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

Version 1.1 (last update: 19.02.2024)

Revision History

Revision	Revision Date	Comment	By
1.0	21 December 2021	Document created	Steeve Ebener, Izay Pantanilla
1.1	19 February 2024	Update the layout and the URLs across the document	Steeve Ebener

Authors

Steeve Ebener[1]
Izay Pantanilla[1]

Acknowledgements

Our gratitude goes to Dr Richard J. Maude[2,3,4] for his contributions to this document.

1. Health GeoLab, Manila, Philippines
2. Mahidol-Oxford Tropical Medicine Research Unit (MORU), Mahidol University, Bangkok, Thailand
3. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK
4. Harvard TH Chan School of Public Health, Harvard University, Boston, USA

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 general and public health in particular.

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

GCS	Geographic Coordinate System
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
HGL	Health GeoLab
HIS	Health Information System
OSM	OpenStreetMap
PCS	Projected Coordinate System
RS	Remote Sensing
SDG	Sustainable Development Goal
WMS	Web Mapping Service

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 that are organized as follows:

- Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle [1]
- Part 2 - Implementing the geospatial data management cycle:
  - 2.1 Documenting the process and defining the data needs [2]
  - 2.2 Defining the terminology, data specifications, and the ground reference [3]
  - 2.3 Compiling existing data and identifying gaps [4]
  - 2.4 Creating geospatial data
    - 2.4.1 Extracting vector format geospatial data from basemaps (the present document)
    - 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

Creating vector format geospatial data or filling geospatial data gaps in existing vector format datasets is an important step of the geospatial data lifecycle [1].

This step generally follows the compilation of existing datasets and the identification of potential gaps [4] and can be performed in two ways depending on the data in question and the availability of resources:

1. Extracting vector format geospatial data from basemaps
2. Collecting the data in the field.

The present guidance focuses on the former while Volume 2.4.2 of the series [5] focuses on the latter.

Data extraction refers to the act or process of retrieving data out of data sources for further data processing or data storage.[2]

We can mention the following examples of application of vector format data extraction pertaining to public health:

- Obtaining the geographic coordinates of health facilities using satellite images when the use of GNSS-enabled devices is not possible
- Improving or completing the GIS format layer containing the road network to assess how physically accessible existing health services are to the population or how such accessibility could be improved
- Preparing for foci investigation by mapping visible features before the intervention team goes in the field
- Identifying how many households, and indirectly, how many people, are living in a particular area to support the planning of routine or emergency vaccination campaigns
- Delineating or modifying the boundaries of an administrative or operational division based on paper maps

Such extraction is therefore always performed in relation to specific geographic features and represented as geographic objects in a GIS format dataset.

[2] https://en.wikipedia.org/wiki/Data_extraction

3. Data extraction process

The generic process to be observed when extracting geospatial data in vector format is as follows:

1. Define the data specifications associated with the geographic features to be extracted
2. Identify the available and most appropriate basemap(s) from which the geographic features could be extracted
3. Decide on the most appropriate method to perform the extraction
4. Implement the method

The following sections describe each of these steps in more detail.

3.1 Define the data specifications associated with the geographic features to be extracted

The first step in the data extraction process consists of defining the specifications to which the final dataset should comply.

These specifications are based on the 6 dimensions of data quality defined by the Data Management Association International (DAMA): completeness, uniqueness, timeliness, validity, accuracy, and consistency [3].

Defining the data specifications is a step that comes very early in the geospatial data life cycle [1] and should therefore normally have been defined by the time there is a need to extract data from other sources.

However, it can be that the only objective of the project you are working on is to extract geospatial data. In this case, the data specifications may have to be defined before proceeding with the extraction. Please refer to volume 2.2 of the HGL guidance to help you in this regard [3].

It is also important at this stage to identify how the geographic features will be represented as geographic objects in the final dataset as this might have an influence on the method used during the extraction.

The following three types of representation can be considered when it comes to geographic features to be captured in vector format in a geospatial dataset:

1. Points (used to represent features like health facilities, villages, households),
2. Lines (e.g., roads, rivers), or
3. Polygons (e.g., administrative boundaries, operational division boundaries, building footprints)

3.2 Identify the basemap(s) from which the geographic features will be extracted

Once the data specifications and the mode of representation of the geographic objects have been defined, the next step consists of identifying the basemap(s) from which these objects will be extracted.

The basemap(s) in question should comply with the specifications that have been defined at the start of the process (see Section 3.1).

This being said, filling gaps in a specific dataset, or generating a new dataset, might require the combination of several basemaps. As such, it is important to compile all the available and accessible basemaps at the beginning of the process together with their associated metadata. In an ideal case, such metadata should cover the following information about the basemap (please refer to HGL guidance 2.5.1 for more details [6]):

- Source
- Temporal validity
- Spatial reference system
- Resolution (in the case of satellite images) or production scale (in the case of basemaps extracted from satellite images or paper maps)
- Data quality e.g. completeness
- Limitations on the use, including extraction and/or redistribution

Scanned maps and radar imagery can be used as basemaps for extracting building footprints and roads but radar imagery is best used when extracting water features such as rivers, lakes, and flooded areas because it is not affected by cloud cover.

The following sections first describe the two main types of basemaps that can be used for the extraction of vector format features before presenting how to assess the concurrence of each dataset with the data specifications.

3.2.1 Types of basemaps

There are two main types of basemaps from which geospatial data in vector format can be extracted:

1. Remote sensing imagery
2. Paper maps

Other types of basemap, accessible through web mapping services for example, are not covered here as they are generally generated using vector format GIS layers (e.g., OpenStreetMap).

The following sections describe these sources in more detail.

3.2.1.1 Remote sensing imagery

Remote sensing (RS) is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft)[3] in contrast to on-site observation. As such, RS images are also known as satellite (spaceborne) images or aerial (airborne) photographs.

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

1. Institutions giving free access to native images or imagery through a Web Mapping Service (WMS)[4] or a Web Map Tile Service (WMTS)[5]
2. Companies offering commercial imagery. Depending on the option you purchase, you will either have access to the native images or be able to access them through a WMS or WMTS.
3. Online GIS platform like ArcGIS Online, online viewing platform such as Google Maps or Bing Maps, or online extraction platforms like those used by OpenStreetMap (OSM) which use satellite imagery as a basemap.

In relation to the above, it is important to emphasize that:

- Data available through a WMS/WMTS are generally not downloadable by the operators
- Both native images and images accessible through a WMS/WMTS will require desktop GIS software such as QGIS (open source) or ArcGIS (proprietary) or an online GIS platform like ArcGIS Online to be viewed and used
- Many different types of sensor exist. Each of these sensors will present specific characteristics (active vs passive, panchromatic (one band of data) vs multispectral (several bands), grayscale vs color composite) and lead to different types of images. It is therefore important to identify the appropriate sensor and, indirectly, the image for the purpose of the extraction that needs to be performed. While natural color composite images (passive and multispectral sensor) are generally preferred for the extraction of vector format features, the following two web pages give an overview of what is available:
    - Remote sensing satellite and data overview: https://en.wikipedia.org/wiki/Remote_sensing_satellite_and_data_overview
    - Interpreting optical remote sensing images: https://crisp.nus.edu.sg/~research/tutorial/opt_int.htm
- Images available through viewing platforms such as Google Maps benefit from contextual information appearing on top of the images. This information includes the name of places, roads, rivers, or infrastructures like health facilities, schools, or shopping malls that allow the operator to locate himself on the images. This being said, it is important to remember that in some cases, the contextual information has been generated by the public and, as

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

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

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

such, might not always be located on top of the correct feature or might not be correct at all.

- Each image or photograph acquired by the sensor covers a specific area on the ground. The extent of this area, also referred to as a scene, will vary from one sensor to another. Several adjacent scenes, forming a mosaic, will generally be needed in order to cover the entire zone of interest for a project. In certain cases, the images composing the mosaic could have been taken at different dates (due to the presence of clouds for example) leading to the period of validity for the complete dataset to extend over a long time.
- The resolution[6] of the images or photographs can vary from one source to another depending on the sensor that has been used. Such resolution will have an important impact on which features can be visible during the extraction.
- Satellite images do not only have to be georeferenced[7] but also orthorectified[8] to ensure the best accuracy possible when extracting features. It is therefore important to find the information on how these two operations have been performed and, indirectly, on the expected horizontal accuracy of the images before using them. Describing these two operations is beyond the scope of the present guidance. Please refer to the following web page for more information about this: https://apollomapping.com/blog/g-faq-orthorectification-part

3.2.1.2 Paper maps

While the digitalization of data continues to expand, some maps remain only available in paper or scanned form. This is, for example, the case for maps showing the situation observed in the less recent past, maps containing features not visible on a satellite image (e.g., boundaries of administrative divisions), or maps that have been sketched on a piece of paper in the field.

It is also important to mention that, in some cases, printed maps unfortunately remain the only expression of previously existing GIS format datasets which have been lost. One more consideration to be made when using paper maps is to give priority to official sources.

Before being able to extract features from a paper map using one of the methods described in the present guidance, there is a need to:

1. Scan the map to convert it into a digital image. Two things are important during this step:
    a. As much as possible the map should be scanned at once which might pose a certain challenge when the paper map is big and would require a large size scanner. If such a scanner is not available, the map will have to be divided into areas, each of these areas being scanned individually and then combined (mosaicked) together before the extraction.

[6] Size of the cells in a raster dataset

[7] Process aiming at aligning the image with reference on the ground. This is done by assigning coordinates (e.g. latitude and longitude) to the image

[8] Process of removing internal and external distortions to assign more accurate coordinates to the final image. Once done, each pixel in the image will appear as if it was viewed directly from above, thereby removing the effect of the topography (hills, valley,..)

   b. The map should be scanned at the highest resolution available to ensure the highest level of accuracy possible when extracting features out of it. This parameter is a function of the scanner you will use and of its available settings. Please refer to its user manual for more information about this.
2. Georeference the digital image of the map. Similar to satellite images, this step is necessary to assign geographic coordinates (i.e., latitude and longitude) to the image. In order to perform this step, points of reference (e.g., landmark, map grid[9],…) have to be clearly visible on the scanned map and identifiable on the RS images used as ground reference. The following two links provide information on how to perform such georeferencing in the two GIS software covered in the present document:
   a. ArcMap: http://bit.ly/3394Gtx
   b. QGIS: http://bit.ly/38IwXIQ

When georeferencing a scanned map, it is critical that the RS images you use as ground reference have a high level of accuracy. The following web page can help you test the accuracy of the remotely sensed images you plan to use if you think that such a check is necessary: http://bit.ly/2wGWDbh

Please also note that a scanned paper map does not need to be orthorectified as, by definition, the features on this kind of map already appear as if directly seen vertically from above.

3.2.2 Concurrence with the data specifications

Once all the available and accessible basemap(s) have been compiled, the next step consists of identifying which one(s) are the most suitable to extract the features of interest.

This evaluation is performed by assessing the concurrence of each of these basemaps against the data specifications defined at the start of the process (see Section 3.1).

In an ideal case, the basemap, or combination of datasets, being used for the extraction will comply with all these specifications.

Unfortunately, it can happen that such concurrence is not reached for all the data quality dimensions. When this is the case:

- Other basemap(s) will have to be found
- If no other basemap(s) are available and/or accessible, then:
    - Either field data collection will have to be considered as an alternative or a complement to data extraction. Please refer to guidance 2.4.2 in this case [5]. Or,
    - Gaps will remain at the end of the extraction. The type and the impact of such gaps will depend on the data quality dimension(s) that is/are not fulfilled.

[9] A network of evenly spaced horizontal and vertical lines used to identify locations on a map

The following sections describe what to assess across the 6 dimensions of data quality covered by the specifications in question for the 2 types of basemaps presented in Section 3.2.

3.2.2.1 Completeness

Completeness refers to the state or condition of having all the necessary or appropriate parts.

In the present context, this corresponds to the need for the basemap(s) to contain all the features that have to be extracted.

Verifying that this is indeed the case is not necessarily straightforward and will mainly depend on the objective of the extraction, the type of features to be extracted as well as the source data being used. More specifically:

1. If the objective is to extract all the instances for a given type of features (for example, all the buildings, roads, or rivers in an area), then the limiting factors against reaching completeness will be as follows:
    - When using RS images as the basemap:
        - The resolution of the image (coarse resolution will make it difficult to identify features accurately)
        - Image contrast (too dark or too bright especially with monochromatic images will result in the same difficulty)
        - The presence of elements masking the features in question in the area of concern. Such elements can be clouds or vegetation cover for example (Examples in Figure 1)

a) b)

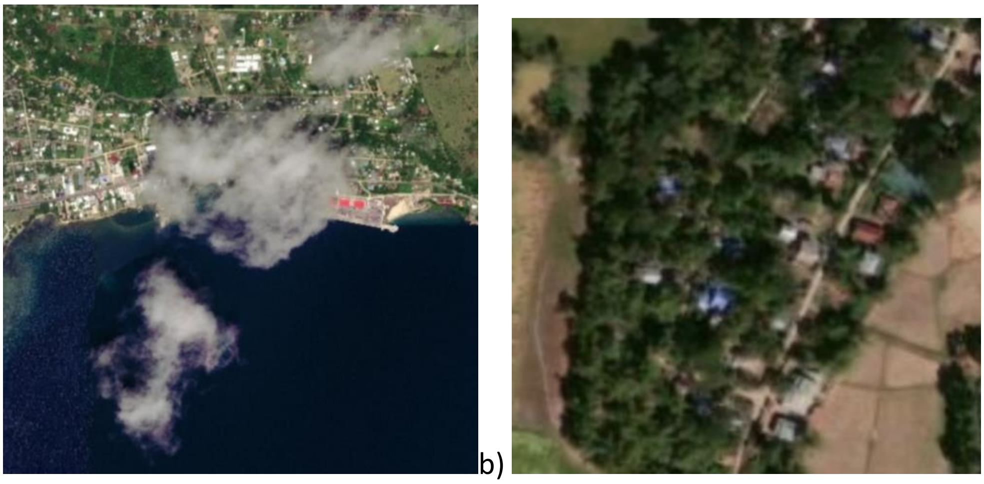


Figure 1 – Example of features masked by clouds (a) or vegetation (b) in an RS image

    - When using a scanned version of a paper map as basemap:
        - The level of completeness of the paper map itself

2. If the objective is to extract a selected number of instances based on a list or reference (e.g. master list of health facilities) then the limiting factors against reaching completeness are:

- The same as those listed under point 1 but this time only applied to the features included in the list or reference
- The absence of contextual information that would allow the operator to identify the considered features on the RS images

The use of incomplete dataset(s) can have a significant impact on any kind of analysis being conducted.

The preference should therefore be given to dataset(s) that ensure completeness and any gap in completeness should be properly documented in the metadata of the extracted dataset.

3.2.2.2 Timeliness

Timeliness refers to the degree to which data represent reality from the required point in time.

In other words, the temporal validity of the basemap(s) being used for the extraction needs to correspond to the temporal validity defined in the data specifications (see Section 3.1).

While the temporal validity of a unique RS image corresponds to the date on which it has been taken, the temporal validity of a mosaic composed of several images will correspond to the period covering the collection dates of all the images that comprise it. Please make sure to find this information before starting the extraction.

It is also important to be aware that significant temporal discrepancies between different scenes of a given mosaic can lead to feature discontinuities between scenes as in the example presented in Figure 2. In this figure, we can see that a building has been erected during the period between the capture of the two scenes (the image on the right having been collected prior to the construction in question).

Figure 2 – Example of feature discontinuity between two scenes due to significant temporal discrepancies between two scenes

The temporal validity of a vector format dataset extracted from RS images follows the temporal validity of the images from which it has been created. Operators should nevertheless be careful not to confound the temporal validity of the RS images with the date on which the dataset has been created or released. It is really the temporal validity of the RS images in which we are interested here and hopefully the information in question can be found in the metadata associated with the datasets.

For paper maps, the distinction has to be made between the date on which the map has been produced and the temporal validity of the data it contains: the latter being the information to be used when it comes to assessing the timeliness of the map in question. The guidance document on making good thematic maps shows where these two elements are generally located on a map [8].

Important temporal discrepancies within or between datasets can lead to incoherent analytical results.

The preference should therefore be given to datasets presenting a temporal validity that is exactly the same as, or close to, the one reported in the data specifications and the period of validity of an extracted dataset should be documented in its metadata.

3.2.2.3 Uniqueness

Uniqueness refers to the quality of being the only one of its kind.

In the present context, this means that none of the features to be extracted are appearing more than once in the basemap (no duplicates).

This dimension very seldom applies to RS images unless the georeferencing and/or orthorectification of the images that are part of a mosaic led to some discontinuities between scenes like in the example presented in Figure 3.

Figure 3 - Example of discontinuity between two scenes of a given mosaic of RS images which results in the same building appearing twice

This being said, if contextual information (place name, for example) is associated with the RS images, it is important to make sure that the information itself does not contain duplicates. While such identification can be more easily performed when using a vector format layer containing the contextual information in question in desktop GIS software, this exercise is more difficult when the contextual information is embedded within the RS imagery in an online application (Google Map for example).

The presence of duplicates in a scanned paper map is also rare. Apart from the same issue as the one reported here above for RS imagery when having to assemble the pieces of a paper map that could not be scanned at once, the other most common occurrence for this situation would be the use of a scanned map based on a GIS format dataset that contained duplicates and that is no longer accessible.

Identifying duplicates before performing the extraction is important in order to avoid these duplicates from being transferred to the extracted datasets and used for different analyses.

Preference should therefore be given to basemap(s) that do not contain duplicates.

3.2.2.4 Validity

Data is considered as valid if it conforms to the syntax (format, type, range) of its definition.

When it comes to geospatial data, HGL guidance 2.2 has defined the following elements as being included under this specific data quality dimension and therefore expected to be included in the data specifications:

- Geographic Coordinate System (GCS) and map projection (Projected Coordinate System (PCS))
- Geographic extent of the area being covered
- Language(s) included in the data
- File format(s) for sharing data
- Metadata standard used to document the data

When it comes to the GCS and PCS, and while modern geospatial technologies contain algorithm that allows us to re-project any geospatial data on the fly against a specific GCS and PCS specified by the operator, overlaying datasets presenting a different datum[10] will result in horizontal discrepancies and therefore horizontal errors when extracting features. It is therefore critical that all the data used as part of the extraction process present the same datum and, ideally, the same GCS and PCS as defined in the data specifications.

