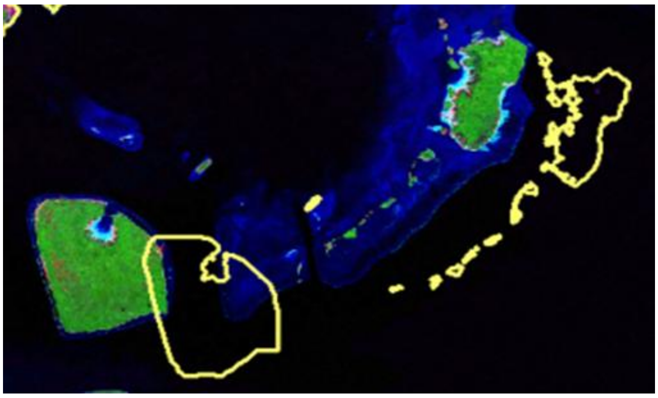
Guidance for the collection and use of geospatial data in health

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

Version 1.1 (01.05.2017)



In collaboration and with the support of:

Author

Steeve Ebener[1]

Acknowledgements

Susann Roth[2]
Mark Landry[3]
Izay Pantanilla[1]

1. AeHIN GIS Lab, Manila, Philippines
2. Asian Development Bank, Manila, Philippines
3. World Health Organization, South-East Asia Regional Office, Delhi, India

Table of Contents

Purpose and audience

The purpose of the AeHIN GIS Lab guidance documents is to inform concerned practitioners about the key elements they need to be aware of when it comes to collecting and using geospatial data in health and guide them through the process to be followed in order for the data in question to be appropriate and of quality.

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

Please note that some of the sections in this guide require basic understanding of concepts pertaining to geospatial data and GIS.

Abbreviations

ADB	Asia 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
IEA	International Epidemiological Association
UNAIDS	United Nations Programme on HIV/AIDS
WHO	World Health Organization
WGS	World Geodetic System

1. Background

The Asian eHealth Information Network (AeHIN) GIS Lab[1] has been established with the support of ADB and WHO to strengthen the technical capacity of the health sector in countries (government and partners) to fully benefit from the power of geography and geospatial technologies through the geo-enabling of their Health Information System (HIS).

Maps produced through the use of a GIS represent key tools for decision makers not only by providing them with key information to investigate, understand, and communicate health issues but also to analyze where, why, and how resources can be allocated to improve the health of people and places.

However, the quality of the information contained in these maps and the analysis which is made out of them is as good as the quality of the geospatial data[2] which is being used to generate them.

Unfortunately, the collection and use of geospatial data often happens without having the necessary processes and protocols in place. This results to geospatial data and products which are not of sufficient quality for their intended use.

The guidance documents generated by the AeHIN GIS Lab in collaboration and with the support of partners are therefore meant to provide such processes and protocols to the concerned practitioners in order for them to be in the position to improve the quality of the geospatial data and products being generated in the health sector.

The complete series of these documents is organized as follow:

- Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle;
- Part 2 - Implementing the geospatial data management cycle:
    - 2.1 Documenting the process and defining the data needs;
    - 2.2 Defining the vocabulary, the data set specifications, and the ground reference (the present document);
    - 2.3 Compiling existing data and identifying gaps;
    - 2.4 Creating geospatial data
        - 2.4.1 Extracting data from other sources
        - 2.4.2 Collecting data in the field;
    - 2.5 Cleaning, validating, and documenting the data; and
    - 2.6 Distributing, using, and updating the data.

[1] http://aehin.org/Resources/GISLab.aspx
[2] Any data that can be mapped

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 through the use of a common dictionary,
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 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 [3], 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 [4] 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 vocabulary

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

Two types of dictionaries can be differentiated when dealing with geospatial data, namely those covering:

- GIS related terms, and
- Thematic terms.

While different GIS dictionaries do exist, the one included in the GIS.com wiki [3] remains the most comprehensive online resource and therefore the one being recommended by the AeHIN GIS Lab [4]. Another option is the GIS dictionary maintained by Esri and which is accessible either online [5] or in print.

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

- The UNAIDS terminology guideline when it comes to HIV/AIDS [6],
- The IEA dictionary of epidemiology [7], and
- CDC's Malaria glossary [8].

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.

[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.

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

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

Last but not least, language can be an issue, meaning that global or regional dictionaries 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 and covering the following 6 dimensions[5] as defined by the Data Management Association International (DAMA) [9]:

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

The standards against which geospatial data is being assessed to measure these dimensions are being captured by the data specifications and the ground references as presented in Table 1.

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

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

The standards in question are being presented in the following sections.

[5] A Data Quality (DQ) Dimension is a recognised term used by data management professionals to describe a feature of data that can be measured or assessed against defined standards in order to determine the quality of data.
[6] The proportion of stored data against the potential of "100% complete"
[7] No thing will be recorded more than once based upon how that thing is identified.
[8] The degree to which data represent reality from the required point in time.
[9] Data are valid if it conforms to the syntax (format, type, range) of its definition.
[10] The degree to which data correctly describes the "real world" object or event being described.
[11] The absence of difference, when comparing two or more representations of a thing against a definition.

