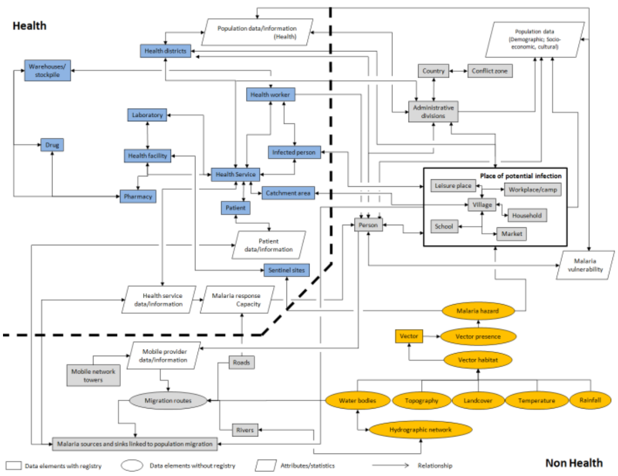
# Guidance for the management and use of geospatial data and technologies in health

# Part 2 - Implementing the geospatial data management cycle: 2.1 Documenting the process and defining the data needs

## Version 1.4 (last update: 19.02.2024)




**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 | Added Annex 2 to describe the process that can be followed to identify core geographic objects and the relationships between them; inclusion of the reference to the new HGL guidance documents | Steeve Ebener |
| 1.4 | 19 February 2024 | Change layout and update of URLs | 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


1. Background .......... 4
2. Introduction .......... 5
3. Documenting the process .......... 5
4. Defining the data needs .......... 6
References .......... 11
Annex 1 – Example of conceptual data model - Malaria elimination .......... 13
Annex 2 – Process aiming at identifying core geographic features, defining them, and capturing the relationships existing between them during a workshop .......... 14


# 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 |
| DOH | Department of Health |
| GIS | Geographic Information System |
| GMS | Greater Mekong Subregion |
| HGL | Health GeoLab |
| WHO | World Health Organization |

# 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 guidances 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 (the present document)
    - 2.2 Defining the terminology, the data specifications, and the ground reference [2]
    - 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 Developing and implementing the appropriate data policy [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

Generating and maintaining good quality geospatial data as well as data products (tables, graphs, and maps) require for proper data management standards, processes, and protocols to be defined and implemented.

In addition to that, these data and data products need to be part of the overall data-information-knowledge continuum in order to support geographically-based decision making, and therefore a more systemic approach to solving public health problems.

The present document's objective is to describe in more details and provide recommendations for the first two steps of the geospatial data management cycle presented in the first volume [1]:

1. Documenting the process, and
2. Defining the data needs.

These two steps are critical: the first one because documenting is what allows replicating the overall process in other geographies and/or context and the second one because it serves as the basis for the implementation of the whole chain.

The present document 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. Documenting the process

The geospatial data management cycle contains 13 different steps.

Implementing these steps might not only take a long time but also involves different persons. The only way to ensure that the process can be replicated is to document each of these steps as precisely as possible.

While some of these steps might just require describing the choices that have been made and why, others might require a lengthier description, in particular:

- Compiling existing data depending on the number of sources that have been considered,
- Collecting or extracting data as well as cleaning and validating it. It is important to include a good description of the method/s that has/have been used as this directly has an influence on the quality of the data, and
- Using the data, as combining Geospatial, and statistical, data coming from different sources, conducting spatial analysis, and/or applying spatial models are not necessarily straight forward exercises.

When it comes to the structure of the document, there is no template to be followed in particular but it is important to ensure that all the key elements from the chain are being captured. A possible approach is to use the steps in the chain as the headers for the different sections of the document in question.

# 4. Defining the data needs

Geo-enabling the HIS in an effective way requires to apply an object-oriented approach in which the management of the geographic features, stored as geographic objects, is separated from the management of the data elements (statistics, information) attach to each these geographic features through time in the information system.

Implementing this approach goes through the development of a data model which is an abstract model that organizes and document the relationships that exists between geographic features and the data elements attached to each of them.

Three main levels of data models can be differentiated[2] as:

1. Conceptual data model which identifies the geographic features with the highest-level relationships existing between them [3];
2. Logical data model which describes add the data elements and primary keys to the conceptual data model to detail the data in as much detail as possible, without regard to how they will be physically implemented in the database; and
3. Physical data model which represents how the model will be implemented in the database.