This being said, re-projecting RS imagery is not always straightforward and might lead to some distortions and, indirectly, horizontal errors. Operators might therefore have to consider using the

[10] Model of the shape of the earth (reference ellipsoid) to define a geographic coordinate system

GCS and PCS of the RS imagery during the extraction process and to then re-project the final extracted dataset according to the GCS and PCS defined in the data specifications.

In addition to the above, some old GIS format datasets do not automatically carry the PCS information and would need to be attributed.

The re-projection or the attribution of the GCS/PCS to a basemap are processes that can be done using tools in desktop GIS software such as ArcMap or QGIS.[11,12,13] The software will make the necessary computations and changes in the dataset to render it correctly.

Scanned maps are projected during the georeferencing process done in GIS software. They adopt the GCS/PCS of the map layer used as reference in georeferencing.

In any case, the GCS and PCS of the basemap(s) must be known in order to be able to perform the necessary adjustments before and after the extraction process and the final GCS and PCS should be included in the metadata of the extracted dataset.

The basemap, or combination of datasets, should cover the entire geographic extent of the area of interest as defined in the data specification. This ensures that all the geographic objects within this area are extracted. Consequently, only the geographic objects within this extent should be extracted. Doing this not only saves time and resources but also prevents extraneous data from being included in the analysis and/or mapping. The use of geospatial data containing the geographic extent such as an administrative boundary or a bounding box with the min/max coordinates of the extent can be used to help delineate the area of interest during the extraction process.

There are cases when certain attributes have to be attached to the features being extracted. While there is no language attached to the RS imagery itself, the language of contextual information associated with it and scanned maps can be of the language defined in the data specifications or one that needs to be translated.

In conclusion, preference should be given to basemap(s) that:

- Contain the complete specification of their CGS and PCS
- Cover the full extent of the area of interest
- When it applies, sees the contextual information displayed in the language defined in the data specifications.

[11] https://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/project-raster.htm
[12] https://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/define-projection.htm
[13] https://docs.qgis.org/3.4/en/docs/user_manual/working_with_projections/working_with_projections.html

3.2.2.5 Accuracy

Accuracy refers to the degree to which data correctly describe the "real world" object or event.

HGL guidance 2.2 [3] includes several elements under this data quality dimension. Among those, the following are directly pertinent to the topic covered in the present guide:

- Scale: Ratio or relationship between a distance or area on a map and the corresponding distance or area on the ground;
- Spatial resolution: Size of the cells or pixels in a raster dataset; and
- Positional accuracy: Quantifiable value that represents the positional difference between a geospatial layer and reality.

The positional accuracy expected from the extraction generally drives the definition of the other parameters reported here above.

Positional accuracy is directly related to scale (Table 1) and, as such, should influence the choice of the paper map to be used to perform the extraction based on its scale. For example, a paper map with a 1:50,000 scale would allow the extraction of features with an expected positional accuracy of around 26 meters.

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 1 - Relation between scale and expected positional accuracy [3]

If data extraction is meant to take place electronically, the selected paper map will have to be scanned (see Section 3.2.1.2). Scanning a paper map results in its rasterization and might therefore affect its original positional accuracy. The final positional accuracy to be considered depends on the spatial resolution that is obtained once the scanned map has been georeferenced and this is because a direct relation also exists between scale and the expected resolution of a raster layer (Table 2).

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 2 - Relationship between scale and the corresponding raster resolution [3]

In order not to alter the level of positional accuracy of the original paper map, it is important to ensure that the resolution obtained through the scanning and georeferencing of such a map corresponds to the same, if not a larger, scale as the one of the original paper map. For example, in order not to lose the positional accuracy of a 1:100,000 scale paper map, the final resolution of the scanned and georeferenced map should not be more than 50 meters (Table 2).

Two dimensions need to be taken into account when it comes to the positional accuracy of remote sensing imagery as ground reference for the extraction.

First, a specific level of positional accuracy is attached to the imagery itself. Such accuracy depends on several factors including, but not limited to, the satellite positioning technology, terrain relief, and sensor viewing angle and therefore the georeferencing and orthorectification process that has been applied to it (see Section 3.2.1.1).

In addition to that, the direct relation that exists between scale and the expected resolution of a raster layer to be representative at that scale (Table 2) and then the relationship between scale and position accuracy (Table 1) define the minimum resolution that the imagery should have in order to obtain specific positional accuracy when used as ground reference to extract features. For example, imagery with a resolution of 50 meters corresponds to a scale of 1:100,000 [Table 2], a scale at which a positional accuracy of around 52 meters can be expected [Table 1].

The final expected positional accuracy of features extracted from specific imagery is obtained by summing the positional accuracy of the imagery with the accuracy measure linked to its resolution. For example, the use of 25-meter resolution imagery presenting a positional accuracy of 50 meters would result in extracted features with a positional accuracy of 76 meters.

Last but not least, the resolution of the RS imagery can vary from one source to another and/or from one area to another, resulting in significant differences in resolution once mosaicked (Figure 4).

In this case, the expected accuracy obtained for the extracted features will change from one scene to the other. As such, priority should be given to imagery with the same resolution across the entire area of interest.

a)  b) 

Figure 4 – Examples of RS imageries with the same resolution (a) and a different resolution (b) for adjacent scenes

3.2.2.6 Consistency

Consistency refers to the absence of apparent contradictions in a database and needs to be looked at from both a geographic and temporal perspective. This data quality component is particularly important when extracting different features from separate data sources and/or combining the extracted features with other pre-existing datasets.

When possible, extracting different features from the same scanned paper map or imagery can help in addressing the question of geographic consistency among extracted features as long as the extraction method used for each of the features conserves such consistency.

The geographic consistency between extracted features and other pre-existing datasets can be addressed by overlaying the dataset in question over the basemap (scanned paper map or RS images) prior to any extraction work and checking if any geographic discrepancies between the two are observed (example: roads in the pre-existing dataset not overlapping on the roads visible in the RS images). If this is the case, the origin of such discrepancy should be identified and might lead to the decision not to use a given basemap or pre-existing dataset or, when possible, to its adjustment (for example when the discrepancy is linked to a difference in datum).

Temporal consistency can be obtained by ensuring that the temporal validity of the data sources used for the extraction (paper maps and/or RS images) as well as that of the other pre-existing datasets matches the temporal validity set for the project or in the data specifications, or at least is close enough to it to still be representative. When doing such an evaluation, it is important to remember that some features do change more rapidly on the ground than others and will not allow for such temporal discrepancy to be too important.

3.3 Decide on the most appropriate method and tool to perform the extraction

Data extraction can be done in different ways: manually, automatically, or a combination of the two. The choice of the most appropriate method will depend on the format, quality, and mode of access of the basemap(s), the volume of features to be extracted, and resources and time at disposal of the project.

The following sections describe the different methods that are currently available to users in order to choose the most appropriate one(s).

3.3.1 Manual extraction

In the context of the present guide, manual extraction refers the human-guided digitizing of geographic features from a basemap.

There are two types of manual digitizing[14]:

1. Hard-copy digitizing: A hardcopy paper map is taped onto a digitizing table connected to a computer. A digitizing puck is used to trace the objects and it feeds the coordinates and codes into a computer.
2. On-screen digitizing: the features are digitized from the selected basemap using a desktop or online GIS-based application.

Hard-copy digitizing is seldom used nowadays (paper maps being scanned and georeferenced to allow for on-screen digitizing) and will therefore not be discussed further in this document.

On-screen digitizing can mainly be performed through the use of:

- Desktop GIS software: Most of the commonly used GIS software, including ArcMap and QGIS, can be used for on-screen digitizing nowadays. Both RS images and scanned paper maps can be used as basemaps in this case. The general process used for both ArcMap and QGIS is essentially the same and described in Annex 1.
- An online application: This approach does not require software installation on your desktop and instead uses an online application which provides on-screen digitizing capabilities, including the visualization of basemap(s). Two types of online platforms can be distinguished here:
  - Those primarily designed for data visualization but that can also be used to digitize geographic features. Google My Maps is one of these platforms. It uses Google satellite images as basemap and is easy and intuitive to use. This platform allows the downloading of all or part of the data created. However, the data are downloaded as KML or KMZ file that need to be converted to the correct vector format. For point-type data, it is also possible to copy all or part of the associated data table to be able

[14] Adapted from https://learn.canvas.net/courses/464/pages/unit-5-dot-4-manual-digitizing

to extract the geographic coordinates. Annex 2 describes the process to be followed when using Google My Maps to digitize geographic features.

- Those that have primarily been designed for on-screen digitizing. Among these platforms, we can mention those used by volunteers in the context of the OpenStreetMap (OSM) project[15]. These tools (JOSM editor, Potlatch,…) allow the volunteers to digitize or edit features directly in the OSM dataset using high resolution RS images as basemap. It is important to mention that, in this particular case, users are bound to a specific data model and therefore cannot use these online tools to digitize features that would not be captured in the OSM dataset.

In conclusion, manual digitizing might be quite time-consuming depending on the area to be covered and/or the number of geographic features to be digitized. However, the result of the method might be of better quality than some automatic extraction methods due to human involvement.

3.3.2 Interactive extraction

Interactive extraction of vector format features, also referred to as interactive vectorization, consists of manual extraction of features assisted by the automatic snapping to raster cells functionalities provided by the GIS software or application being used.

This type of capability, which is available under ArcMap (https://bit.ly/3IzgE4f), facilitates the work of the operator by setting different parameters that will, for example, define the maximum raster line width for tracing as well as directly snap the vector feature being extracted to a given position on the raster line.

3.3.3 Automatic extraction

Automatic extraction is the computer-guided, automated technique for converting raster data into vector features. Computer software uses algorithm and/or Artificial Intelligence (AI) to perform this kind of operation which is referred to as vectorization.

The algorithm can be set to extract different types of features including, but not limited to, roads, rivers, or building footprints.

This process relies on the operator’s input to control how the vectorization is performed. The issue of accuracy can be improved by spending adequate time in calibrating the algorithm used for the data extraction.

This usually means doing the process on a small area first and continuously adjusting the settings and checking the result until the sufficient accuracy is achieved. It is also important to remember

[15] https://www.openstreetmap.org/

that the algorithm will differ from one area to another: the algorithm that may work to extract data from a desert area may not necessarily work as well in a mountainous area.

Other factors such as image resolution, amount of noise in the image, and the actual content of the scanned document all play a role in determining the outcome of the vectorization.

While both RS images and scanned maps can be used for the automatic extraction process, scanned maps when available for the particular geographic object needed usually produce better extracted data. This is because scanned maps tend to consist of two colors that delineate the foreground and background values: with the foreground values denoting the object that needs to be traced and the background value usually a single solid color. This makes it easier for the algorithm to trace these objects.

Automatic vectorization can be performed through the use of:

- Desktop GIS software: Automatic vectorization can be done by using specific functions in ArcMap or QGIS:
    - For ArcMap, the extension ArcScan allows interactive or automatic vectorization. ArcScan supports two vectorization methods: centerline and outline. Centerline vectorization will generate vector features along the center of the raster linear elements. Outline vectorization will generate vector features at the border of the raster linear elements.[16] ArcScan also allows the extraction of data from the whole image or only part of it.
    - For QGIS, the Polygonize (Raster to Vector) tool is used to extract data. This tool creates vector polygons for all connected regions of pixels in the raster sharing a common pixel value. Each polygon is created with an attribute indicating the pixel value of that polygon.[17]

A. Stand-alone or online application: Operators can choose from the different stand-alone software available for purchase that best meets their data extraction needs and do the extraction on their own. One example is Picterra. It is an AI-based software as a service (SaaS) platform which allows operators to interactively create a personalized AI for detecting, localizing, and counting any objects from satellite and aerial imagery. Another example is the R2V software for automated map digitizing, GIS data capture, and CAD conversion applications.

---

[16] https://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/arcscan/what-is-arcscan-.htm

[17] https://docs.qgis.org/3.4/en/docs/user_manual/processing_algs/gdal/rasterconversion.html?highlight=polygonize

3.4 Implement the method

This section describes the elements to be looked into before, during, and after the implementation of the method finally selected to perform the extraction.

These elements are described in the following subsections.

3.4.1 Before the implementation

There are certain elements that need to be looked at depending on the methods described in section 3.3. More specifically:

1. Method implemented through an online service:
    a. An internet connection with sufficient bandwidth and transfer speed
2. Desktop-based method
    a. The software needed to run the method. This can either be GIS software (ArcGIS, QGIS) or stand-alone data extraction software.
    b. A computer with the minimum requirements to run the desktop software. This information is generally provided by the software company.
3. Internet connection: An internet connection of sufficient bandwidth and speed is needed to implement online extraction of data. It is also needed in extracting data using GIS software when using online basemaps such as ArcGIS Online, Google, and Bing instead of scanned maps. These online basemaps continuously refresh their rendering whenever the operator navigates around the map requiring the internet connection.

Aside from these elements, the following are needed for all methods:

1. Standard Operating Procedure (SOP) document: a document describing the steps to be followed when extracting data. This could also include how to determine which objects in the map are applicable to be extracted for the required geographic object (e.g., when extracting household objects, should only permanent structures be captured? How about tents or camps?).

b. Master list of the object being extracted: It is the authoritative, standardized, complete, up-to-date, and uniquely coded list of all active records for a given object. The master list will be the source of the official name and unique ID of the object being extracted. The unique code will allow the extracted data to be joined with statistical data.

In some cases, it is possible or necessary to edit the basemap to eliminate unwanted raster elements prior to extraction. This is called raster preprocessing.

Skilled manpower is important for all of the methods. The people doing the extraction should be familiar with the method to be used. If this is not the case, training sessions should be done using the SOP document to ensure that the process is done according to the defined steps.

3.4.2 During the implementation

During the implementation of the method identified in the previous section, the SOP should always be followed. In the event that a certain case is not captured in the SOP (e.g., if a certain structure could be considered for the object being extracted), this should be noted, the action/s taken agreed upon, and added to the SOP for future reference and use.

Periodic checks should also be made to ensure the quality of data. This may include checking that the captured object indeed belongs to the required geographic object, within the required map extent, has the correct official name and unique ID in the attribute table, etc.

While spot-checks for data extracted by GIS software (manual or automatic) and stand-alone software can only be done by checking the data on the local computer or by sharing the file, Google My Maps allows the remote checking of the data being extracted through the internet. An example of this was the extraction of the boat routes for Vanuatu. The work done by HGL in Manila was remotely checked by the Ministry of Health of Vanuatu in Port Vila and comments and corrections were communicated in a timely manner.

3.4.3 After the implementation

Once the extraction process is completed, it is important to conduct a final validation of the resulting extracted data. If the previous steps are done properly and the periodic checks done during the data extraction allowed spotting of any error in the process or the data itself and correcting it, then the validation of the results would be a much easier and quicker process. However, the opposite is true as failure to follow the process correctly could lead to spending more time and resources in validating and correcting the results.

When validating the results, the geospatial data should be checked against the data specifications as discussed above as well as other elements that may have been included in the SOP:

- Check that the final GCS/PCS, geographic extent, and file format of the resulting geospatial data are as defined in the data specification.
- Check that the captured object indeed belongs to the required geographic object (roads are indeed roads and not rivers and vice versa, trees or other objects are not digitized as buildings, etc.).
- Check that map features being considered for a particular geographic object are captured (e.g., all types of household being considered are extracted)
- Check for the completeness and uniqueness again the object's master list
- Check that there are no missing entry in the attribute table
- Check that the unique ID and official name of each object are correctly attributed (spelling, code format, etc.)
- For point-type object, check that the geographic coordinates are captured/calculated up to five (5) decimal places.

Once the resulting data is validated, ensure that it is properly documented in the metadata to facilitate its distribution and use [6].

References

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

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

[3] 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, data specifications, and the ground reference. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_2.pdf [Accessed 19 February 2024]

[4] 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]

[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.J. (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.

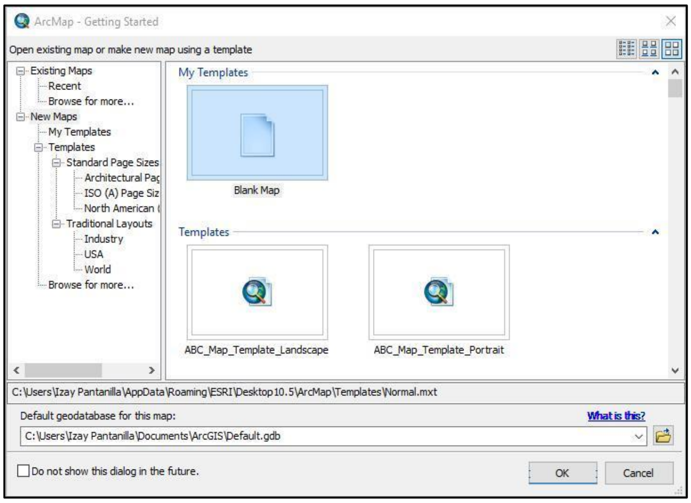
Annex 1 – Process to be followed when using GIS software to manually digitize geographic features

This annex describes the steps to be followed when using GIS software (ArcMap or QGIS) to manually digitize geographic features.

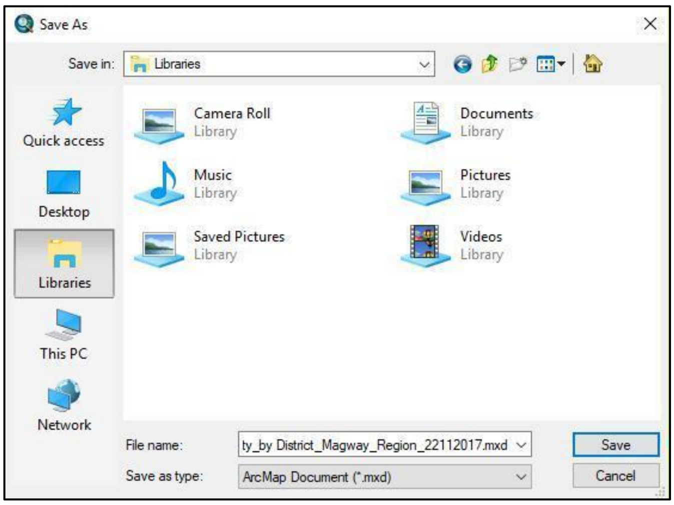
The example used for the following sections is the extraction of point-type object into a shapefile. The same process can be applied to extract line/polyline or polygon by changing the type of features in step 11.