4.1 Geospatial data specifications

The geospatial data specifications document contains all the features a geospatial dataset should comply with in order to be considered of quality and fulfil the original purpose and expected outcomes [1].

Such document should at least contain the following information to be considered as standards (example in Annex 1):

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 layers)
    b. Spatial resolution (raster layers)
    c. Positional accuracy (vector layers)
    d. Positional accuracy (GPS readings)
    e. Precision level
3. Timeliness:
    a. Period for which the data is being considered as relevant

To reinforce the notions of completeness, uniqueness, and consistency, the document can also make references to the satellites images and master lists that should be used as ground reference.

Among the above mentioned elements, it is important to provide more information regarding the concepts of projections, accuracy. This is done in the following sections.

The concept of metadata is itself being covered in volume 2.5 of the AehIN GIS Lab guidance document series (currently under preparation). This being said, it is crucial to decide on the metadata standard as well as develop the metadata profile that to be used at the moment of defining 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.[12] 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:

[12] http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/What_are_geographic_coordinate_systems/003r00000006000000/

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 [5]
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 Annex 1.

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.[13]

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

- Equal-Area: Conserves the size of a feature,
- Conformal: Conserves the shape of features,
- Equidistant: Conserves the distance between two features, and
- True Direction: Conserves the direction between two 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://www.icsm.gov.au/mapping/about_projections.html#types

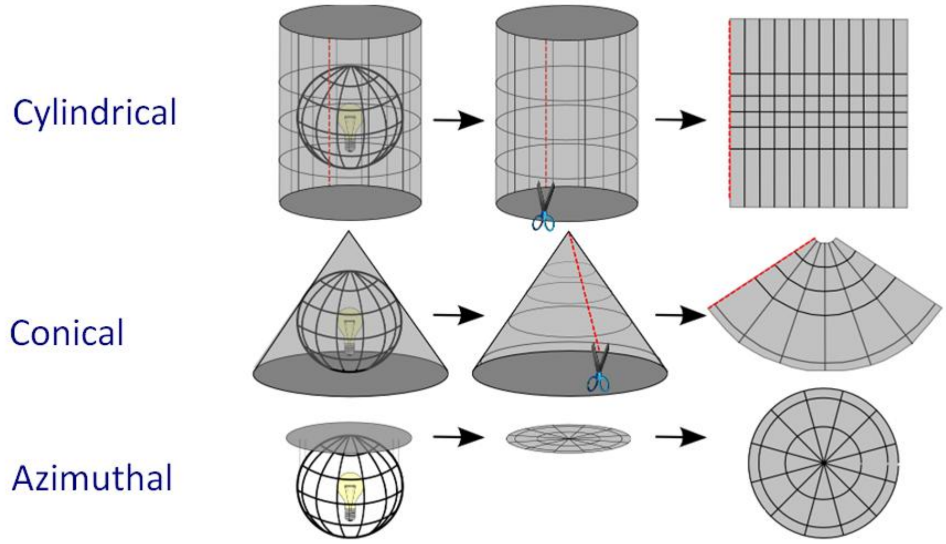


Figure 1 - Basic projection techniques

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 Annex 1), 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, accuracy, resolution, and precision

The following definitions are considered in the context of the present guidance [5]:

- 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.
- Accuracy: The degree to which a measured value conforms to true or accepted values. Accuracy is a measure of correctness.

- Resolution (raster format): The dimensions represented by each cell or pixel in a raster.
- 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[14, 15] (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 [10]. This is the reason why the positional accuracy for GPS readings reported in Annex 1 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 [11] through the following rule: divide the denominator of the map

[14] http://www.colorado.edu/geography/gcraft/notes/error/error_f.html
[15] 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

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.

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 the case of Myanmar (Annex 1), 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[16]. 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°).[17]

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

[16] 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
[17] 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.

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

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[18]).

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, or ground truth, as used in the context of the present guidance refers to two different concepts:

1. The actual location of a given feature on the surface of the Earth. As it is not possible to check all the available geospatial data directly in the field, high resolution satellite or orthophoto images represent the best option.
2. Master list which can be defined as the authoritative, standardized, complete, up-to-date, and uniquely coded list of all active records for a given object [2].

Both of these 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 type of object (point, line, polygon) when it comes to vector format geospatial data as reported in Annex 2.

The next sections provide more details on what needs to be considered when choosing an appropriate ground reference to answer the initial objectives and expected outcomes.

4.2.1 Remote sensing imagery

While the numbers of satellites providing images that can be used as ground reference continue to slowly increase and that more and more countries are having the resources to cover their whole territory through orthophotos, it is important to remember that[19]:

[18] One of a set of ordered x,y coordinate pairs that defines the shape of a line or polygon feature.
[19] http://www.gearthblog.com/blog/archives/2016/07/imaging-satellite-list.html

1. Most of these images are not freely available and come with a cost (the higher the resolution the higher the price in general).
2. The resolution of these images change from one satellite to the other, making them not always appropriate for the scale of work that has been chosen (Table 4).

