 20,075 km. Each degree along the equator is then equivalent to 11,320 meters (40,075 km / 360°).[19]

Number of captured digits	Example (Longitude)	Maximum potential error (m)	Precision level
1	120.9	11,132	
2	120.93	1,113	Kilometre
3	120.037	111	Hectometre
4	120.9376	11	Decametre
5	120.93761	1	Metre

Table 5 - Relation between the numbers of decimal digits and the corresponding precision level for geographic coordinates taken at the level of the equator

[18] Coordinates are expressed in different units once a map projection is being used. The units in question should be converted into the metric system before being able to determine when the precision level reaches the meter

[19] The meridional circumference - from pole-to-pole - is of 40,009 km and the circumference of the Earth changes as you move towards the poles, therefore having an impact on the measurements presented in Table 5.

As per Table 5, a precision level down to the meter is reached when using coordinates with 5 decimal digits. Such precision level is recommended by the AeHIN GIS Lab both during data collection in the field (use of GNSS enabled devices) and when generating or extracting vector format geospatial data (precision level of vertices[20]).

To summarize:

- The purpose behind the use of geospatial data will guide the choice of a specific scale of work (Table 3);
- This scale will directly influence the positional accuracy (Table 3) and spatial resolution (Table 4) that should be used when compiling, collecting, or extracting geospatial data;
- The highest accuracy possible should be sought when using GNSS-enabled devices to allow for the largest use possible of the resulting data; and
- A precision level down to the meter (5 digits in decimal degrees) is being recommended.

4.2 The ground reference

Ground reference can be defined as the closed reference to the reality against which data quality can be measured.

In the context of the HGL guidance, the ground reference is provided by the following:

1. High resolution orthorectified remote sensing imagery (e.g. satellite and orthophoto images).
2. Georeferenced master lists which can be defined as the unique, authoritative, officially curated by the mandated agency, complete, up-to-date and uniquely coded list of all the active (and past active) records for given types of geographical feature/object (e.g. health facilities, administrative divisions, villages) [2].

Both elements are necessary to evaluate the completeness, uniqueness, timeliness, accuracy, and consistency of geospatial data (Table 1).

The quality dimensions these elements will allow to measure depend on the format of geospatial data (vector, raster) as well as the way the geographical feature is being represented once captured as a geographic object (point, line, polygon) when it comes to vector format geospatial data as reported in Annexes 1 and 2.

Please refer to HGL's guidance 2.3 [3] to learn more on how to use high resolution imagery and master lists to assess the quality of geospatial data.

The next sections provide more details on what needs to be considered when choosing an appropriate imagery or establishing master lists.

4.2.1 Remote sensing imagery

Remote sensing (RS) is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or

[20] One of a set of ordered x,y coordinate pairs that defines the shape of a line or polygon feature.

aircraft)[21] in contrast to on-site observation. As such, RS images are also known as satellite (spaceborne) images or aerial (airborne) photographs.

With the advancement in technology, more and more RS images are being collected for various uses. These images can be accessed for free or for a fee through different channels, including but not limited to:

1. Institutions giving free access to native images or imagery through a Web Mapping Service (WMS)[22] or a Web Map Tile Service (WMTS)[23]
2. Companies offering commercial imagery. Depending on the option you purchase, you will either have access to the native images or be able to access them through a WMS or WMTS.
3. Online GIS platform like ArcGIS Online, online viewing platform such as Google Maps or Bing Maps, or online extraction platforms like those used by OpenStreetMap (OSM) which use satellite imagery as a basemap.

Please refer to HGL guidance 2.4.1 [4] for more details regarding these different sources of imagery.

4.2.2 Georeferenced master lists, registry, and common geo-registry

Georeferenced master lists are central to the Health Information System (HIS) as they represent the reference ensuring data consistency among data sources. At the same time, master lists:

- Provide the denominator for data collection (including for sampling), monitoring, and evaluation.
- Represent one of the pillars to geo-enable the HIS [19].
- Form the reference to assess the quality of geospatial data as reported in Annex 3.
- Minimize duplicate reporting and improve transparency.
- Support better analysis and synthesis of data and consequently, decision making as well as health system functioning.
- Serve as the official source of geographic coordinates for point type geographic objects when this information is being captured.