A1.1 Using ArcMap

1. Open ArcMap.
2. When the ArcMap – Getting Started window opens, click *New Maps* from the left side menu then click the icon for the *Blank Map*. Click *OK*.



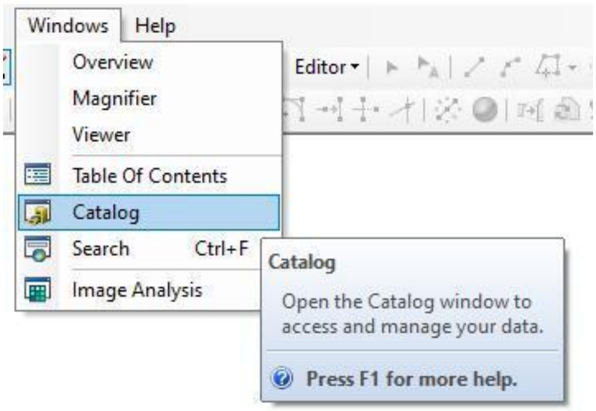
3. Save your new map document by going to the main menu, click *File* > *Save* or by clicking the *Save* button .
4. Since this is the first time you are saving the document, the Save As window will open. Browse to the location you want to save your map document in.

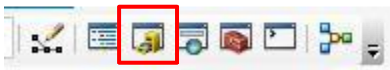


5. Type in the name for your map document in the File name field. Click *Save*.

   IMPORTANT NOTE: Remember to periodically save your map document as you go through the next steps. This ensures that your work is saved even if the GIS software or your computer suddenly crashes.

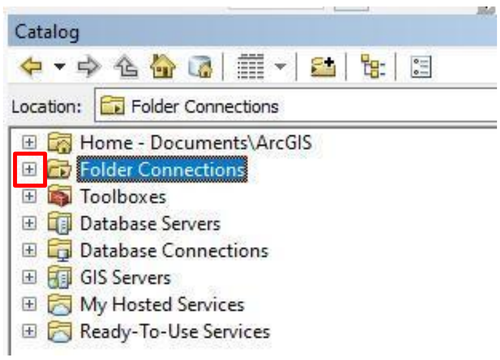
6. Create an empty map layer (e.g., shapefile) that will contain the extracted data. From the Windows main menu, choose *Catalog* or click the Catalog button from the Standards toolbar to open the Catalog window.



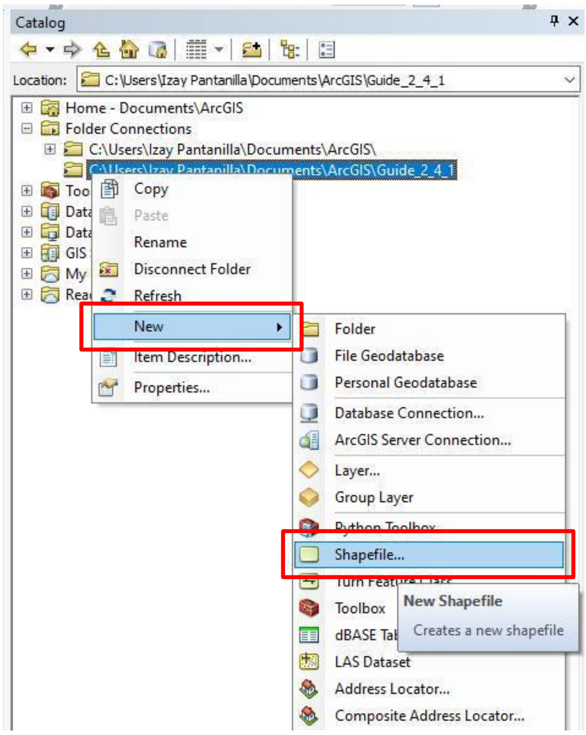


7. The Catalog window opens (usually on the right side of the ArcMap screen). Go to the folder you want to save your file in by clicking the *Connect To Folder* button . In the window that opens, browse to the location of the folder you want and click *OK.*

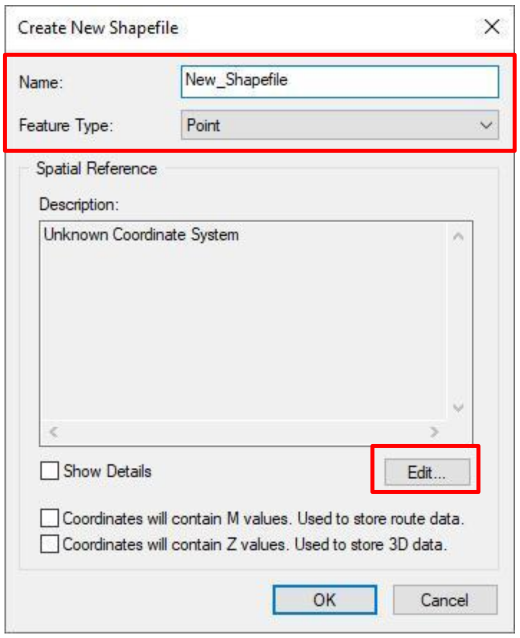
8. Click the "+" button beside Folder Connections to expand it. You will now see your chosen folder path there.



9. Right click on your folder and choose *New > Shapefile*.



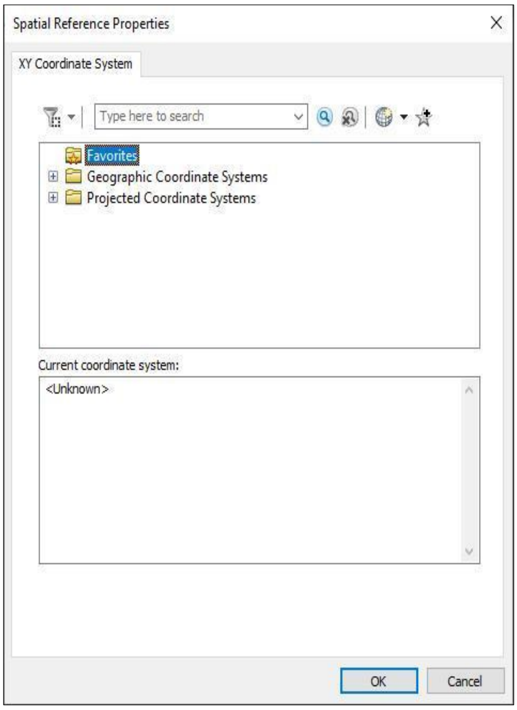
10. In the window that opens, type in the name of your shapefile with the date of the day you created it to help you keep track of your file and for your reference when creating the metadata for the file (e.g., Health_Facility_01012020).



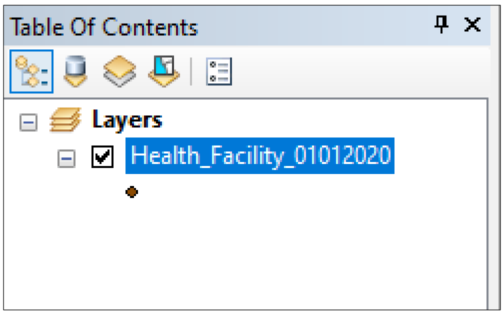
11. Choose the correct feature type (i.e., point, line /polyline, or polygon).

12. To assign a coordinate system to your shapefile, click *Edit*.

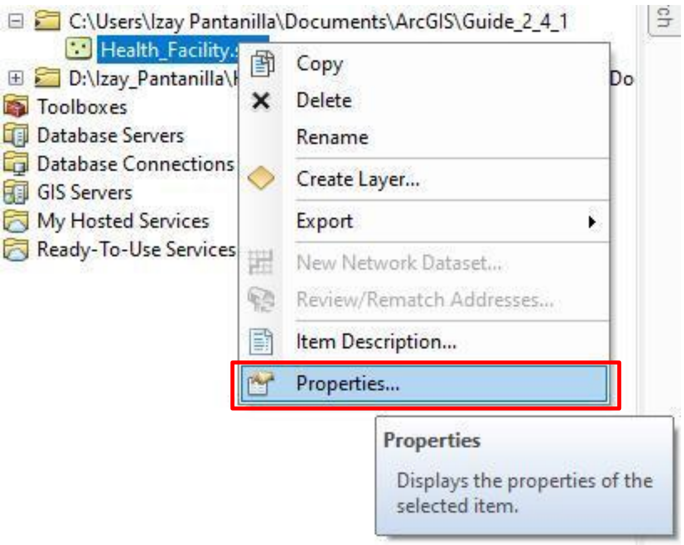
13. In the window that opens, expand the Geographic Coordinate Systems or Projected Coordinate Systems folder and choose the correct GCS/PCS as defined in the data set specifications. (You may also use the search bar on top of the window.) Click *OK*.



14. Click *OK* in the Create New Shapefile window. Your new shapefile will appear in the Table Of Contents (ToC) panel.

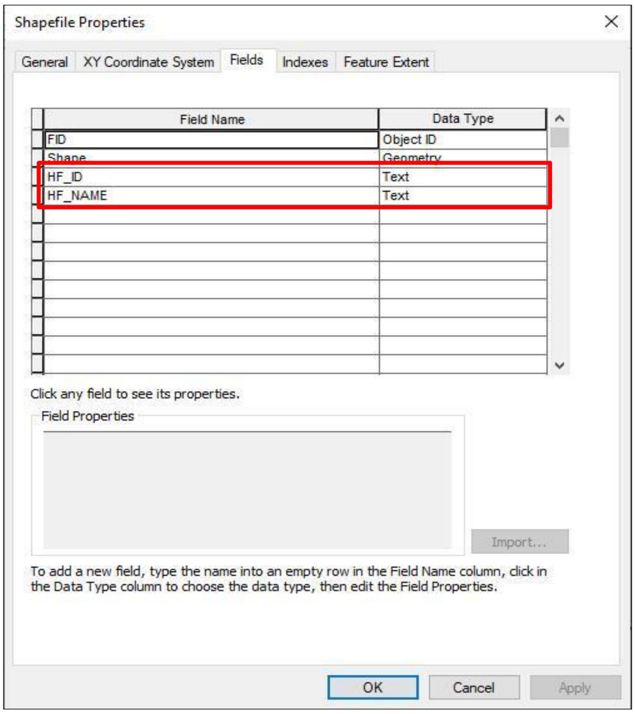


15. In the Catalog window, right click on your shapefile and select *Properties…*



16. In the window that opens, click the Fields tab. This shows the field you will see in the attribute table of the shapefile.

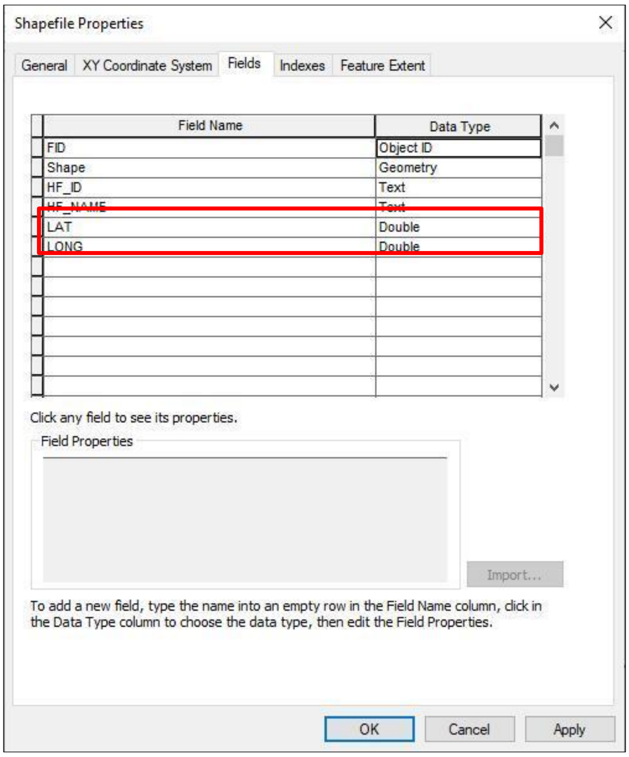
17. We will add fields that will be in the attribute table and will contain the unique code and official name from the master list of the objects being extracted. This will facilitate the joining of the shapefile with the master list once the extraction is done for the shapefile to contain all the needed information for the objects. Type the name of the field you want to add in the Field Name column (e.g., *HF_ID* for the health facility unique code, *HF_NAME* for the official health facility name) and the corresponding type in the Data Type field (e.g., *Text* for both the health facility unique code and name).



Note: The number of characters allowed for the field names in ArcMap is limited to 10. The field names should therefore be abbreviated. It is advisable that the field names be adapted from the master list and/or data dictionary of the object being extracted if they have been structured similar to the data dictionary in the HIS geo-enabling: Guidance on the establishment of a common geo-registry for the simultaneous hosting, maintenance, update, and sharing of master lists core to public health appendices 2 and 3.[18]

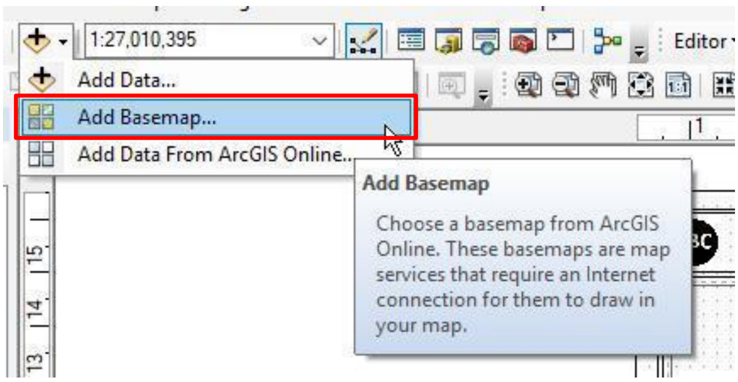
[18] https://healthgeolab.net/DOCUMENTS/Guidance_Common_Geo-registry_Ve1.pdf

18. As you are extracting point-type object in this example, we will add the fields that will later contain the latitude and longitude of the points. Type in *LAT* and *LONG* separately in the Field Name column. Choose *Double* as the Data Type for both. Click *OK*.



Note: The fields containing the latitude and longitude are not necessary when extracting line- or polygon-type objects.

19. Add the basemap by going to *File > Add Data > Add Basemap…* or by clicking the dropdown button beside the *Add Data* button and clicking *Add Basemap…*



Note: An internet connection is needed to add a basemap on ArcMap.

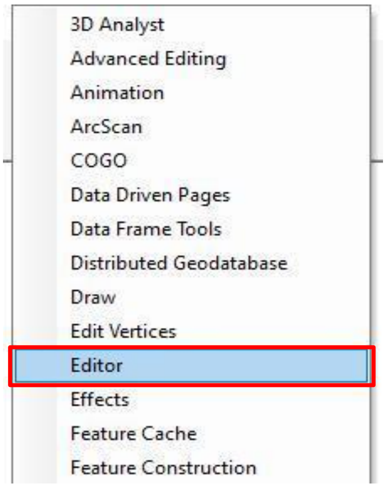
20. In the window that opens, select *Imagery* (or Imagery with Labels). Click *Add*.

21. The basemap will appear in the Table Of Contents (ToC) panel.

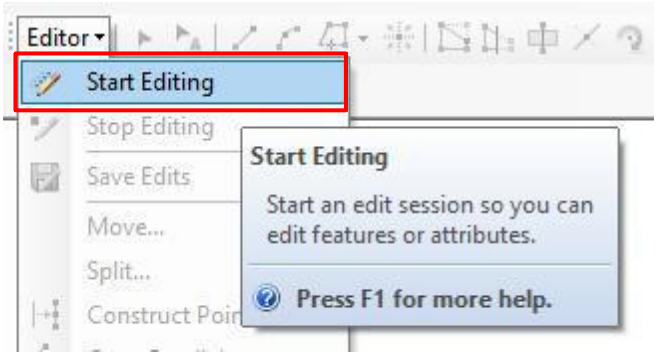
22. Zoom in on the area of interest by using the Zoom In button  from the Tools toolbar. Make sure to zoom in to the required map scale or within the scale range as defined in the data set specifications. You can see the scale in the Standards toolbar.

Note: You may zoom in closer if this allows you to see the object of interest better.

23. To start extracting geospatial, make sure that the Editor toolbar is shown. If not, right click anywhere in the main menu area and the list of available toolbars will appear. Click on *Editor*.



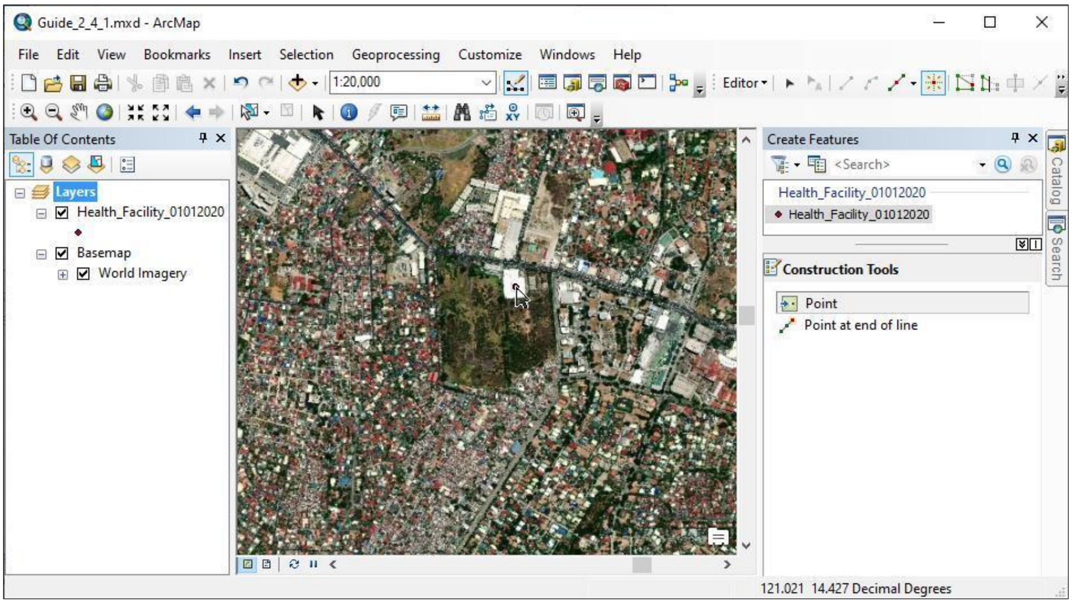
24. The Editor toolbar will now appear in the toolbars area. Start the editing session by clicking on the *Editor* dropdown button and choosing *Start Editing*.