Each of these models fulfils a different function and contains different items as summarized in Table 1[4].

| Included item | Data model | | |
|---|---|---|---|
| | Conceptual | Logical | Physical |
| Entity Names | X | X | |
| Entity Relationship | X | X | |
| Attributes[5] | | X | |
| Primary Keys[6] | | X | X |
| Foreign Keys[7] | | X | X |
| Table Names[8] | | | X |
| Column Names | | | X |
| Column Data Types[9] | | | X |

Table 1 - Items included in each of the 3 levels of data modelling

The above lead naturally to the identification of the geospatial and statistical data needed to address the objectives of the project, activity, program for which they are being used [1].

Due to their respective functions, the different models are generally being developed in the order presented in the table 1, meaning by starting with the conceptual data model and ending with the physical one. A consultative process among all stakeholders is generally used to develop these models.

[2] http://www.1keydata.com/datawarehousing/data-modeling-levels.html
[3] Refers to any person, place, or thing that data can represent on a map
[4] http://www.1keydata.com/datawarehousing/data-modeling-levels.html
[5] Data or information attached to a particular entity/object.
[6] Key in a relational database that is unique for each record
[7] Set of one or more columns in a table that refers to the primary key in another table
[8] Name of the table containing the information about a particular entity/object
[9] Format in which the data/information is captured for each field in the table (integer, character, date,...)

The types of geographical features[10] considered in Public Health can be separated into four groups when looking at how they would be captured as geographic objects[11] in a GIS (Figure 1), namely those that are:

- fixed and for which the geography can be simplified by a point when converted into a geographic object (examples: household, health facility, village when boundaries are not available,...). The geography of these geographic objects is obtained through their geographic coordinates.
- fixed as well but for which the geography has to be represented by polygons due to their much larger extent (Examples: administrative divisions, health districts,...) or by a line (Example: road, river,....) when converted into a geographic object.
- mobile (Examples: individuals, patients, vehicles,...). The geography of these geographical features would either be obtained by considering them attached to a fixed geographical feature or by simplifying them as a point that would be located through its geographic coordinates (latitude and longitude) taken at a given time.
- Continuous: some elements of our environment are not defined objects per say and not associated with one specific location but are rather distributed spatially. These are better represented using a continuous surface (e.g. terrain, land surface attributes, population distribution)