Establishing and maintaining a good quality georeferenced master list requires for a lot of elements to be in place (example for the health facility master list in [20] or community health workers in [21].

In the meantime, it is important to provide here the list of elements that characterizes a master list to be used as ground reference. A master list should:

1. Cover the core set of fields that would allow uniquely identifying, locating and, when appropriate, contacting each active record in the list.
2. Originate from the governmental entity officially mandated to develop and maintain such master list.
3. Be complete and up-to-date.
4. Contain an official and unique Identifier (ID) for each of the records.

[21] https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used?qt-news_science_products=0#qt-news_science_products

[22] A standard protocol for serving georeferenced map images over the Internet

[23] A standard protocol for serving pre-rendered or run-time computed georeferenced map tiles over the Internet

5. Make the link with other master lists when appropriate (for example the name and unique code of administrative divisions to be included in the health facility master list).

A master list should at least be available for the geographical features identified as being core for public health (health facilities, communities/settlements (city, towns, villages, hamlets), administrative and reporting divisions) [2].

It is also important to mention here that for health facilities, the concept of Health Facility Master List (HFML) has evolved over the past decades to now cover two components under the Master Facility List (MFL) concept [22, 23]:

1. The signature domain - a set of information that permit the unique identification and location as well as the capture of the contact information of a given health facility.
2. The service domain - attributes used to capture the availability of services and capacity of each health facility.

While the information collected as part of the service domain component is important, the signature domain represent the component that contains all the information necessary from a geospatial data management perspective.

In addition to that, the information included in the service domain might be under the responsibility of different entities/programs within the Ministry of Health, making it sometime difficult to be compiled and regularly updated. Integrating this information with the signature domain might also result in having it being stored in different databases and therefore generate different versions.

In view of the above, the Health GeoLab recommends to:

- Consider the health facility master list (HFML) as being only composed of the signature domain data elements [22] and to have it placed under the responsibility of one single entity within the MOH (generally the one in charge of the Health Information System).
- Keep the information contained in the service domain in separated databases under the responsibility of the program in charge of their development and maintenance. The link between these databases and the HFML (signature domain) is ensured through the unique identifier attached to each facility.

The concept of registry does itself refer to the IT solution that allows storing, managing, validating, updating and sharing a master list while the master list is itself the standardized data stored in that solution [Adapted from 20]. An Example of registry is the online platform developed by the Department of Health of the Philippines to manage its health facilities master list[24].

The need to consider several types of geographical features when implementing public health programs and the relationship between these geographical features resulted in the development of a new type of registry - a common geo-registry - to host, manage, regularly update and share the master lists as well as associated hierarchies and geospatial data for the geographic objects core to development in general and public health in particular [24].

[24] https://nhfr.doh.gov.ph/Home

References

[1] Ebener S. (2016): Guidance for the management and use of geospatial data and technologies in health. Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management chain. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part1.pdf [Accessed 19 February 2024]

[2] Ebener S. (2016): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.1 Documenting the process and defining the data needs. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_1.pdf [Accessed 19 February 2024]

[3] Pantanilla I., Ebener S., Mercado C.E., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.3 Compiling existing data and identifying gaps. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_3.pdf [Accessed 19 February 2024]

[4] Ebener S. (2021): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.4 Creating geospatial data - 2.4.1 Extracting vector format geospatial data from basemaps. Health GeoLab document: https://healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_4_1.pdf [Accessed 19 February 2024]

[5] Ebener S., Maude R.J., Gault P. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.4 Creating geospatial data - 2.4.2 Collecting data in the field. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_4_2.pdf [Accessed 19 February 2024]

[6] Ebener S., Pantanilla I., Mercado C.E., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.5 Cleaning, validating, and documenting the data - 2.5.1 Documenting the data using a metadata profile. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_5_1.pdf [Accessed 19 February 2024]

[7] Ebener S., Pantanilla I. (2019): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.5 Cleaning, validating, and documenting the data - 2.5.2 Using advanced Microsoft Excel functions. Health GeoLab document: https://healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_5_2.pdf [Accessed 19 February 2024]

[8] Pantanilla I., Ebener S., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.1 Creating good thematic maps using desktop GIS software. Health GeoLab document:

https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_6_1.pdf [Accessed 19 February 2024]

[9] Ebener S. (under preparation): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.2 Using thematic maps for decision making. Health GeoLab document.