25. Click the *Create Features* button  on the Editor toolbar.

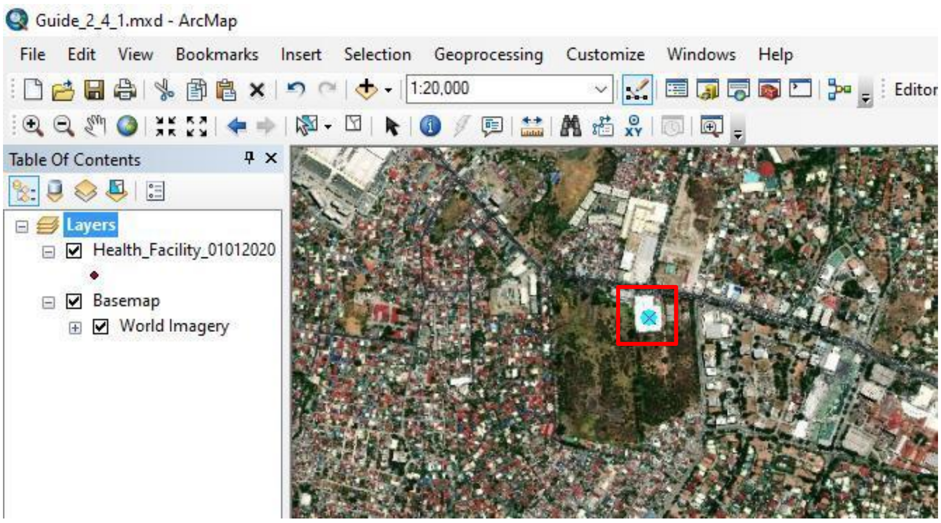
26. The Create Features side panel will open. Click on the name of your shapefile and the available tools for your shapefile will appear in the Construction Tools section below. Click on the correct tool ("Point" for this example).

27. In the map canvas, trace from the basemap the object you want to extract. As you are trying to capture the location of health facilities in this example, a point-type object, just click once on top of the object in the basemap that corresponds to the structure of the health facility.



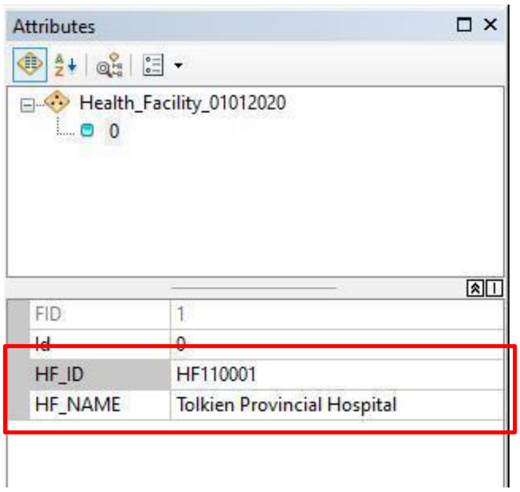
Note: For line-type objects (such as roads, rivers, boat route, etc.), you will have to trace along the length of the object you are extracting, usually along the center line. For polygon-type objects (such as administrative boundaries, operational division boundaries, building footprints, etc.), you will have to trace the boundary of the object you are extracting.

28. The new point you created will appear on the map canvas (highlighted and with an "X" mark).

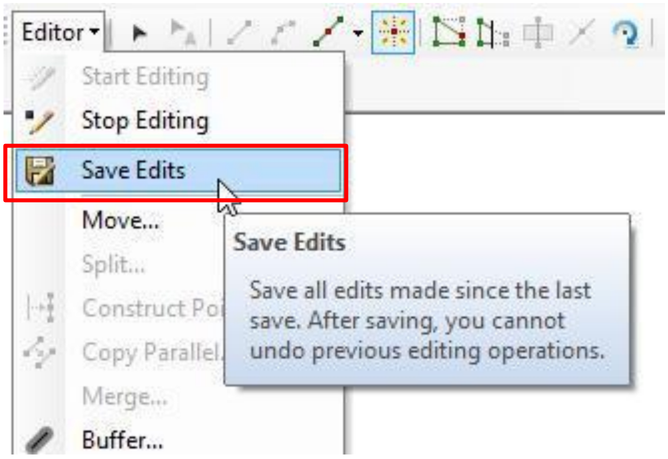


29. Click the *Attributes* button  on the Editor toolbar to be able to input the information regarding the object.

30. The Attributes window opens on the right side of the ArcMap screen. Type in the unique code and official name of the object as taken from the master list. Close the Attributes window.



31. Click the *Editor* dropdown button on the Editor toolbar and select *Save Edits* to save your work.



Note: Although this step can also be done just after every few geographic objects are captured, it is a good practice to do this after every captured object to ensure that the work is saved in case the GIS software or your computer suddenly crashes.

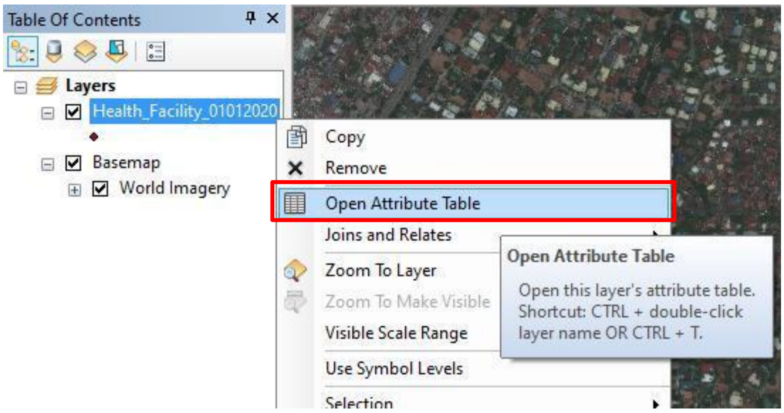
32. Navigate to the next object on the basemap by using the map navigation buttons in the Tools toolbar.

33. Repeat steps 24-32 to extract the remaining geographic objects from the basemap.

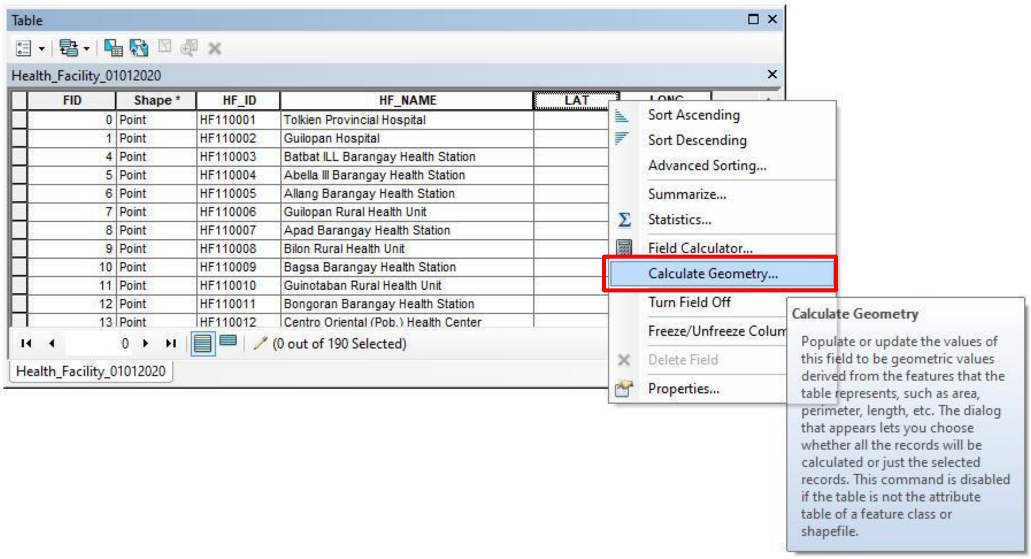
34. Once all geographic objects are extracted, save your edits.

Note: Steps 35-43 are for when extracting point-type objects. If you are extracting line- or polygon-type objects, skip to step 44.

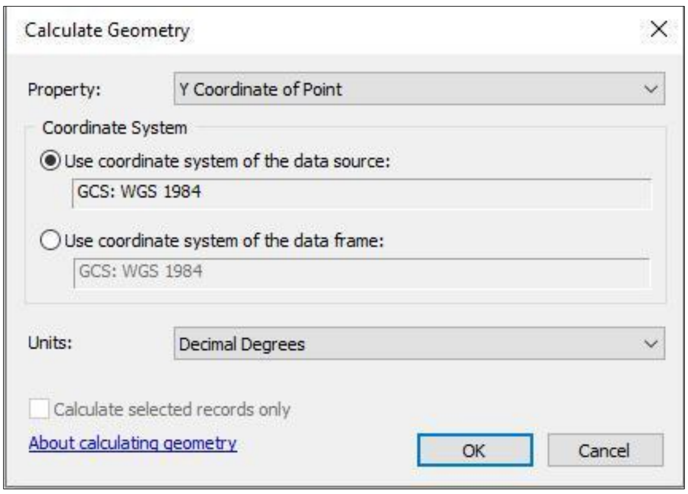
35. In the Table of Content, right click on your point shapefile and choose *Open Attribute Table*.



36. The Attribute Table opens. Right click on the LAT (latitude) field name and choose *Calculate Geometry…*



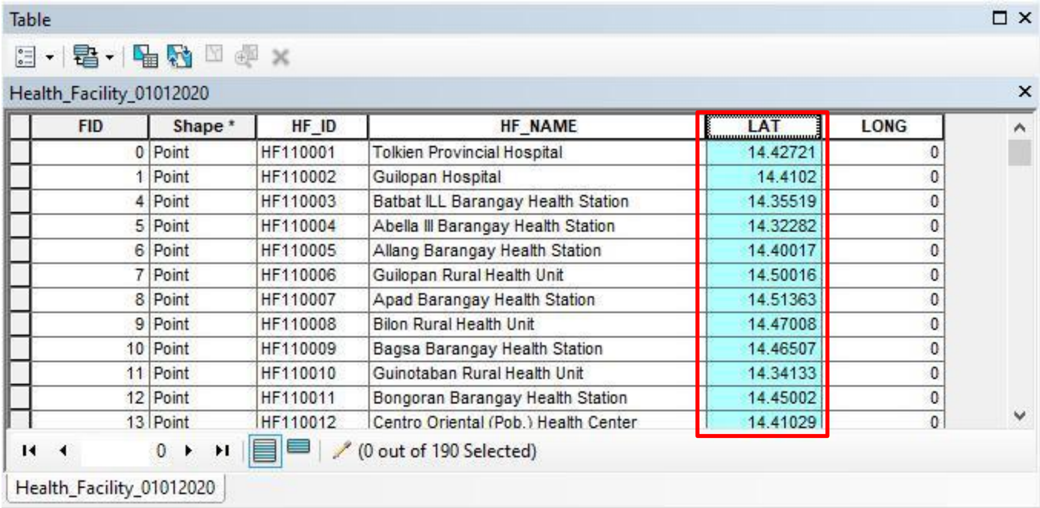
37. The Calculate Geometry window opens. In the Property field, choose *Y Coordinate of Point*.



38. In the Coordinate System section, choose *Use coordinate system of the data source* (which is the coordinate system you assigned to your shapefile in Step 13).

39. In the Units field, choose the appropriate units based on the data specifications (usually decimal degree). Click *OK*.

40. The Attribute Table will now show the calculated latitude for each point.



Table

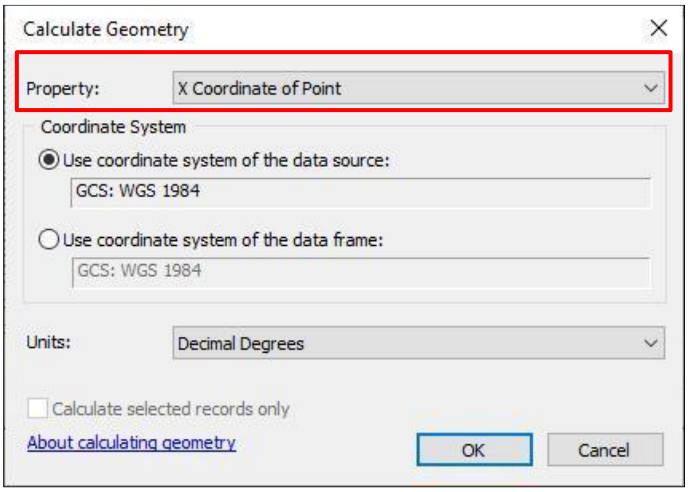
Health_Facility_01012020

FID	Shape *	HF_ID	HF_NAME	LAT	LONG
0	Point	HF110001	Tolkien Provincial Hospital	14.42721	0
1	Point	HF110002	Guilopan Hospital	14.4102	0
4	Point	HF110003	Batbat ILL Barangay Health Station	14.35519	0
5	Point	HF110004	Abella III Barangay Health Station	14.32282	0
6	Point	HF110005	Allang Barangay Health Station	14.40017	0
7	Point	HF110006	Guilopan Rural Health Unit	14.50016	0
8	Point	HF110007	Apad Barangay Health Station	14.51363	0
9	Point	HF110008	Bilon Rural Health Unit	14.47008	0
10	Point	HF110009	Bagsa Barangay Health Station	14.46507	0
11	Point	HF110010	Guinotaban Rural Health Unit	14.34133	0
12	Point	HF110011	Bongoran Barangay Health Station	14.45002	0
13	Point	HF110012	Centro Oriental (Pob.) Health Center	14.41029	0

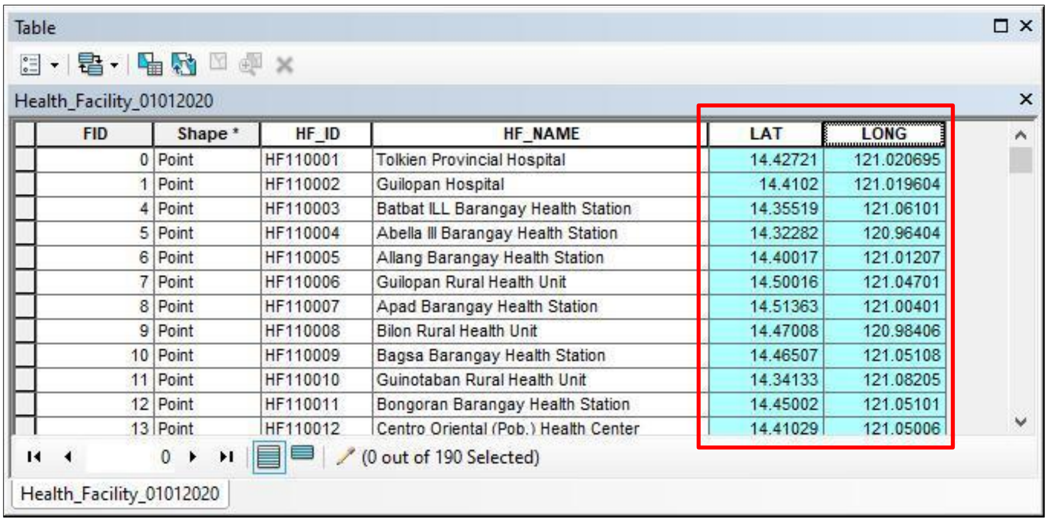
(0 out of 190 Selected)

Health_Facility_01012020

41. Repeat the process for the LONG (longitude) field. Right click on the LONG field name and choose *Calculate Geometry...* This time, choose *X Coordinate of Point* in the Property field. Click *OK*.



42. This calculates the longitude. Now both latitude and longitude of the points are shown.



Table

Health_Facility_01012020

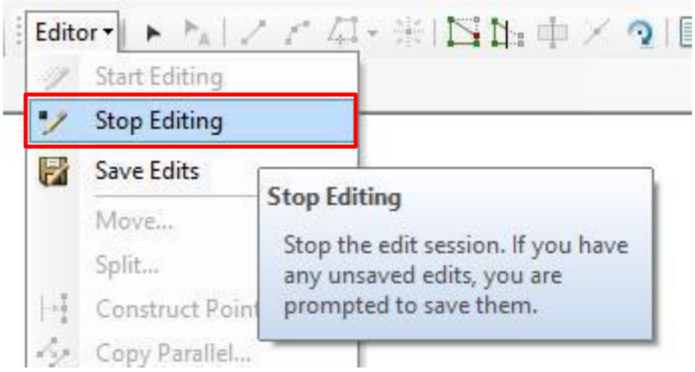
FID	Shape *	HF_ID	HF_NAME	LAT	LONG
0	Point	HF110001	Tolkien Provincial Hospital	14.42721	121.020695
1	Point	HF110002	Guilopan Hospital	14.4102	121.019604
4	Point	HF110003	Batbat ILL Barangay Health Station	14.35519	121.06101
5	Point	HF110004	Abella III Barangay Health Station	14.32282	120.96404
6	Point	HF110005	Allang Barangay Health Station	14.40017	121.01207
7	Point	HF110006	Guilopan Rural Health Unit	14.50016	121.04701
8	Point	HF110007	Apad Barangay Health Station	14.51363	121.00401
9	Point	HF110008	Bilon Rural Health Unit	14.47008	120.98406
10	Point	HF110009	Bagsa Barangay Health Station	14.46507	121.05108
11	Point	HF110010	Guinotaban Rural Health Unit	14.34133	121.08205
12	Point	HF110011	Bongoran Barangay Health Station	14.45002	121.05101
13	Point	HF110012	Centro Oriental (Pob.) Health Center	14.41029	121.05006

(0 out of 190 Selected)

Health_Facility_01012020

43. Save your edits.

44. Stop the editing session by clicking the *Editor* dropdown button on the Editor toolbar and choosing *Stop Editing*.



45. Save your project.

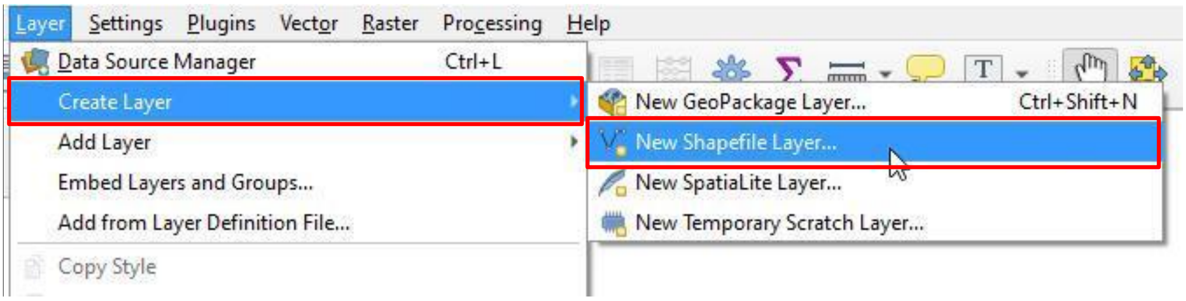
46. The extracted data is now ready for final validation.