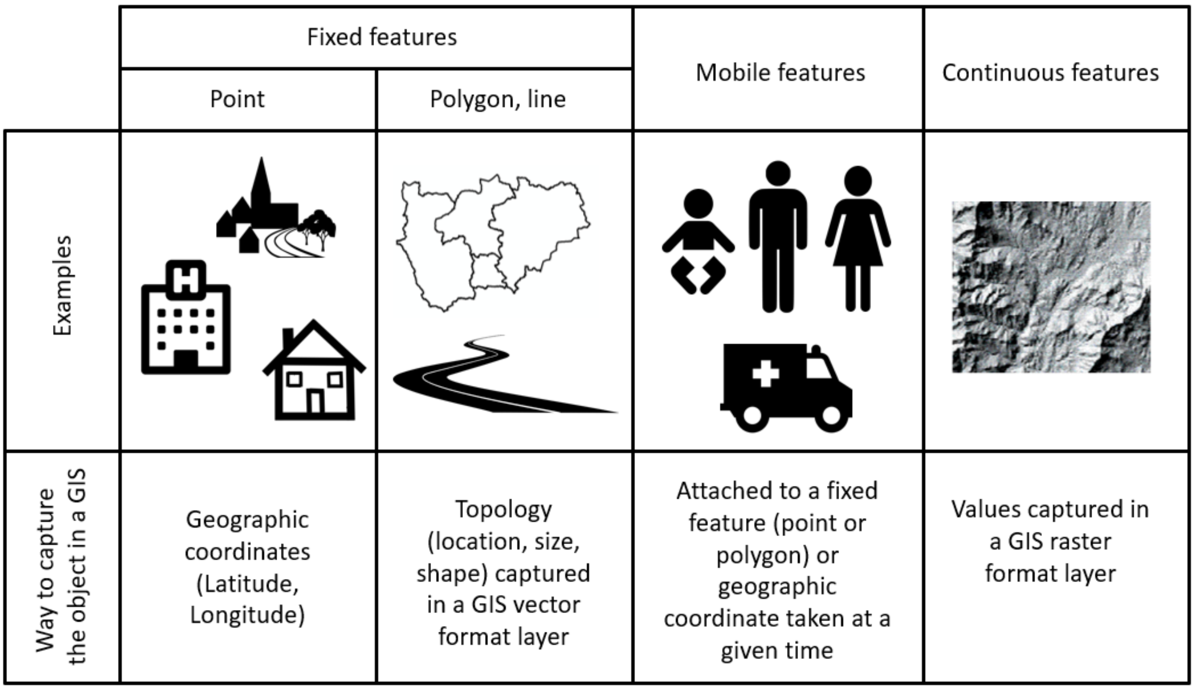



Figure 1 – Geographic features considered in Public Health and the corresponding way to capture them as a geographic object in a GIS

Any health-related data or information collected in the field will be attached to one of the fixed or mobile geographic features reported in Figure 1. In a GIS, the link between each piece of data and the geography of the corresponding geographic feature (object) is generally ensured using a unique Identifier or ID.

[10] Naturally and artificially-created features on the earth. Natural geographical features consist of landforms and ecosystems. For example, terrain types, (physical factors of the environment) are natural geographical features. Conversely, human settlements or other engineered forms are considered types of artificial geographical features
[11] Computer representation of geographical feature

As an example, the conceptual data model presented in Annex 1 has been developed to identify all the geographic features needed to be considered in the context of malaria elimination in general and the implementation of ADB's malaria and dengue risk mapping and response planning in the Greater Mekong Subregion project in particular.

In Annex 1, different shapes have been used to differentiate between:

- Geographic features for which a master list[12] is needed [2]. These are represented by rectangles.
- Geographic features for which a master list is not needed/applicable due to their continuous nature. These are represented by ovals.
- Groups of attributes which are represented by white parallelograms. While generally not included in a conceptual data model (Table 1), this item has been included here as it was adding some value to the overall model without overloading it.

Colors are used to differentiate between features/objects that relate to:

- Health (in blue),
- Spatially distribute malaria hazard (orange), and
- The malaria risk mapping and elimination (in grey).

Arrows are used to indicate a relationship between two geographic features or between a geographic feature and a group of attributes. This relationship can be of different types, including:

- Geographic (is within, lives in),
- Administrative (is reporting to)
- Health related (covers, provides service to, refers to),
- Associative (is part of, defines),

These different types of relationships can be captured in the final database structure. It is important to note here that the fixed geographic features presenting the highest number of relationships in Annex 1 are:  health facilities, communities/settlements (city, towns, villages, hamlets), administrative and reporting divisions. These represent the core geographic features for which a master list should be developed in priority to allow for an effective use of geospatial technologies in public health.

Annex 2 describe the process that can be followed during a workshop to identify these core geographic features, define each of them, as well as capture the relationships that exists between them under the form of hierarchies[13]. These hierarchies that can then be connected to obtain the conceptual data model. To be successful this exercise should involve all the partners having a stake in the public health priority(ies) that the conceptual data model aims at covering.

Once the conceptual data model is developed and agreed upon among all concerned stakeholders, the next step consists of developing the logical data model[14, 15]. It is important to mention here that this model mainly applies to data being stored partly or fully in tabular forms. Continuous

[12] 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 geographic feature/object (e.g. health facilities, administrative divisions, villages)

[13] Arrangement of items (objects, names, values, categories, etc.) in which the items are represented as being "above", "below", or "at the same level as" one another

[14] http://learndatamodeling.com/blog/logical-data-modeling-tutorial/

[15] http://www.1keydata.com/datawarehousing/logical-data-model.html

geospatial data such as the Digital Elevation Model (DEM) or land cover which do not contain a table will therefore remain represented as is in the conceptual data model.