As an example, and taking the above into account, the Landsat ETM+ (Enhanced Thematic Mapper Plus) mosaic freely available from the Earth Science Data Interface (ESDI) at the Global Land Cover Facility [12] with its 30 meter resolution and 50 accuracy represents a good option to assess the quality dimensions listed in Annex 2 when working at a scale between 1:50,000 and 1:100,000 [Table 3 and 4].

Nowadays, high resolution images are also available at no-cost through online platform such as Google Map [13] or Bing Map [14]. This option requires uploading the data into the platform which represent the following limitation:

1. The volume of data that can be imported remains limited.
2. The format of the data has to be modified in order to allow for the upload.
3. The potential modification can't take place into the platform itself and have to be visually implemented in a GIS.

A good alternative to this option is the possibility to have access to such imagery directly from a GIS software through what is called a web mapping service.[20] While the images are not physically on your computer they can be used in the background to perform the necessary assessment. GIS software such as ArcGIS and QGIS offer this possibility, the former giving access to a larger number of sources than the latter but with a cost, cost that however remains less than when purchasing the images.

4.2.2 Master lists and registries

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 [15]
- Form the reference to assess the quality of geospatial data as reported in Annex 2
- Minimize duplicate reporting and improve transparency
- Support better analysis and synthesis of data and consequently, decision making as well as health system functioning; and
- Serve as the official source of geographic coordinates for point type objects when this information is being captured

Establishing and maintaining a good quality master list requires for a lot of elements to be in place. These topics will be covered as part of another forthcoming publication of the AeHIN GIS Lab.

[20] Standard protocol for serving (over the Internet) georeferenced map images which a map server generates using data from a GIS database.

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; and
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 geographic objects 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 Master Facility List (MFL) has evolved over the past decade to now cover two components [17]:

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 AeHIN GIS Lab recommends to:

- Consider the Master Facility List as being only composed of the signature domain 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 MFL (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 18].

References

[1] Ebener S. (2016): Guidance for the collection and use of geospatial data in health. Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle. AeHIN GIS Lab documents: http://www.aehingislab.net/DOCUMENTS/Guide_GIS_Lab_Part1.pdf [Accessed May 1, 2017]

[2] Ebener S. (2016): Guidance for the collection and use of geospatial data in health. Part 2 - Implementing the geospatial data management cycle: 2.1 Documenting the process and defining the data needs. AeHIN GIS Lab documents: http://www.aehingislab.net/DOCUMENTS/Guide_GIS_Lab_Part2_1.pdf [Accessed May 1, 2017]

[3] 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.

[4] 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. Internal document

[5] Esri GIS data dictionary: http://support.esri.com/other-resources/gis-dictionary [Accessed May 1, 2017]

[6] UNAIDS terminology guidelines: http://www.unaids.org/sites/default/files/media_asset/2015_terminology_guidelines_en.pdf [Accessed May 1, 2017]

[7] 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 May 1, 2017]

[8] CDC Malaria glossary: https://www.cdc.gov/malaria/glossary.html [Accessed May 1, 2017]

[9] DAMA UK Working Group on "Data Quality Dimensions" (2013): The six primary dimensions for data quality assessment- Defining Data Quality Dimensions https://www.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf [Accessed May 1, 2017]

[10] Ebener S., Maude J.R., Gault P. (2016): Guidance for the collection and use of geospatial data in health. Part 2 - Implementing the geospatial data management cycle: 2.4 Creating geospatial data - 2.4.1 Collecting data in the field. AeHIN GIS Lab documents: http://www.aehingislab.net/DOCUMENTS/Guide_GIS_Lab_Part2_4_2.pdf [Accessed May 1, 2017]

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

[12] Earth Science Data Interface (ESDI) at the Global Land Cover Facility http://glcfapp.glcf.umd.edu:8080/esdi/index.jsp [Accessed May 1, 2017]

[13] Google map web site: https://maps.google.com/ [Accessed May 1, 2017]

[14] Bing Map web site: www.bing.com/maps/ [Accessed May 1, 2017]

[15] Roth S., Landry M., Ebener S., Marcelo A., Kijsanayotin B., Parry J. (2016): The Geography of Universal Health Coverage: Why geographic information systems are needed to ensure equitable access to quality health care. ADB brief No. 55, April 2016.

[16] 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 May 1, 2017]

[17] Measure Evalution (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 May 1, 2017]

[18] OpenHIE (2015): OpenHIE Health Facility Registry Implementation Guide: https://docs.google.com/document/d/1KaUPHQRiZ9hQ59Irp56oKayvVZbkWngTYkl2AZdafhw/edit [Accessed May 1, 2017]

Annex 1 - 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

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)

Timeliness:

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

Completeness, uniqueness and consistency:

- Priority should be given to geospatial data generated and maintained by official governmental entities;
- When applicable, the content of the layer should match the official registries in terms of completeness, uniqueness and consistency (spelling, codes);

Annex 1 - Quality dimensions measured through the use of 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 feature that can be identified on the satellite image (a river for example)
[3] While such master 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 object
[4] Only if a master list is available as it is difficult to develop and maintain such a master list for this object
[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