A1.2 Using QGIS

1. Open QGIS.
2. Create a new map document by going to the main menu, click *Project > New.*
3. Save your new map document by going to the main menu, click *Project > Save* or by clicking the *Save* button .
4. The Choose a QGIS project file window opens. Browse to the location you want to save your map document in.
5. Type in the name for your map document. Click *Save*.

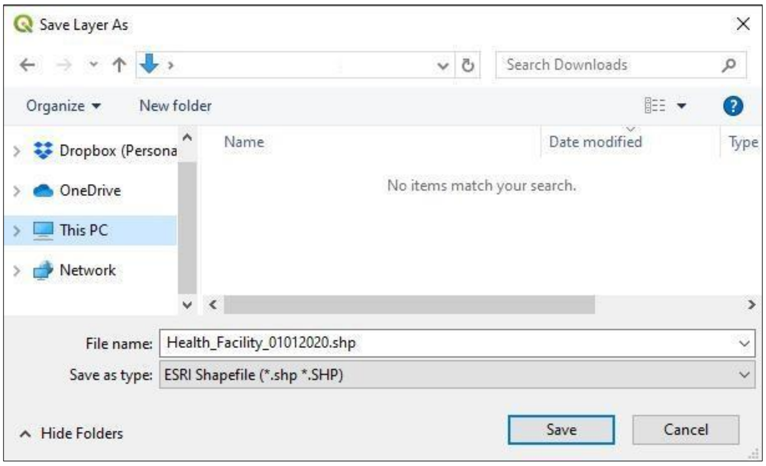
   IMPORTANT NOTE: Remember to periodically save your map document as you go through the next steps. This ensures that your work is saved even if the GIS software or your computer suddenly crashes.

6. You need to create an empty map layer (e.g., shapefile) that will contain the extracted data. From the Layers main menu, choose *Create Layer > New Shapefile Layer...*

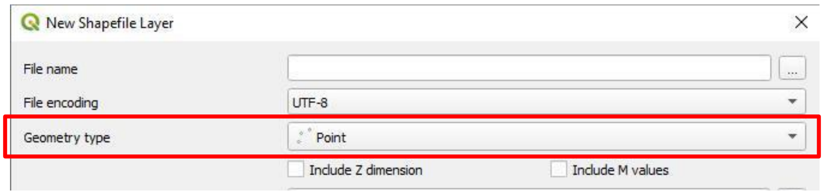


7. The New Shapefile Layer window opens. Click the *Browse* button at the end of the File name field.

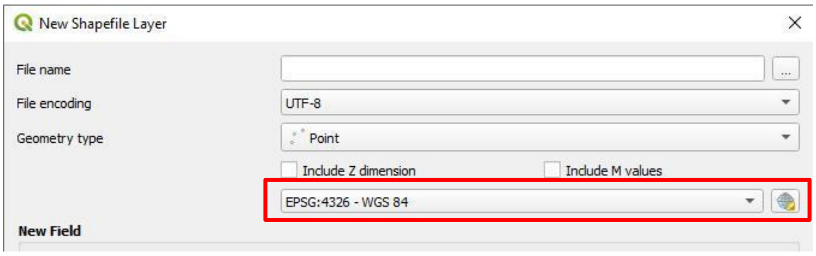
8. The Save Layer As window will open. Navigate to the location of the folder where you want to save your shapefile. In the Field name field, type in the name of your shapefile with the date of the day you created it to help you keep track of your file and for your reference when creating the metadata for the file (e.g., Health_Facility_01012020). Click *Save*.



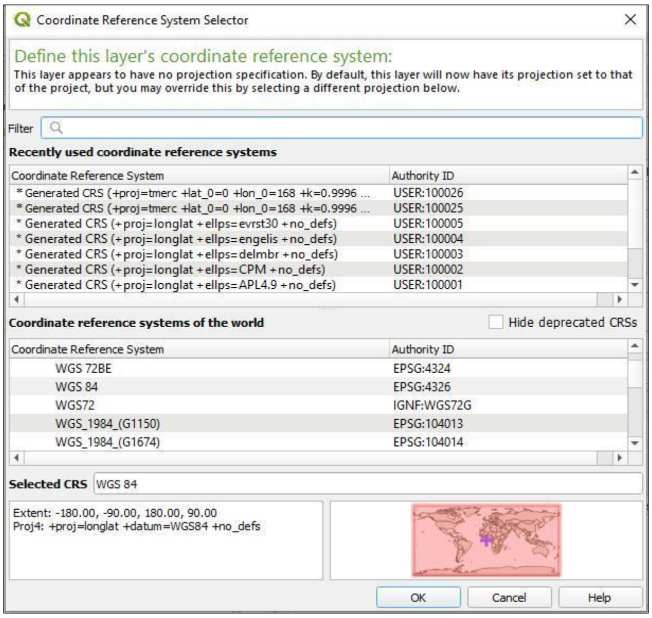
9. Back in the New Shapefile Layer window, choose the correct geometry type (i.e., point, line /polyline, or polygon) for the object being extracted. ("Point" for this example.)



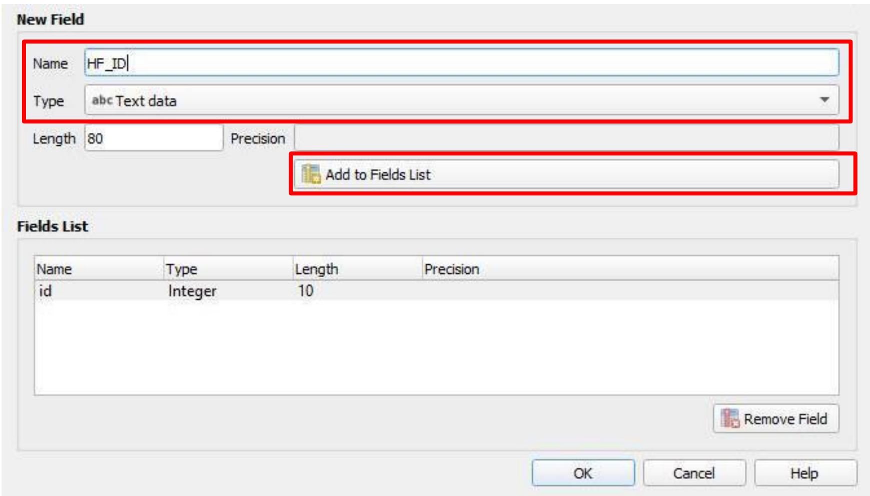
10. Click the Select CRS button  to choose the correct coordinate system as defined in the data set specifications.



11. In the window that opens, scroll up or down in the Coordinate reference systems of the world section to choose the correct coordinate system as defined in the data set specifications. (You may also use the search bar on top of the window.) Click *OK*.



12. We will add fields that will be in the attribute table and will contain the unique code and official name from the master list of the objects being extracted. This will facilitate the joining of the shapefile with the master list once the extraction is done for the shapefile to contain all the needed information for the objects. In the New Field section, type the name of the field you want to add in the Name field (e.g., *HF_ID* for the health facility unique code, *HF_NAME* for the health facility official name) and the corresponding type in the Type field (e.g., Text data for both the health facility unique code and name). Click the *Add to Fields List* button after each entry.

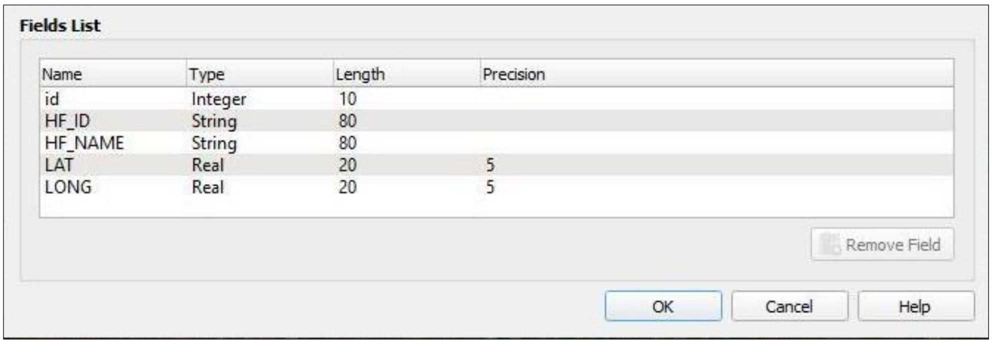


Note: The number of characters allowed for the field name in QGIS is limited to 10. The field names should therefore be abbreviated. It is advisable that the field names be adapted from the master list and/or data dictionary of the object being extracted if they have been structured similar to the data dictionary in the HIS geo-enabling: Guidance on the establishment of a common geo-registry for the simultaneous hosting, maintenance, update and sharing of master lists core to public health appendices 2 and 3.[19]

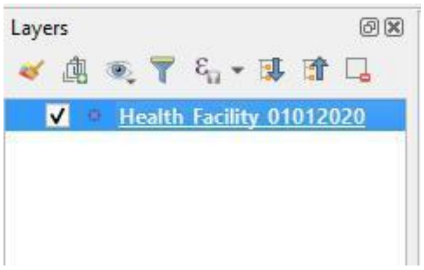
13. As you are extracting point-type object in this example, we will add the fields that will later contain the latitude and longitude of the points. Type in *LAT* and *LONG* separately in the Name field. Choose *Decimal number* as the Type for both and specify the Precision as *5*. Click the *Add to Fields List* button after each entry.

    Note: The fields containing the latitude and longitude are not necessary when extracting line- or polygon-type objects.

14. The newly created fields will now appear in the Fields List section. Click *OK*.



15. Your new shapefile will appear in the Layers panel.



[19] https://healthgeolab.net/DOCUMENTS/Guidance_Common_Geo-registry_Ve1.pdf

16. To add a basemap in QGIS, go to the Browser panel. Right click on *XYZ Tiles* then choose *New Connection…*

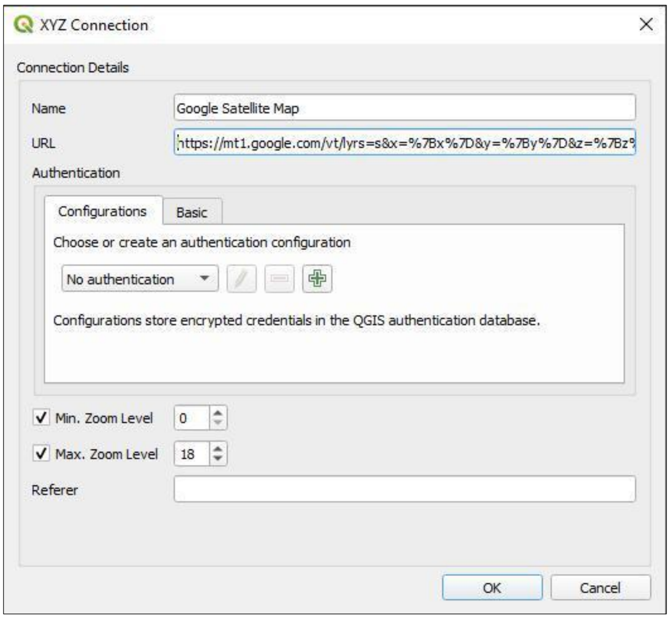
17. In the window that opens, type the name of the basemap you want to use in the Name field. In the URL field, use the URL of the basemap you want to use from the following:

Google Satellite Map:
https://mt1.google.com/vt/lyrs=s&x=%7Bx%7D&y=%7By%7D&z=%7Bz%7D

Esri Satellite Map:
https://server.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer/tile/%7Bz%7D/%7By%7D/%7Bx%7D

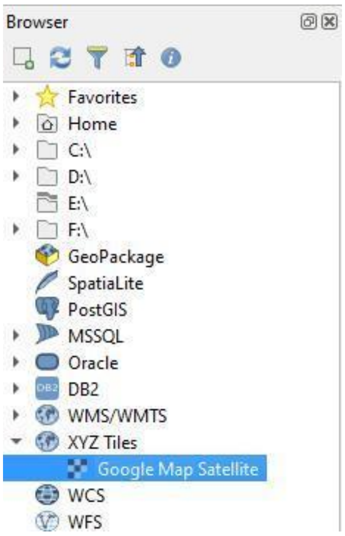
Bing Virtual Earth:
http://ecn.t3.tiles.virtualearth.net/tiles/a%7Bq%7D.jpeg?g=1

18. Do not change the other default settings. Click *OK*.



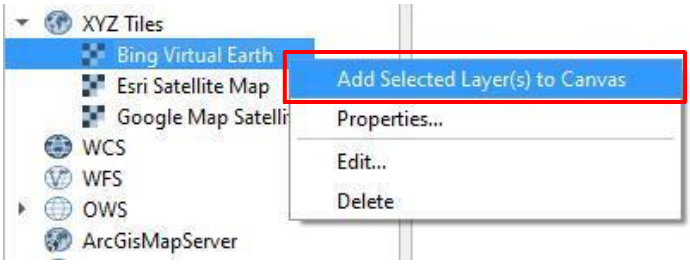
Note: An internet connection is needed to add a basemap on QGIS.

19. The basemap will now appear in the Browser under XYZ Tiles.

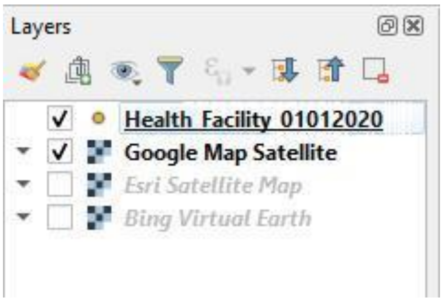


20. If using more than one basemap, follow steps 16-18 to add the other basemap/s.

21. Add the basemap to the map canvas by right clicking the basemap name and *choosing Add Selected Layer(s) to Canvas*. Or by double-clicking on the basemap.



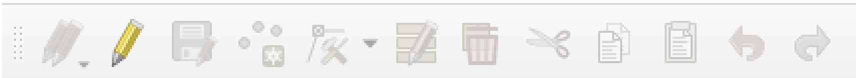
22. The basemap/s will appear in the Layers panel. Click and hold to drag it below the shapefile you created so the shapefile will appear on top of the basemap in the map canvas.

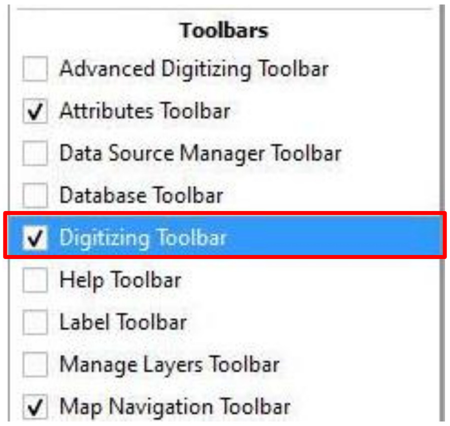


23. Zoom in on the area of interest by using the Zoom In button  from the Map Navigation toolbar. Make sure to zoom in to the required map scale or within the scale range as defined in the data set specifications. You can see the scale in the Status bar at the bottom of the QGIS window.

Note: You may zoom in closer if this allows you to see the object of interest better.

24. To start extracting geospatial data, make sure that the Digitizing Toolbar is shown. If not, right click anywhere in the main menu area. The list of available toolbars will appear. Click on *Digitizing Toolbar*.



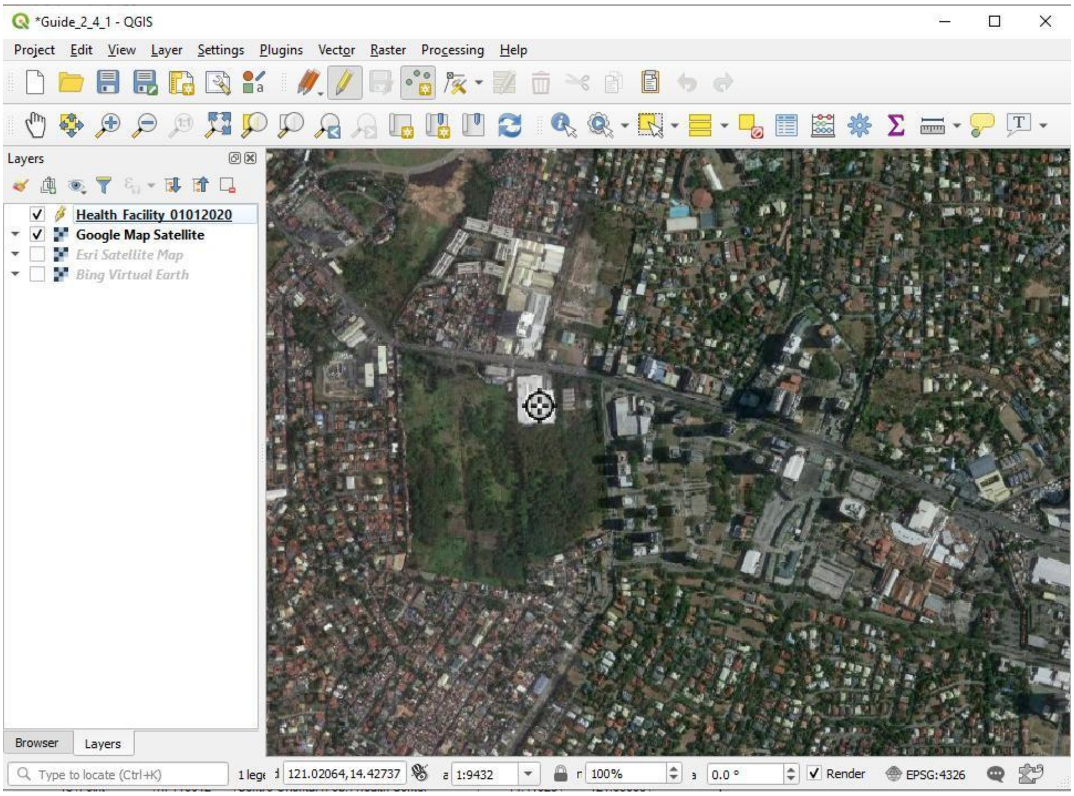


25. The Digitizing Toolbar will now appear in the toolbars area. Start the editing session by clicking on the *Toggle Editing* button .