As described in Table 1, such model does also contain attributes as well as primary and foreign keys. At the same time, relationships between entities/objects are specified using primary keys (PK) and foreign keys (FK) therefore specifying what attributes are used for this relationship. Due to the size that this type of data model can reach, Figure 2 only presents the primary keys (PK), foreign keys (FK) attributes, and some of the relationship for 4 of the entities/objects reported in Annex 1, namely: health services, administrative divisions, health workers, and laboratories.

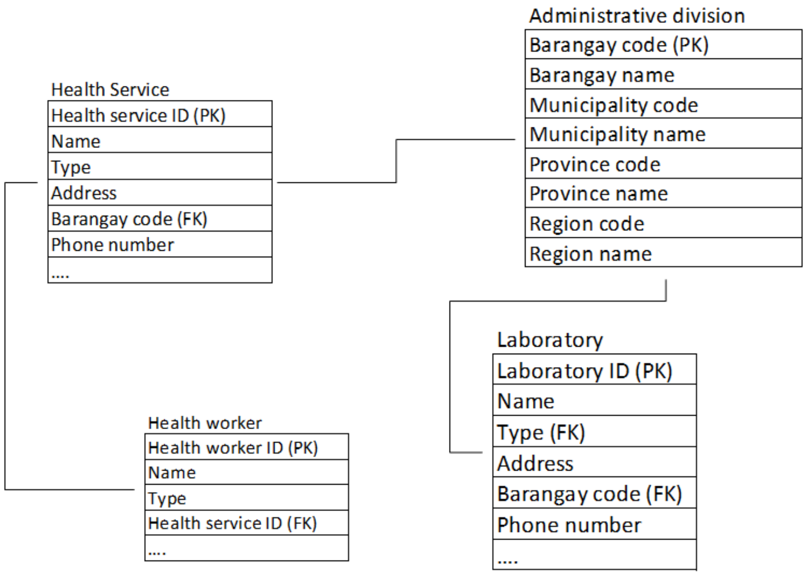



Figure 2 - Example of logical data model extracted from Annex 1

The final step in the process consists of expanding and slightly modifying the logical data model to become a physical data model[16, 17] which will be used to build the database itself.

This is done by doing the following on the logical model:

- Converting entities into specific tables names
- Converting attributes into column names
- Specifying the data type for each of the column

Figure 3 present an example of such transformation for the entities/objects reported in Figure 1.

[16] http://learndatamodeling.com/blog/physical-data-modeling-tutorial/
[17] http://www.1keydata.com/datawarehousing/physical-data-model.html

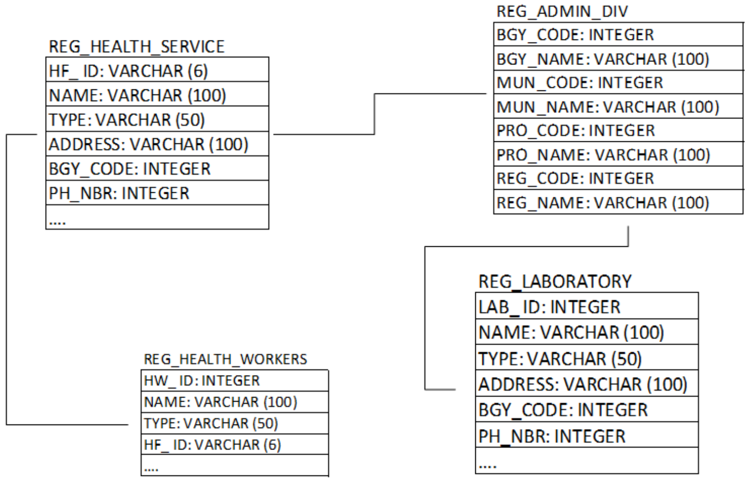



Figure 3 - Example of physical data model extracted from Figure 1

The different models presented here should not only evolve as the project implementation progresses but should also be improved, completed, and potentially updated through a consultative process among involved stakeholders.