[10] Ebener S. (under preparation): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.3 Developing and implementing the appropriate data policy. Health GeoLab document.

[11] Ebener S., Stokes C., Terborgh C.J., & Lance K. (2006): Moving Towards Global Spatial Data Infrastructure: Improving Data Integration at the Global Level, 9th Global Spatial Data Infrastructure (GSDI) Conference, Santiago de Chile, 3-11 November 2006: https://healthgeolab.net/DOCUMENTS/Global_Data_Integration_final.pdf [Accessed 19 February 2024]

[12] Department of Health of the Philippines (2014): GIS/GPS guidelines and standards. Internal document developed with the support of the WHO Country Office in the Philippines.

[13] Esri GIS data dictionary: https://support.esri.com/en-us/gis-dictionary [Accessed 19 February 2024]

[14] UNAIDS terminology guidelines: http://www.unaids.org/sites/default/files/media_asset/2015_terminology_guidelines_en.pdf [Accessed 19 February 2024]

[15] International Epidemiological Association (2014): A dictionary of epidemiology - Sixth Edition edited for the International Epidemiological Association by Miquel Porta. Oxford University Press: http://www.irea.ir/files/site1/pages/dictionary.pdf [Accessed 19 February 2024]

[16] CDC Malaria glossary: https://www.cdc.gov/malaria/glossary.html [Accessed 19 February 2024]

[17] DAMA UK Working Group on "Data Quality Dimensions" (2013): The six primary dimensions for data quality assessment- Defining Data Quality Dimensions https://www.sbctc.edu/resources/documents/colleges-staff/commissions-councils/dgc/data-quality-deminsions.pdf [Accessed 19 February 2024]

[18] Tobler W. (1987): Measuring Spatial Resolution, Proceedings, Land Resources Information Systems Conference, Beijing, pp. 12-16.

[19] Roth S., Landry M., Ebener S., Marcelo A., Kijsanayotin B., & Parry J. (2016): The Geography of Universal Health Coverage. ADB brief No. 55, April 2016: https://www.adb.org/publications/geography-universal-health-coverage [Accessed 19 February 2024]

[20] WHO (2018): Master Facility List Resource Package: guidance for countries wanting to strengthen their Master Facility List. https://www.who.int/publications/i/item/-9789241513302 [Accessed 19 February 2024]

[21] Clinton Health Access Initiative, Community Health Impact Coalition, Global Fund, Living Goods, health GeoLab Collaborative, UNICEF (2021): Implementation support guide: Development of a National Georeferenced Community Health Worker Master List Hosted in a Registry. https://www.unicef.org/documents/implementation-support-guide-development-national-georeferenced-community-health-worker [Accessed 19 February 2024]

[22] Health Facility Assessment Technical Working Group (2007): The Signature Domain and Geographic Coordinates: A Standardized Approach for Uniquely Identifying a Health Facility. Measure Evaluation publication WP-07-91: http://www.cpc.unc.edu/measure/resources/publications/wp-07-91/at_download/document [Accessed 19 February 2024]

[23] Measure Evaluation (2015): Guidelines for data management standards in routine health information systems. Measure Evaluation publication WP-07-91MS-15-99: http://pdf.usaid.gov/pdf_docs/PA00KB8N.pdf [Accessed 19 February 2024]

[24] Health Geolab (2022): HIS geo-enabling: Guidance on the establishment of a common geo-registry for the simultaneous hosting, maintenance, update and sharing of lists as well as associated hierarchies and spatial data (version 2.0): https://healthgeolab.net/DOCUMENTS/Guidance_Common_Geo-registry_Ve2.pdf [Accessed 19 February 2024]

