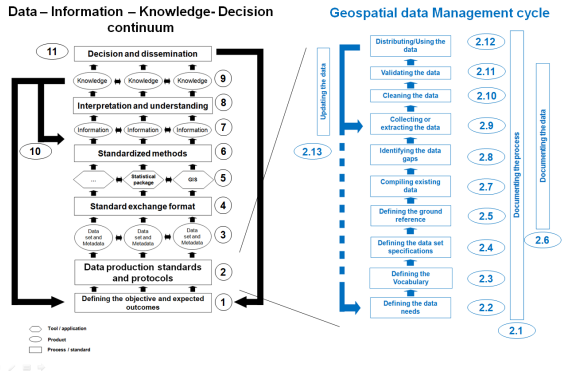
Guidance for the collection and use of geospatial data in health

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

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

1. Background 4
2. Introduction 5
3. The data, information and knowledge management continuum 5
4. The geospatial data management chain 6
References 7
Annex 1 – The data-information-knowledge continuum and the geospatial data management chain 8

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 the 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
DOH	Department of Health
GIS	Geographic Information System
GMS	Greater Mekong Subregion
R-CDTA	Region-Capacity Development Technical Assistance
WHO	World Health Organization

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 (the present document);
- 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;
    - 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

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 in order to generate, maintain, use, and share quality geospatial data.

The present document builds on previous publications [1], 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 [2] 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: Facts and statistics collected for reference or analysis[3]
- Information: Facts provided or learned about something or someone[2]
- Knowledge: Facts, information, and skills acquired through experience or education[2]

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 enough to ensure the production of quality data, information, and knowledge to support decision making.

This continuum is actually 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.

[3] http://www.oxforddictionaries.com/definition/english/data

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

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.[4] 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. [1] to illustrate the different elements and steps to be considered in order 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) [3] 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."

[4] 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 in order 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 (projection, accuracy, format,...) in order to ensure compatibility among 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. Collect or extract data in order to fill the identified gaps;
10. Clean the collected data and assess their validity based on the ground reference and data specifications;
11. Distribute and/or use the data based on the original overall needs; and
12. Update the data to ensure its constant timeliness.

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., 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.

[2] 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

[3] WHO (2015): Strategy for Malaria Elimination in the Greater Mekong Subregion (2015–2030): http://iris.wpro.who.int/bitstream/handle/10665.1/10945/9789290617181_eng.pdf?sequence=1 [Accessed May 1, 2017]

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