26. Click the *Add Point Feature* button  on the Digitizing Toolbar.

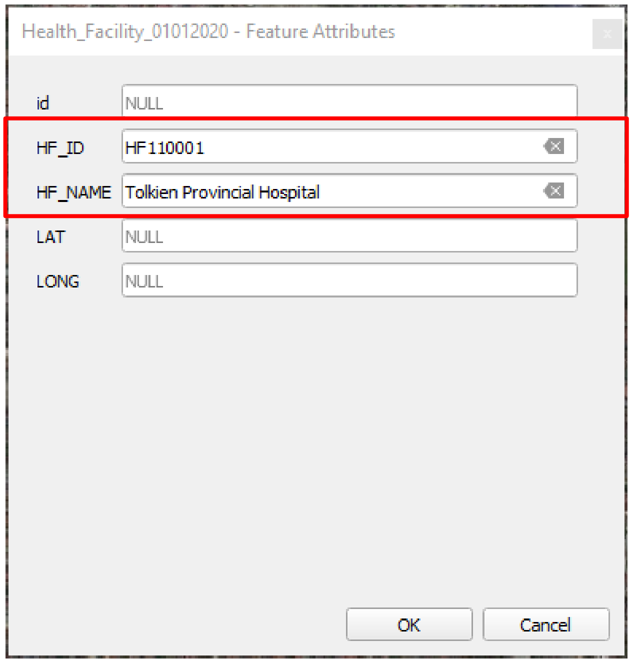
Note: This button becomes the *Add Line Feature*  or *Add Polygon Feature*  when extracting line- or polygon-type object, respectively.

27. In the map canvas, trace from the basemap the object you want to extract. As you are trying to capture the location of health facilities in this example, a point-type object, just click once on top of the object in the basemap that corresponds to the structure of the health facility.

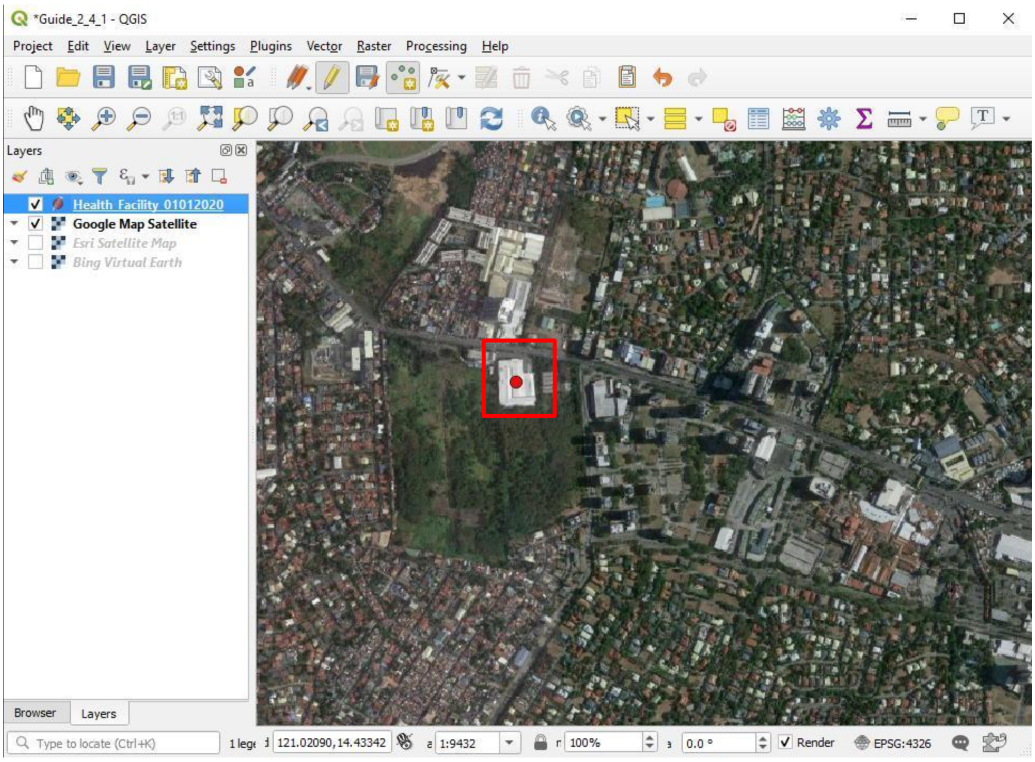


Note: For line-type objects (such as roads, rivers, boat route, etc.), you will have to trace along the length of the object you are extracting, usually along the center line. For polygon-type objects (such as administrative boundaries, operational division boundaries, building footprints, etc.), you will have to trace the boundary of the object you are extracting. You can learn more about creating line-and polygon-type objects from this link: https://docs.qgis.org/3.4/en/docs/training_manual/create_vector_data/create_new_vector.html

28. The Feature Attributes window will open. Type in the unique code and official name of the object as taken from the master list in the HF_ID and HF_NAME fields, respectively. Click *OK*.



29. The new point you created will appear on the map canvas.



30. Click the *Save Layer Edits* button  on the Digitizing toolbar to save your work.

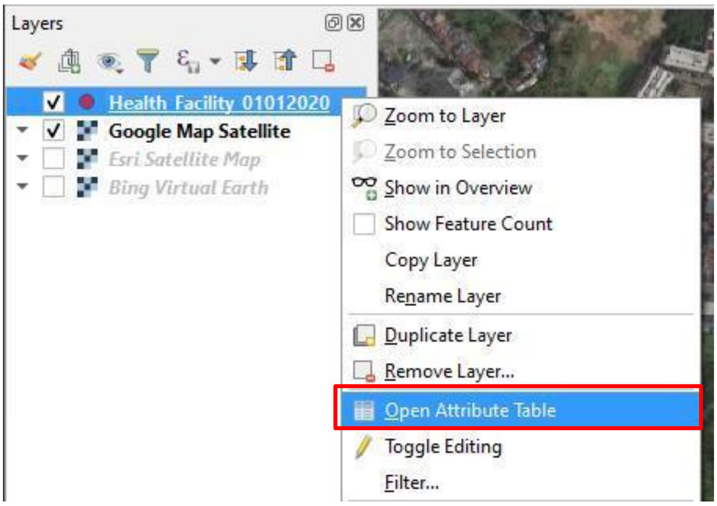
Note: Although this step can also be done just after every few objects are captured, it is a good practice to do this after every captured object to ensure that the work is saved in case the GIS software or your computer suddenly crashes.

31. Navigate to the next object on the basemap by using the buttons in the Map Navigation toolbar.

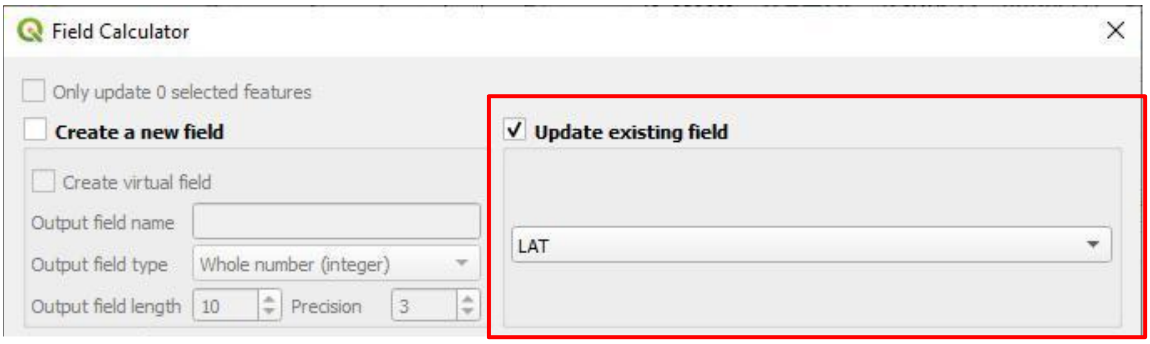
32. Repeat steps 26-31 to extract the remaining geographic objects from the basemap.
33. Once all geographic objects are extracted, save your edits.

Note: Steps 34-44 are for when extracting point-type objects. If you are extracting line- or polygon-type objects, skip to step 45.

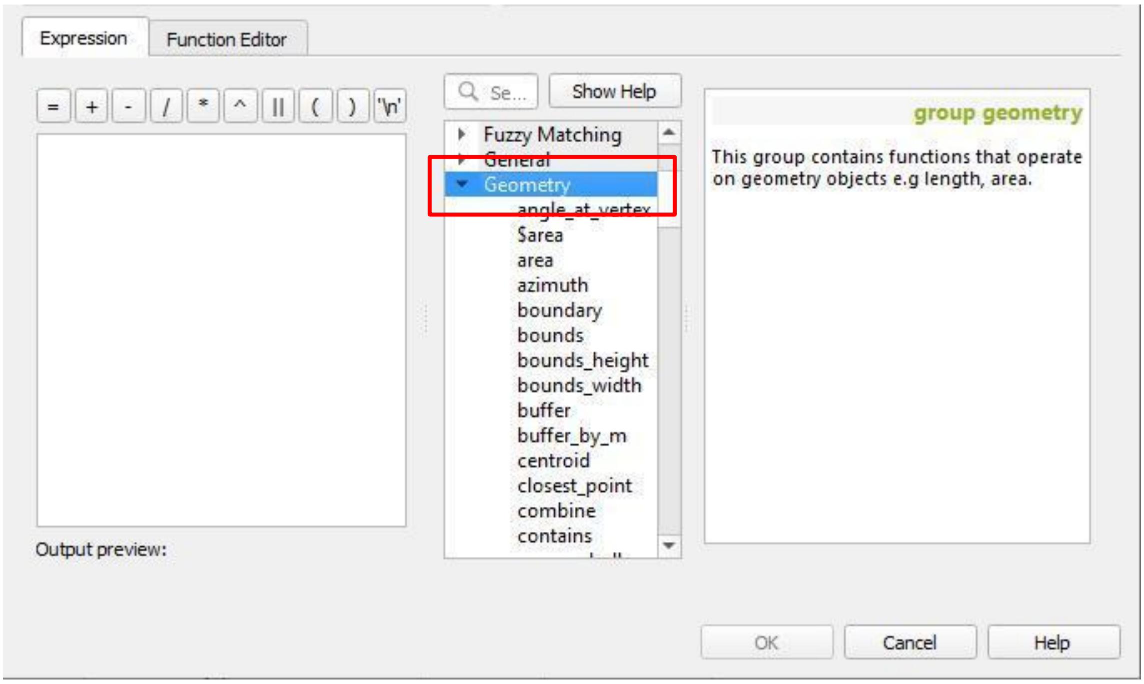
34. In the Layers panel, right click on your point shapefile and choose *Open Attribute Table*.



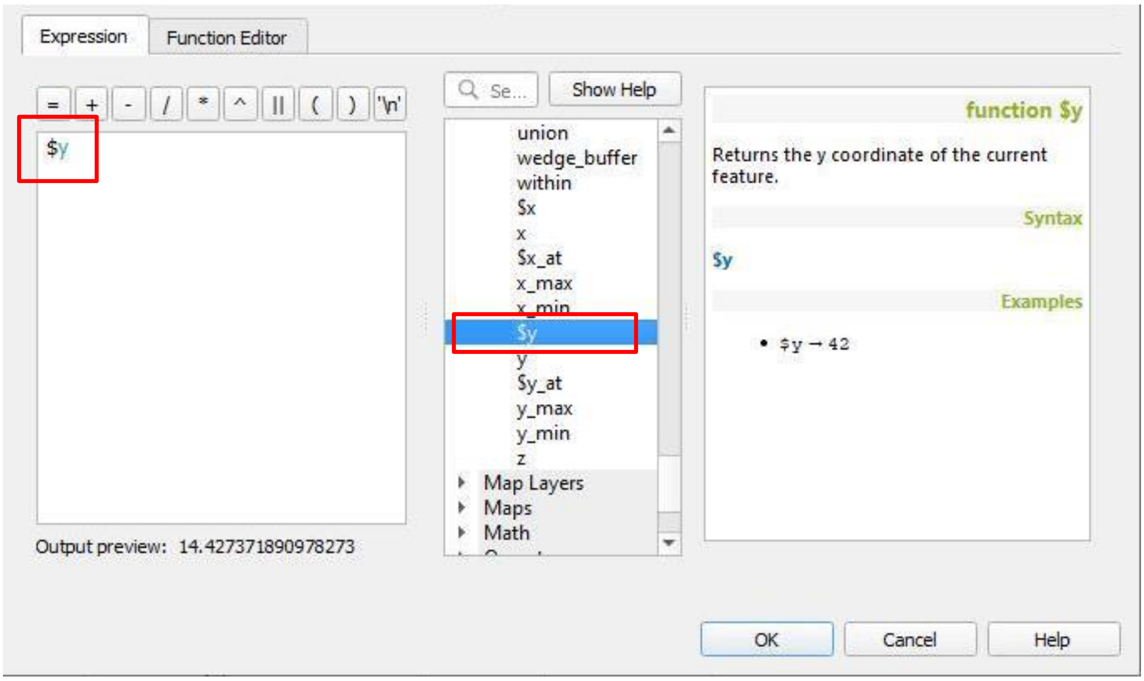
35. The Attribute Table opens. Click the *Open field calculator* button  .
36. The Field Calculator window opens. Check *Update existing field* and choose *LAT* (latitude) from the dropdown menu below.



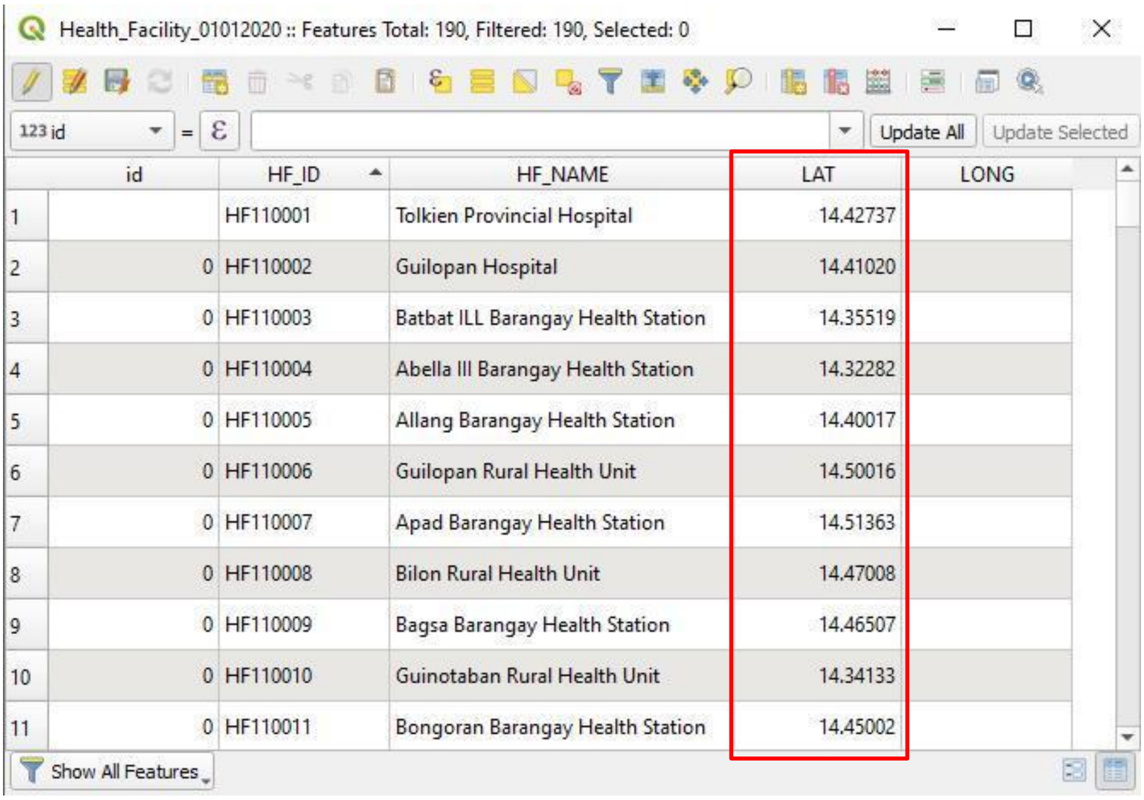
37. In the Expression section, expand Geometry by clicking the small triangle beside the name.



38. Scroll down and double click on *$y* to compute for the latitude. This will appear in the blank section on the left side. Click *OK*.



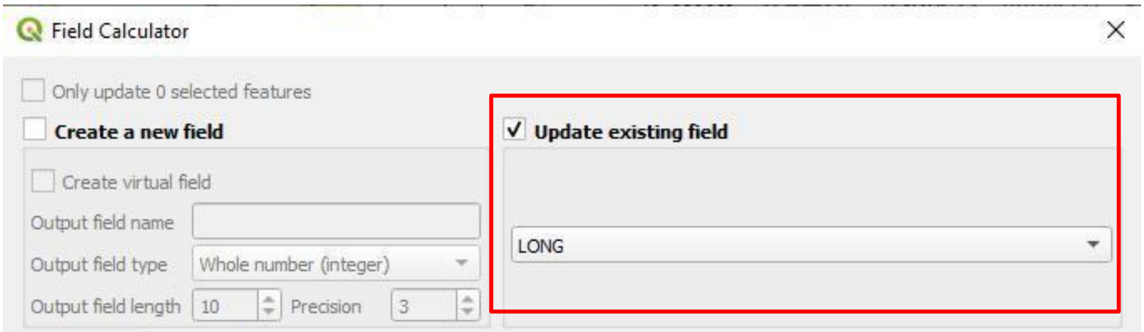
39. The Attribute Table will now show the calculated latitude for each point.