## 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: 2.2 Defining the terminology, the data specifications, and the ground reference. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_2.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: www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part_2_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]

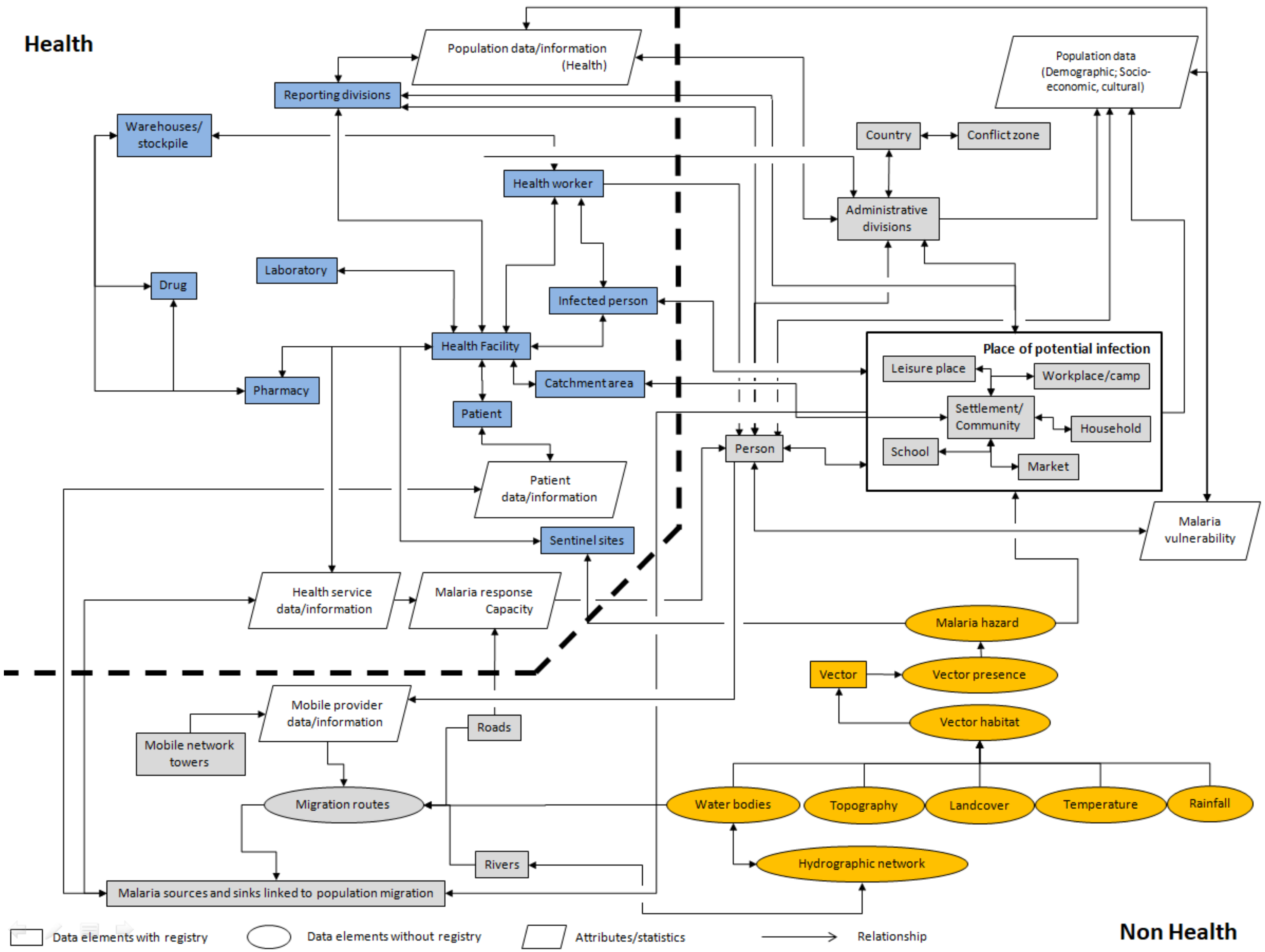
[9] Ebener S. (under preparation): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.2 Using thematic maps for decision making. Health GeoLab document.

[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.

# Annex 1 – Example of conceptual data model - Malaria elimination



# Annex 2 – Process aiming at identifying core geographic features, defining them, and capturing the relationships existing between them during a workshop

**Material to prepare for the workshop:**

- A set of introductory slides presenting the objectives of the consultation, describing what is meant by geographic features, relationships, hierarchies as well as the process that will be followed to identify them.
- A white or paper board per group, plus at least two more boards to analyze the results of the different exercises.
- Sticky notes (one color per group)
- Markers to write on the sticky note and boards.
- A mobile phone to take pictures.