Annex 1 - Generic data specifications for georeferenced master lists, geospatial data and statistical data

Georeferenced master lists

<table>
<tr><th colspan="2">Validity</th></tr>
<tr><td>Language</td><td>English and local language (Unicode)</td></tr>
<tr><td>File format</td><td>MS Excel</td></tr>
<tr><td>Data dictionary</td><td>Covers the minimum set of data elements included in the corresponding master list data dictionary</td></tr>
<tr><td>Metadata</td><td>Covers the minimum set of fields included in the corresponding master list template</td></tr>
<tr><td>Geographic coordinate system</td><td>Geographic Coordinate System: GCS_WGS_1984
Angular Unit: Degree (0.0174532925199433)
Prime Meridian: Greenwich (0.0)
Datum: D_WGS_1984
Spheroid: WGS_1984
Semimajor Axis: 6378137.0
Semiminor Axis: 6356752.314245179
Inverse Flattening: 298.257223563</td></tr>
<tr><th colspan="2">Accuracy (geographic coordinates)</th></tr>
<tr><td>Positional accuracy</td><td>15 meters</td></tr>
<tr><td>Precision</td><td>meter (5 digits)</td></tr>
<tr><th colspan="2">Timeliness</th></tr>
<tr><td>Temporal validity</td><td>Data older than 1 year should be avoided</td></tr>
<tr><th colspan="2">Completeness</th></tr>
<tr><td>Records</td><td>All the currently active records are included in the list</td></tr>
<tr><td>Data element</td><td>A value is available for all the data elements included in the data dictionary</td></tr>
<tr><th colspan="2">Uniqueness</th></tr>
<tr><td>Duplicates</td><td>No duplicate records</td></tr>
<tr><th colspan="2">Consistency</th></tr>
<tr><td>Data elements based on a classification table or associated master list</td><td>When applicable, data elements values are consistent with the options included in the corresponding classification table (e.g. health facility type) or associated master list (e.g. administrative unit names)</td></tr>
</table>

Geospatial data

Validity	
Language	English and local language (Unicode)
File format	● Vector: shape file ● Raster: Esri GRID
Attribute table (vector layers)	When applicable, covers all the data elements included in the corresponding master list
Metadata	AIM-Net profile based on ISO 19115: Geographic information – Metadata
Geographic coordinate system	Geographic Coordinate System: GCS_WGS_1984 ● Angular Unit: Degree (0.0174532925199433) ● Prime Meridian: Greenwich (0.0) ● Datum: D_WGS_1984 ● Spheroid: WGS_1984 o Semimajor Axis: 6378137.0 o Semiminor Axis: 6356752.314245179 o Inverse Flattening: 298.257223563
Geographic extent	To be defined based on the program or intervention
Accuracy (geographic coordinates)	
Scale (vector/raster layers)	1:100,000
Spatial resolution (raster layers):	90 m
Positional accuracy (vector/raster layers):	50 meters
Positional accuracy (GNSS reading)	15 meters
Precision (GNSS reading)	meter (5 digits)
Satellite imagery mosaic used as ground reference	Google Map, Bing Map, Esri imagery
Timeliness	
Temporal validity	Data older than 1 year should be avoided
Completeness	
Records (vector layers)	When applicable, the layer contains all the active records included in the corresponding master list
Data element (vector layers)	When applicable, a value is available for all the data elements as included in the corresponding master list
Uniqueness	
Duplicates (vector layers)	No duplicate records using the corresponding master list as ground reference
Consistency	
Data elements based on a classification table or associated master list	Values are consistent with the options included in the corresponding classification table (e.g. health facility type) or associated master list (e.g. administrative unit names)

Statistical data

Validity	
Language	English and local language (Unicode)
File format	MS Excel
Data dictionary	Covers all the data elements included in the file
Metadata	Covers the minimum set of fields included in the defined metadata profile
Timeliness	
Temporal validity	Data older than 1 year should be avoided
Completeness	
Records	Statistics available for all the active records included in the corresponding master list
Data element	A value is available for all the statistical indicators included in the data dictionary
Uniqueness	
Duplicates	No duplicate records using the corresponding master list as ground reference
Consistency	
Values	Values are captured according to the indicator reference sheet (format, unit)