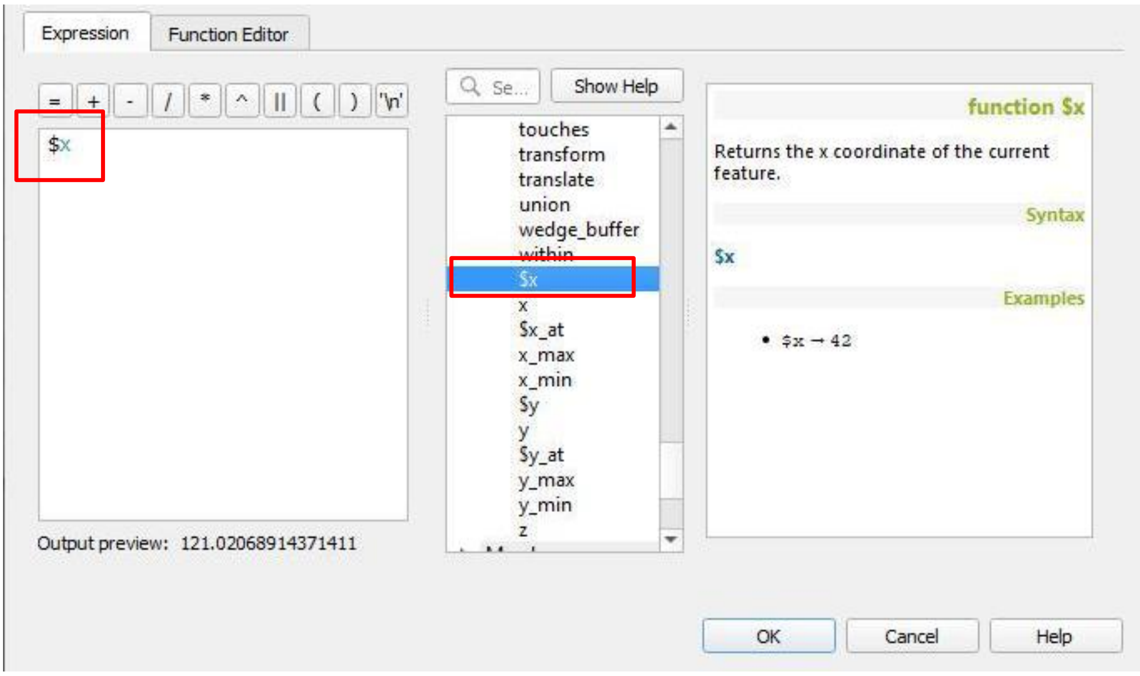
40. Repeat the process for the LONG (longitude) field. Click the *Open field calculator* button


.

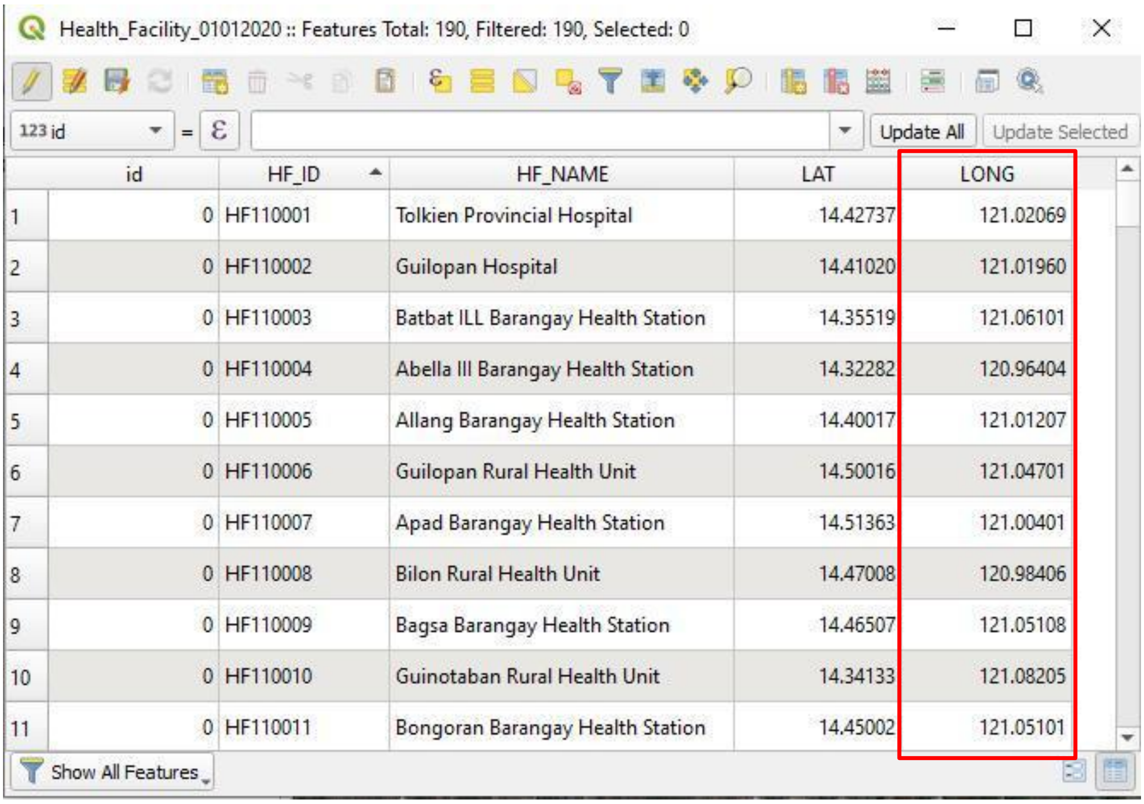
41. This time, choose *LONG* as the field to update.



42. In the Expression section, expand Geometry by clicking the small triangle beside the name. Scroll down and double click on *$x* to compute for the longitude. This will appear in the blank section on the left side. Click *OK*.



43. The Attribute Table will show the calculated longitude for each point. Now both latitude and longitude for each point are shown.



44. Save your edits either by clicking the *Save edits* button in the Attribute Table or clicking the *Save Layer Edits* button  on the main QGIS window.

45. Stop the editing session by clicking the *Toggle Editing* button  on the Digitizing Toolbar.

46. Save your project.

47. The extracted data is now ready for final validation.

Annex 2 – Process to be followed when extracting data from Google My Maps

This annex describes the steps to be followed when extracting data through online extraction application particularly Google My Maps.

The example used for the following sections is the extraction of point-type object.

This process is in two parts:

1. Creating an empty attribute table in Google Drive
2. Extracting the geographic object in Google My Maps

A Google account is required to be able to use Google My Maps and Google Drive. Internet connection is also needed.

While My Maps can also be used as a mobile app in an Android device, this annex focuses on the process followed when using My Maps on a desktop or laptop computer.

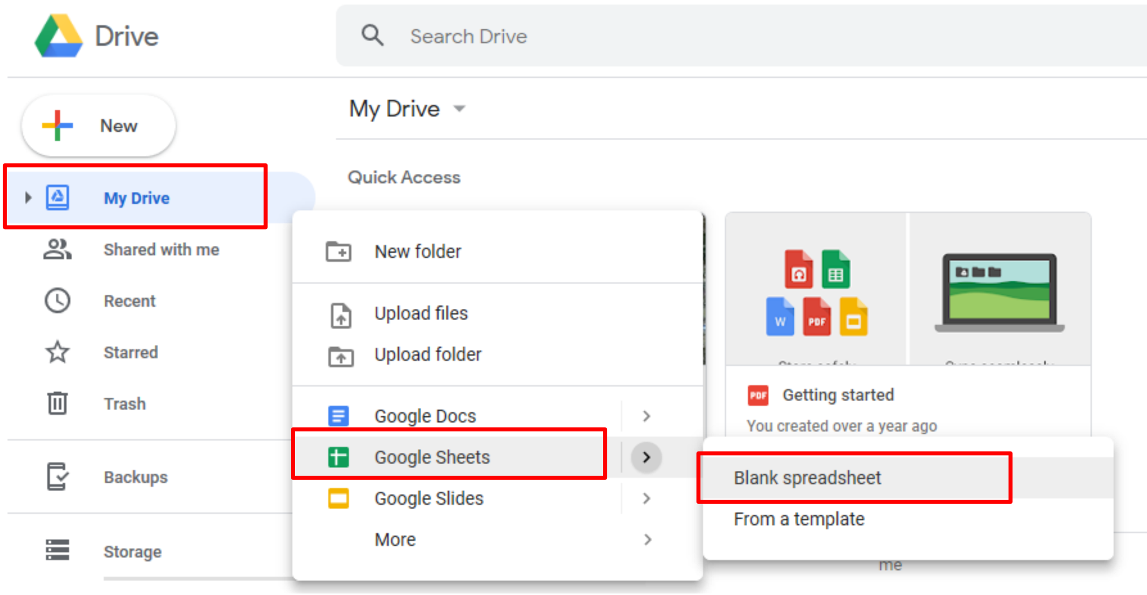
Please note that the process described here covers the extraction of geographic coordinates from Google Maps (point-type feature). When extracting line- or polygon-type object, the LAT and LONG fields will not be populated (nor needed). The extracted data can only be acquired through the download of the KML/KMZ file.

The following sections detail the two parts of the process.

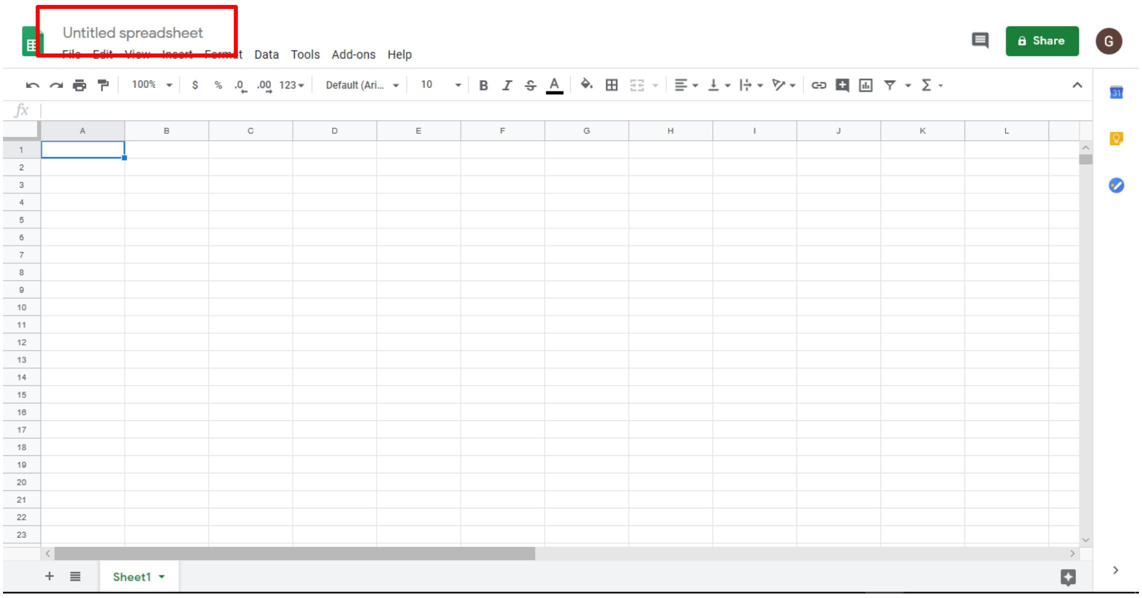
Part 1. Creating an empty attribute table in Google Drive

The following steps should be followed in order to create an empty attribute table in Google drive:

1. On your browser, go to the Google Drive page: https://drive.google.com/
2. Sign in using your Google account credentials.
3. You will be directed to the Google Drive page. Right click on *My Drive* then choose *Google Sheets > Blank spreadsheet*.



4. A blank spreadsheet opens. Click on *Untitled spreadsheet* to enter a title for your spreadsheet which will be the name of the layer you going to extract (e.g., Health Facility).



5. Type the following in the mentioned cells:

   Cell A1: HF_ID

   Cell B1: HF_NAME

   Cell C1: LAT

   Cell D1: LONG

6. In the next row, you will enter a sample entry for each field in your table:

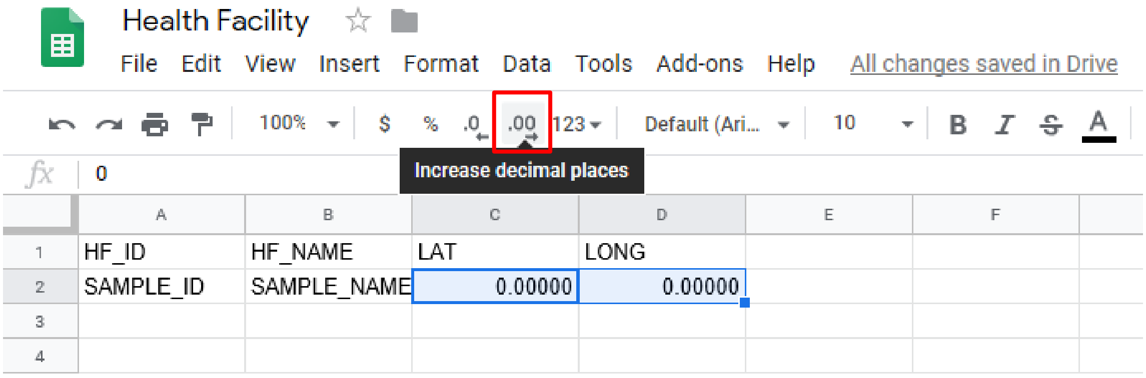
   Cell A2: SAMPLE_ID

   Cell B2: SAMPLE_NAME

   Cell C2: 0.00000

   Cell D2: 0.00000

7. The entries for Cells C2 and D2 (0.00000) will normally just show up as "0". To add the decimal places, select Cells C2 and D2 then click the *Increase decimal places* button repeatedly until there are 5 decimal places.

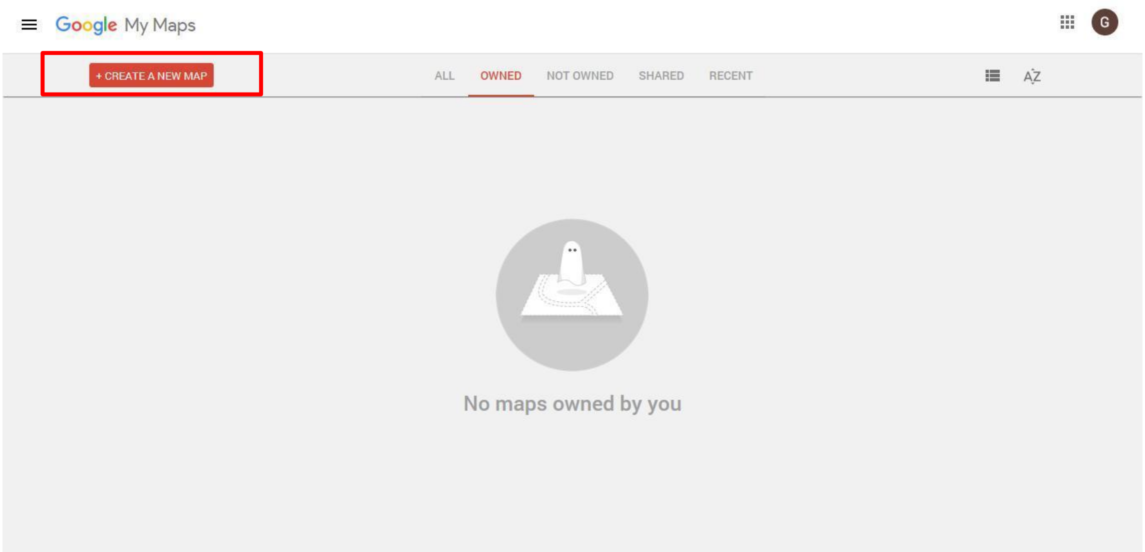


8. Your file and all changes you make on it are automatically saved by and in Google Drive.

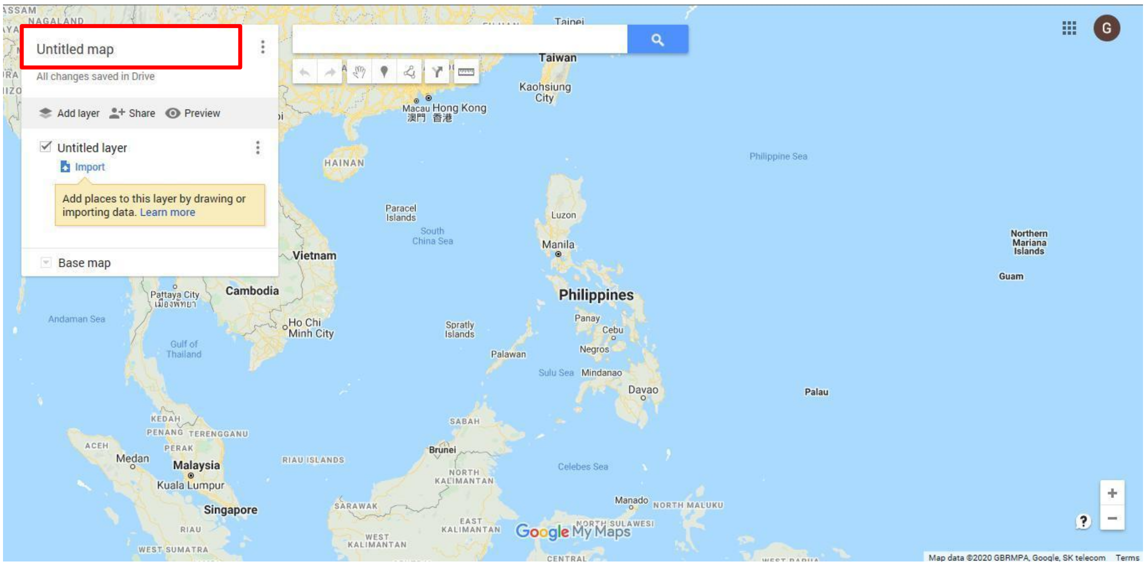
Part 2: Extracting the geographic object in Google My Maps

The following steps are to be followed in order to extract the geographic objects in Google My Maps:

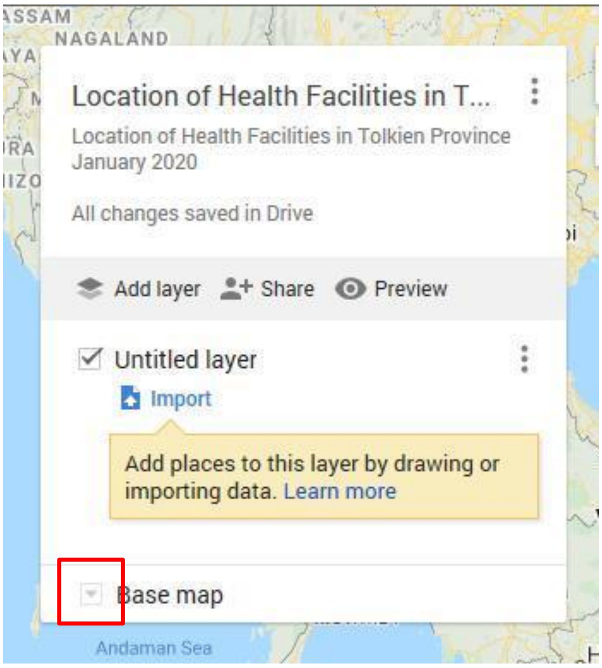
1. On your browser, go to the Google My Maps page: https://mymaps.google.com/
2. Sign in using your Google account credentials.
3. You will be directed to the Google My Maps main page. Click on *+Create a new map* button.