**Process followed during the workshop:**

1. Present the introduction slides (10 minutes)
2. Conduct the first group exercise to identify the core geographic features (45 minutes)
    1. Create groups of 4 to 5 participants (each group would look into a different public health priority in case the conceptual data model is meant to cover more than one).
    2. Attribute one board and color of sticky note per group.
    3. Ask each group to write down one geographic feature they think is core to addressing the public health priority per sticky note and to place them on the group-specific board (give them 20 minutes for that)
    4. Take a picture of each paper board before moving on to the next step.
    5. Bring all the boards in one place and the participants in front of them.
    6. Use two separate and empty boards and:
        i. On the first board: Group together the business or programmatic data (statistics or information) as they are not geographic features but information/data that can be attached to them (these data elements are meant to be captured in the logical data model)
        ii. On the second board, group together the sticky notes containing the following (Figure A2.1):
            - Fixed or mobile features that are present across several, if not all, the groups (blue rectangle in Figure A2.1) and highlight that these are the geographic features you would want to start working on as part of the geo-enablement of the HIS. Having a master list for each of these features and having it used by all the programs would not only reduce duplication of efforts but also facilitate data interoperability and the possibility to fully benefit from the power of geography, geospatial data, and technologies (see volume 2.2 of the HGL guidance series [2])
            - Fixed or mobile geographic features that were only mentioned by one of the groups (yellow rectangle in Figure A2.1). These are important but less central to the priority(ies) covered by the exercises and could therefore be handled at a later stage.
            - Continuous geographic features (red rectangle in Figure A2.1) and explain that these do not require the development of a master list and that they will directly be captured and used under the form of geospatial data.

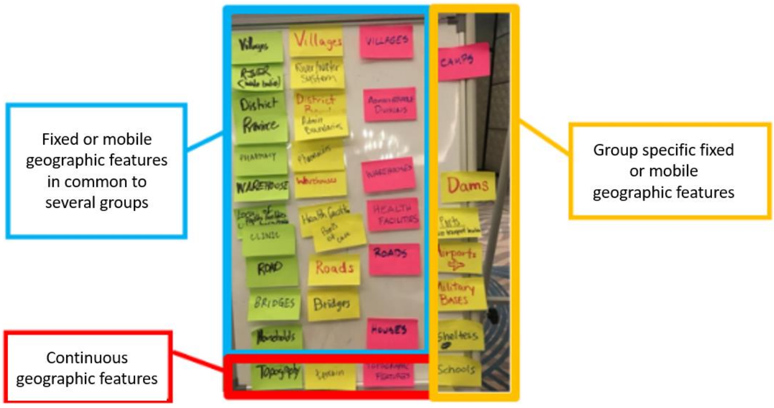



Figure A2.1 – Example of board resulting from the first exercise

7. Comment on the final board which provides a good view of the geographic features composing the geography of the identified public health priorities.
8. Take a picture of the white/paper board before moving on to the next exercise.

3. Conduct the second group exercise aiming at providing a definition for each of the geographic features obtained during the first exercise (45 minutes)
   a. Identify which among the geographic features identified during exercise 1 there might not necessarily be a consensus among the group when it comes to their definition or no official definition exists at this stage (e.g. captured in an official document). Example of such geographic features generally include settlements, community, health facilities, health area, catchment area, vaccination points,…
   b. Use the same groups as with the first exercise.
   c. Ask each group to come up with their own definition for each of the geographic features identified during step 3.a and to capture these definitions on the paper board (20 minutes)
   d. Bring all the boards in one place and the participants in front of them.
   e. Go through the definition captured by each group and identify:
      i. Those for which there is a consensus among the groups. If needed, combined the content of these definitions into one single definition agreed upon by everybody and capture it on a separate blank board.
      ii. Those for which there is no clear consensus. Take the time to identify where the lack of consensus is coming from as this is not only key for the rest of the workshop but also for the implementation of the HIS geo-enabling process in general see volume 2.2 of the HGL guidance series [2]).
         - If a consensus can be obtained, then capture the final definition on the board (this might require expanding the list of geographic features in case the definition captured by one of the groups was actually referring to another geographic feature).
         - If such a consensus cannot be obtained during the workshop, then capture this on the board for resolution after the workshop.