Annex 2 - Example of geospatial data specifications - Myanmar

Validity:

Geographic coordinate system

- Geographic Coordinate System: GCS_WGS_1984
    - Angular Unit: Degree (0.0174532925199433)
    - Prime Meridian: Greenwich (0.0)
    - Datum: D_WGS_1984
    - Spheroid: WGS_1984
        - Semimajor Axis: 6378137.0
        - Semiminor Axis: 6356752.314245179
        - Inverse Flattening: 298.257223563

Geographic extent (Decimal degrees)

- West Boundary: 92.1° E
- East Boundary: 101.2° E
- South Boundary: 9.6° N
- North Boundary: 28.6° N

Language:

- English and Myanmar language (unicode)

File format:

- Vector: shape file
- Raster: Esri GRID

Metadata standard:

- ISO 19115: Geographic information - Metadata

Source

- Priority should be given to geospatial data generated and maintained by official governmental entities;

Accuracy:

- Scale (vector/raster layers): 1:100,000
- Spatial resolution (raster layers): 90 m
- Positional accuracy (vector/raster layers): 50 meters
- Positional accuracy (GPS reading): 15 meters
- Precision: meter (5 digits)
- Mosaic of satellite images used to check for positional accuracy and consistency: Google Map

Timeliness:

- The most recent available data should be used
- Data older than 5 years should be avoided

Completeness, uniqueness and consistency:

- When applicable, the content of the attribute table of the GIS layer should match the content of the corresponding master list in terms of completeness, uniqueness and consistency (e.g. spelling, codes)

Annex 3 - Quality dimensions measured using satellite images and/or master lists.

Vector format data

	Use of satellite/orthophoto images for assessing:	Use of a master list for assessing:
Points type features		
Mobile objects		
People	NA	(Completeness, Uniqueness, Timeliness, Consistency)[3]
Patients	NA	(Completeness, Uniqueness, Timeliness, Consistency)[3]
Health personnel	NA	(Completeness, Uniqueness, Timeliness, Consistency)[3]
Ambulances	NA	(Completeness, Uniqueness, Timeliness, Consistency)[3]
Fixed objects		
Villages	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency[5]	Completeness, Uniqueness, Timeliness, Validity, Consistency
Health facilities	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency[5]	Completeness, Uniqueness, Timeliness, Validity, Consistency
Household	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency[5]	Completeness, Uniqueness, Timeliness, Validity, Consistency
Line type features		
Roads	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency[5]	(Completeness, Uniqueness, Timeliness, Validity)[4]
Rivers	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency[5]	(Completeness, Uniqueness, Timeliness, Validity)[4]
Polygon type features		
Administrative divisions	Accuracy[2], Consistency	Completeness, Uniqueness, Timeliness, Validity, Consistency
Water bodies	Completeness, Uniqueness, Timeliness[1], Accuracy, Consistency	Completeness, Uniqueness, Timeliness, Validity, Consistency

Raster format data (Continuous features)

Digitial Elevation Model	Completeness[6], Accuracy, Consistency[5]	NA
Landcover	Completeness[6], Timeliness[1], Accuracy, Consistency[5]	NA
Population distribution	Completeness[6], Accuracy, Consistency[5]	NA

[1] Will depend on the temporal stamp of the satellite image
[2] When the administrative division boundary follows a natural geographical feature that can be identified on the satellite image (a river for example)
[3] While such master list is key in the Health Information System (HIS) context, it does not necessarily convert into a geospatial dataset due to the mobile nature of the geographical feature
[4] Only if a master list is available as it is difficult to develop and maintain such a master list for this geographical feature
[5] In the sense that the different geospatial datasets have to be geographically consistent among themselves
[6] In the sense that all the land areas reported on the satellite image should be covered by the layer in question