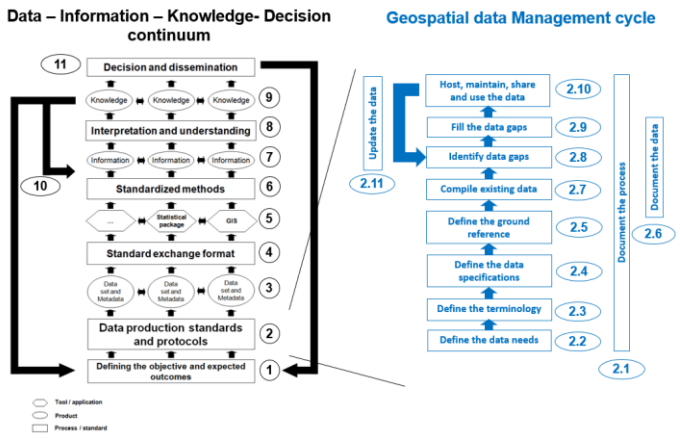
Guidance for the management and use of geospatial data and technologies in health

Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle

Version 1.3 (last update: 22.10.2023)



In collaboration and with the support of:

Revision history

Version	Release date	Comment	By
1.0	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	04 March 2020	Adjustment of the terminologies 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	22 October 2023	Adjustment in the text and simplification of the geospatial data management cycle	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
DOH	Department of Health
GIS	Geographic Information System
GMS	Greater Mekong Subregion
HGL	Health GeoLab
R-CDTA	Region-Capacity Development Technical Assistance
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 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 (the present document)
- Part 2 – Implementing the geospatial data management cycle:
    - 2.1 Documenting the process and define the data needs [1]
    - 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 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

Generating and maintaining good quality geospatial data as well as generating 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 present a framework which aims at linking geospatial data management to the overall data-information-knowledge-decision continuum as well as providing an organized geospatial data management cycle which covers all the steps to be followed to generate, maintain, use, and share quality geospatial data.

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. The data-information-knowledge-decision continuum

Data, information, and knowledge are concepts that can be defined in different ways. In the context of the present series, they are being defined as follows:

- Data: Raw, unorganised facts and statistics collected for reference or analysis
- Information: Data processed, organised, structured or presented in a given context so as to make it useful
- Knowledge: Facts, information, and skills acquired through experience or education

By extension, geospatial data, information, or knowledge can be defined as data, information, or knowledge that has a geographic component to it.

Due to the connection that exists between them, these concepts can be organized in the form of a continuum such as the one presented in the left part of Annex 1.

Even if passing from data to information and then from information to knowledge is not as linear as presented here, this way of looking at the whole process is sufficient to ensure the production of quality data, information, and knowledge to support decision making.

This continuum is a cycle where the learning acquired during each loop constantly improves not only the processes being used but also the quality of the data, information and knowledge being produced, and decisions taken.

Once the needs (objectives and expected outcomes) are defined, step 1 in Annex 1, standardisation, takes place at 3 levels along the chain, namely using:

1. Data production standards and protocols (step 2) to ensure the generation of data compatible among sources (step 3).
2. Standard exchange formats (step 4) to ensure interoperability among systems and software (step 5).
3. Standardised methods (step 6) to ensure the generation of reliable and compatible information among sources (step 7).

From there, the interpretation and understanding of the information by experts (step 8) is the process through which knowledge is being generated (step 9).

At this point, some additional information might be needed, or the initial needs might even have to be redefined (back loop under step 10). If not, then the knowledge which has been generated is disseminated and/or used for decision making (step 11).

Finally, new needs might emerge from the decisions being taken. This is what is illustrated by the other back loop (step 12).

It is important to note here that there is a process behind each of the boxes along this chain (step 1, 2, 4, 6, 8 and 11).

Among those processes, the geospatial data management cycle (point 2.1 to 2.13) is crucial to ensure the production and use of quality and documented geospatial data for decision making.

4. The geospatial data management cycle

Data management comprises all the disciplines related to managing data as a valuable resource.[2] These disciplines apply to geospatial data as well.

Proper geospatial data management requires the implementation of all the steps reported in the geospatial data management cycle (right part of Annex 1). This cycle is an extended version of the one developed by Ebener et al. [8] to illustrate the different elements and steps to be considered for geospatial data to be of quality for decision making.

The starting point of the geospatial data management cycle is the definition of the overall need (objectives and expected outcomes) for the activity/project/strategy which is to be implemented (Step 1 in Annex 1).

As an example, the need considered in the context of the present series of guidance is directly extracted from the WHO's strategy for Malaria elimination in the Greater Mekong Subregion (2015-2030) [13] and more precisely its goal which reads as follows:

"The ultimate goal of the new strategy is to interrupt the transmission of malaria and eliminate the disease, in a sustainable way, within all affected countries of the GMS by 2030. Considering the urgency of action against multidrug resistance in the GMS, Plasmodium falciparum is to be eliminated by 2025."

[2] http://en.wikipedia.org/wiki/Data_management

Once such need is defined, the geospatial data management cycle can be implemented through the following steps (Annex 1):

1. Document the whole process from the beginning. Doing this allows for the process to be replicable in other geographic areas and/or context.
2. Define the data that will be needed to achieve the original objective and expected outcomes,
3. Define the vocabulary to make sure that all the people involved in the collection, maintenance, and use of geospatial data are understanding each other.
4. Define the data specifications to ensure data quality across sources.
5. Define the ground reference that will be used to be in the position to measure the potential geographic and temporal shift that exists between the geospatial data being used and the reality.
6. Define the metadata profile that will be used to document the data and collect the necessary information to fill it.
7. Compile the already existing data to avoid duplication of efforts and reduce data collection to the minimum.
8. Identify potential data gaps based on the ground reference and the data specifications defined earlier.
9. Fill the data gaps by improving the existing data, collecting/extracting new data, and validating the resulting data.
10. Host, maintain, share and/or use the data based on the original defined needs; and
11. Update the data to ensure its timeliness.

The cycle is to be restarted and adjustments made at each of its steps as the data needs evolve.

Each of the above-mentioned steps is described in more details, together with recommendations to produce, use, and maintain quality geospatial data, in the subsequent volumes of the present guidance. Please refer to the background section of the present document for the complete list of documents.

References

[1] 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]

[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 other sources. 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] WHO (2015): Strategy for Malaria Elimination in the Greater Mekong Subregion (2015–2030)

Annex 1 – The data-information-knowledge-decision continuum and the geospatial data management cycle