      A table similar to the one presented in Figure A2.2 should be obtained at the end of the exercise.

| Geographic feature | Agreed upon definition |
|---|---|
| Health facility | Fixed infrastructure where health care is being provided (health center, hospital) |
| Health area | Area around a health facility defined for the purpose of cataloging, budgeting, and managing health resources |
| Catchment area | A geographical area delineated around a health facility, from where the population utilizes its services |
| Community | A group of people living in the same place or having a particular characteristic in common |
| Vaccination point | Place used by the Ministry of Health for the vaccination of the population outside health facilities |

Figure A2.2 – Example of table obtained at the end of the second exercise

f. Take a picture of the different boards before moving on to the next exercise.

4. Conduct the third group exercise aiming at capturing the relationships between geographic features in hierarchies (1 hour)
   a. Use the same groups as with the first exercise.
   b. Ask each group to draw at least one hierarchy that is central to the public priority covered by the group (give them 30 minutes for that):
      i. Place the fixed or mobile geographic features that are part of the first hierarchy on the board (they can use the sticky note).
      ii. Draw the relationships between the geographic features for that hierarchy.
      iii. Identify and write down the type of relationship(s) that links the geographic features in the hierarchies they have drawn (geographic, administrative, health related, associative,…). The board at this end of this exercise should look like the content of Figure A2.3

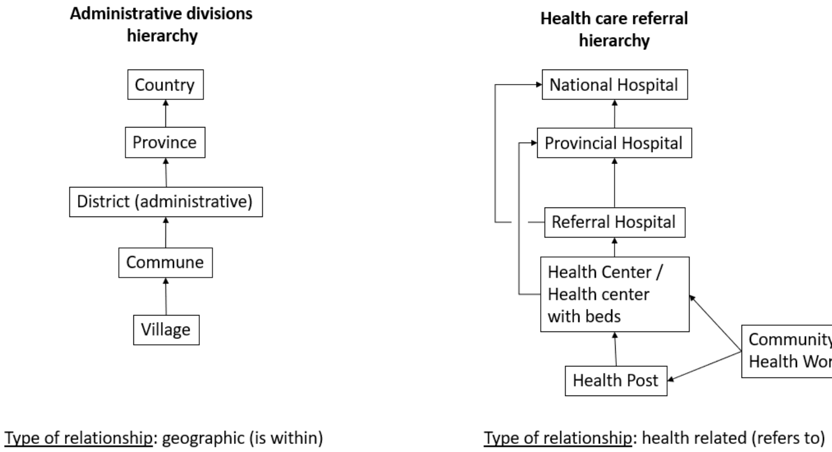



Figure A2.3 – Example of hierarchies resulting from the third exercise

      iv. If time permits, repeat the exercise for additional hierarchies in separate sheets on the board
   c. Bring all the boards in one place and the participants in front of them
   d. Go through the hierarchies captured by each group and:
      i. Make sure that each hierarchy, and relationships, are expressed in a way that it can be easily understood (make the necessary adjustments in collaboration with the concerned group if this is not the case)
      ii. Identify and address potential discrepancies that exist between the same hierarchy being drawn by different groups (for example, difference in the organization of the administrative divisions in the country)
      iii. Highlight hierarchies that are:
         - Being used across different public health scenarios
         - Connecting the same set of geographic features in a different way
         - Containing geographic features that have not been identified during the first or second exercise
      iv. Label each hierarchy for easy reference (e.g., administrative structure hierarchy, health care referral hierarchy,…)
   e. Take a picture of the different boards before moving to the next step

5. Conclude the workshop by:
   a. Summarizing the outcomes of each exercise conducted during the workshop and highlight the importance of each result in the context of the HIS geo-enabling process
   b. Listing the issues that could not be solved during the workshop (definition not finalized, for example)
   c. Presenting the way forward

**Tasks conducted after the workshop:**

1. Group hierarchies into a draft conceptual data model and share it with the workshop participants for review and inputs before considering it as final (example in Annex 1).
2. Address the question of the definitions for which a consensus could not be obtained during the workshop.
3. Capture the final list of geographic features, definitions, hierarchies, and conceptual data model in a document to be part of the guidelines the Ministry of Health will use to support the geo-enablement of its HIS.