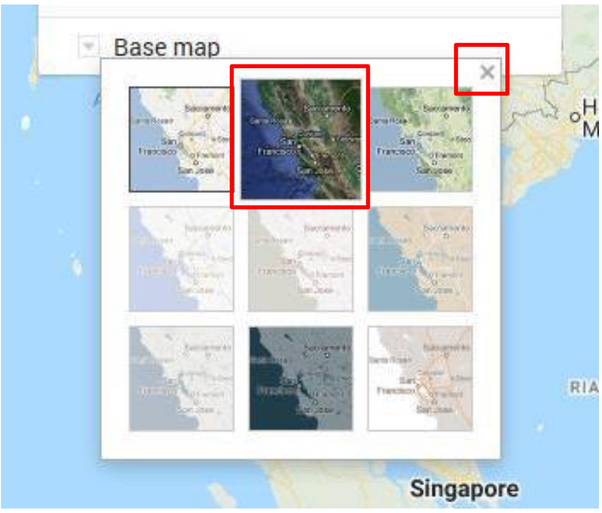
4. This will open the map page. Click on the *Untitled map* text to enter a title and description for your map (e.g., Location of Health Facilities in Tolkien Province). This will help you identify your map from others you might create in the future. Click *Save*.



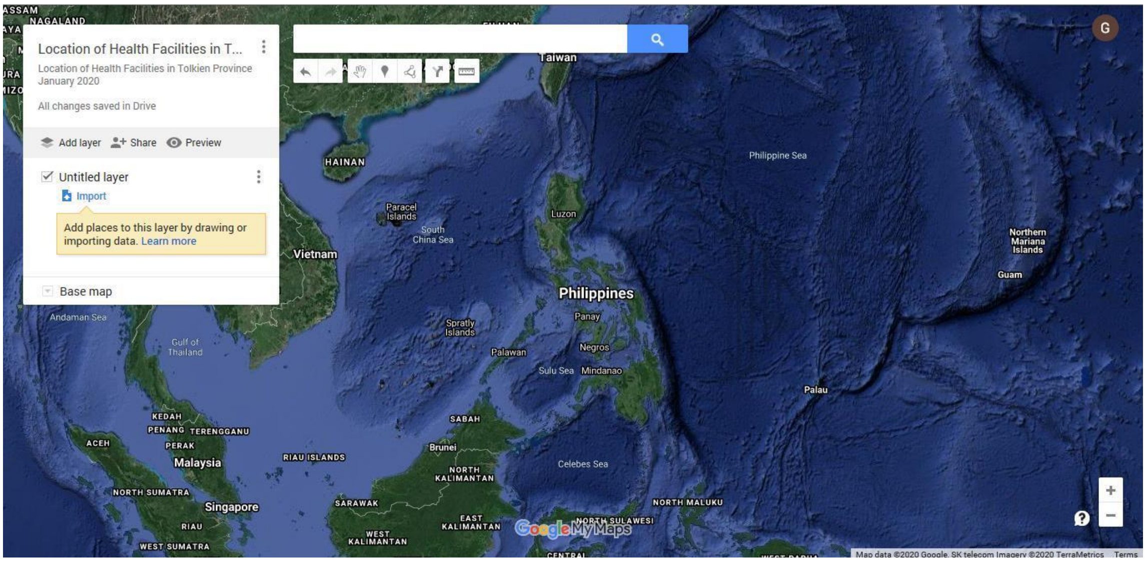
5. Click on the dropdown arrow beside Base map.



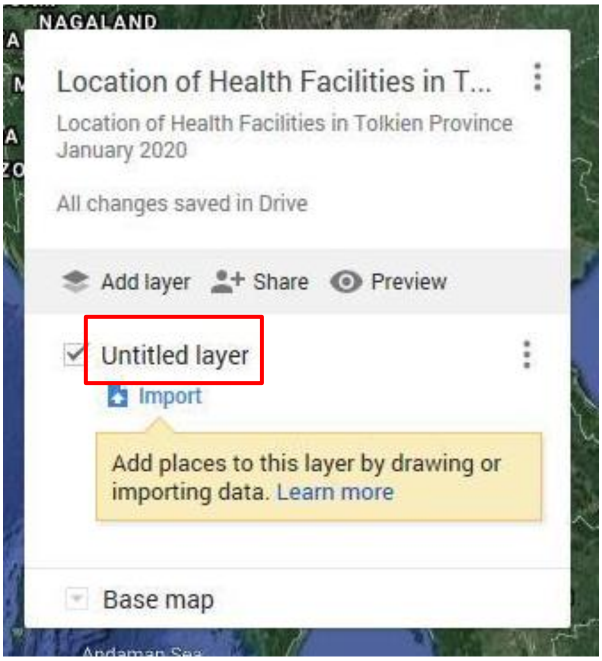
6. This will show the choices of available basemaps. Click on the “Satellite” basemap.



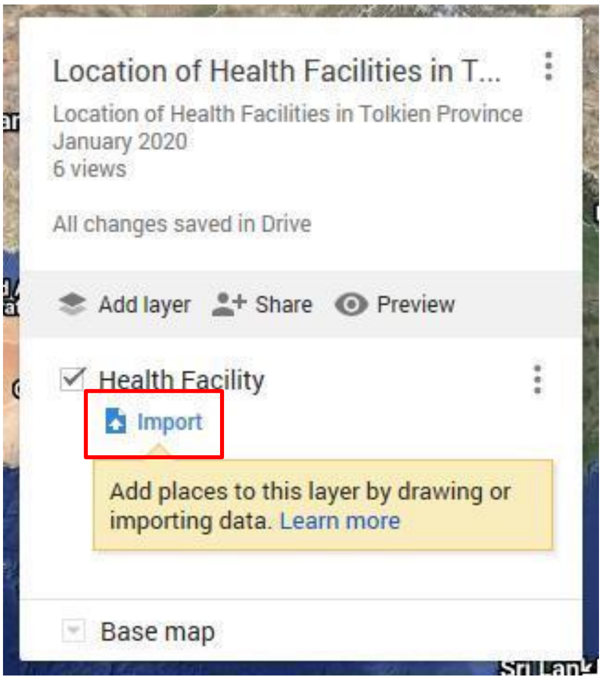
7. Close the basemap choices by clicking the “X” on the top right corner. Your map will now show the Google satellite image as basemap.



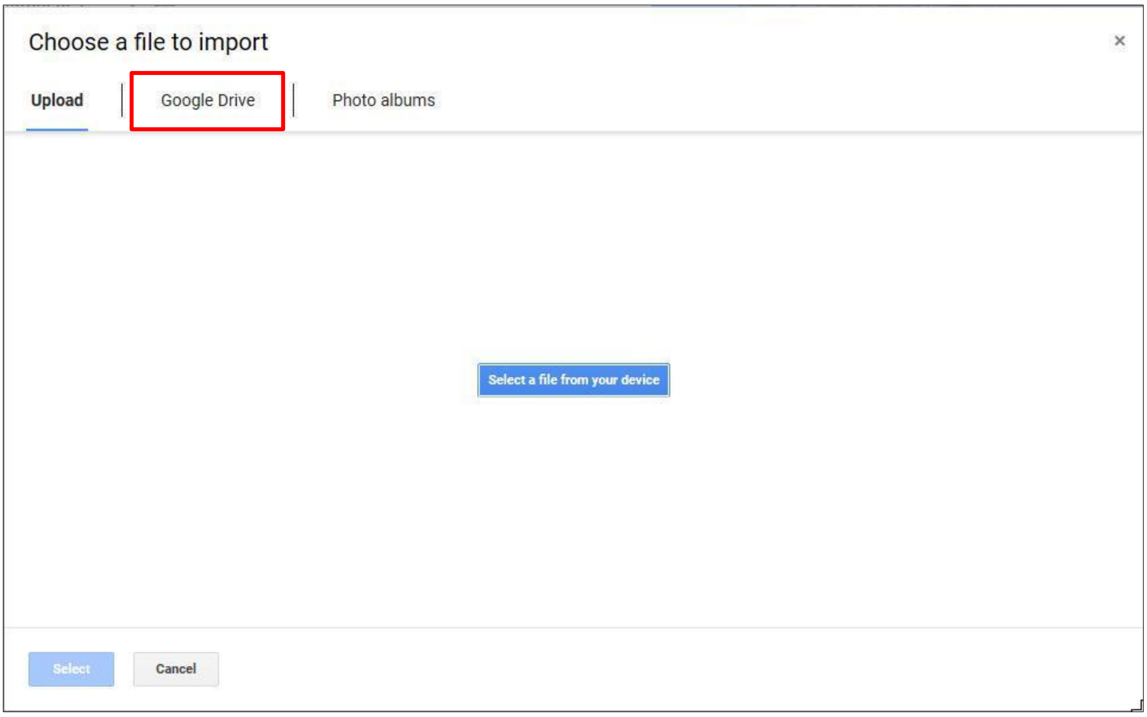
8. Click on *Untitled layer* to enter the name of the layer/geographic object you are going to extract (e.g., Health Facility). Click *Save*.



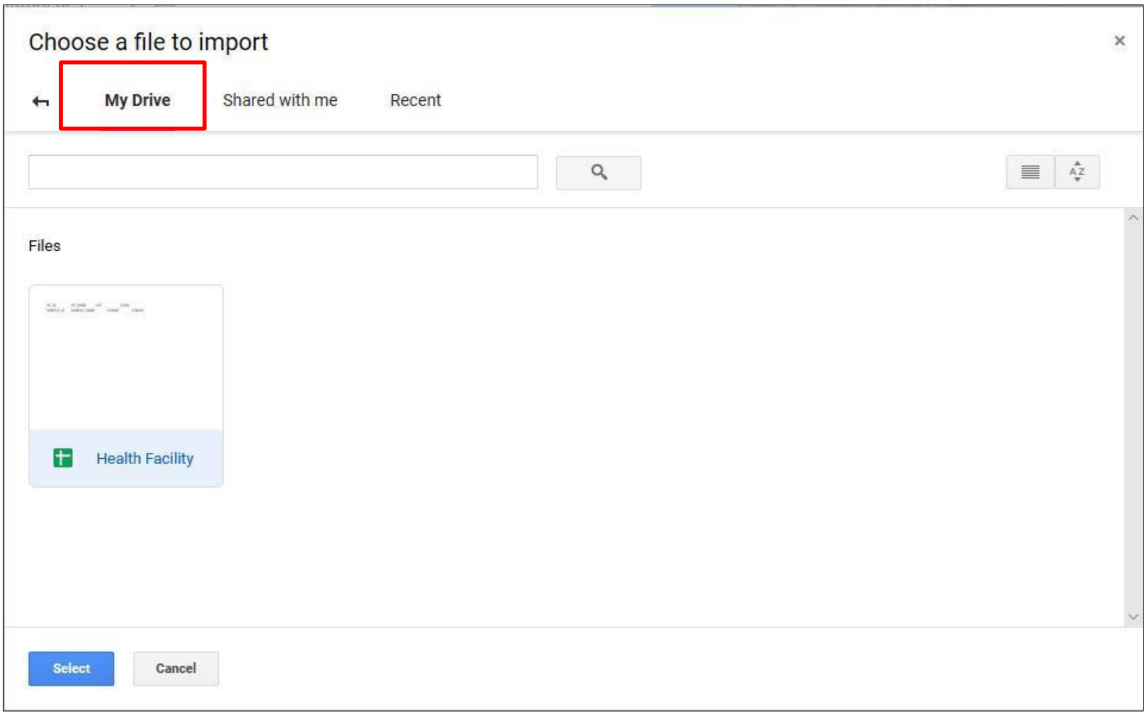
9. Click *Import* under the layer name.



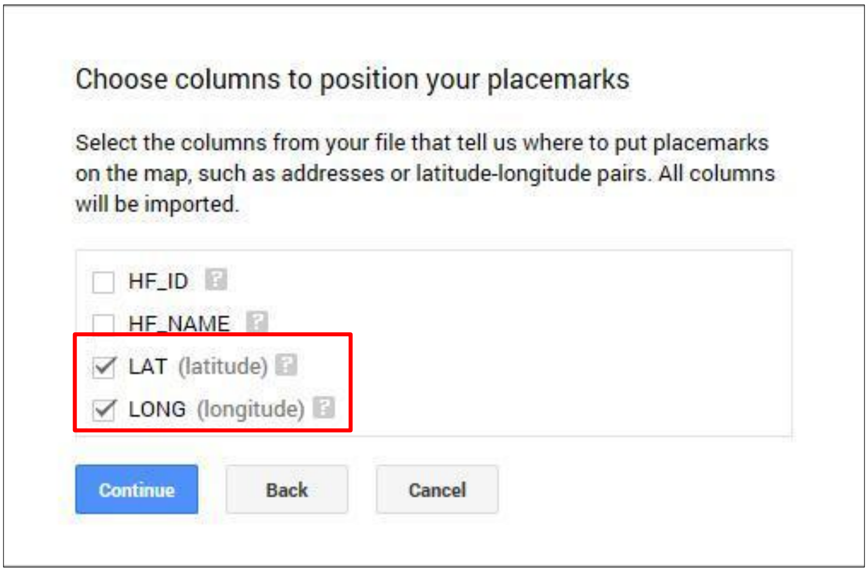
10. A window will open to help you choose the file to import. Choose the *Google Drive* tab.



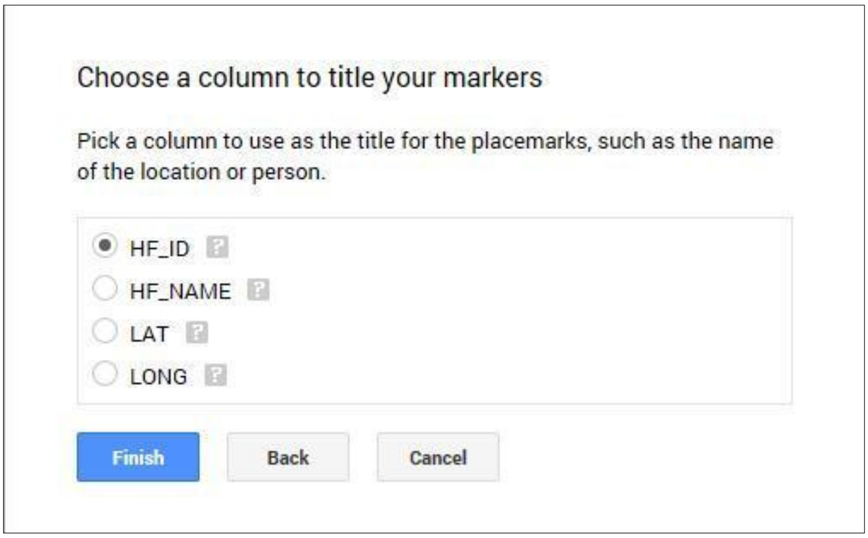
11. In the *My Drive* tab, find and click the table you created in Part 1. (You may also use the search bar.) Click *Select*.



12. In the new window that opens, make sure that LAT (as latitude) and LONG (as longitude) fields are checked. These fields are where the geographic coordinates of the points will be saved once you start the extraction. Click *Continue*.

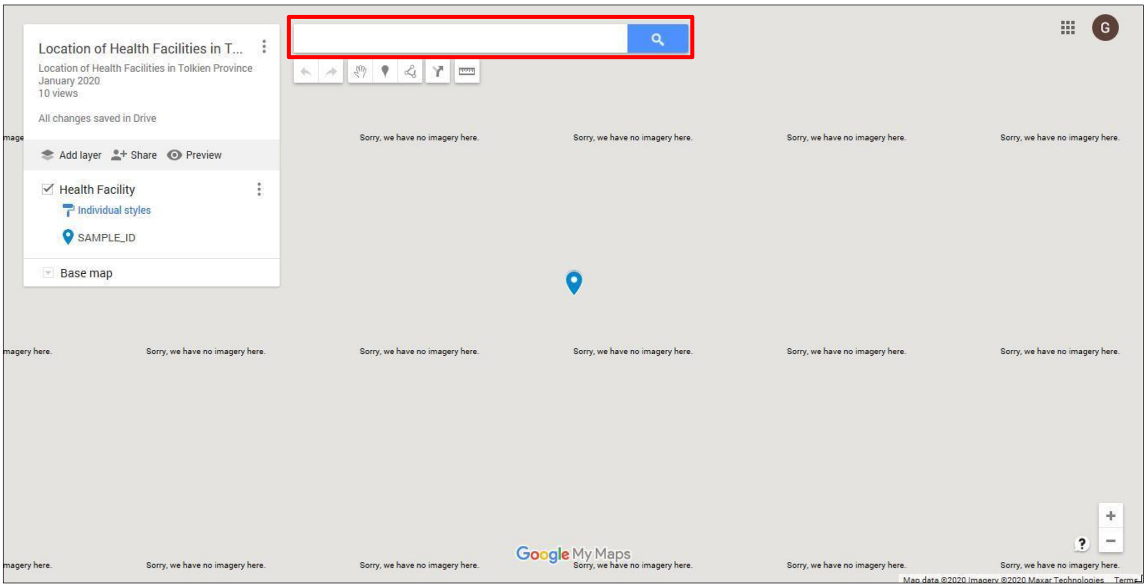


13. Choose a column which entry will serve as the title for each individual point (e.g., HF_ID). Click *Finish*.

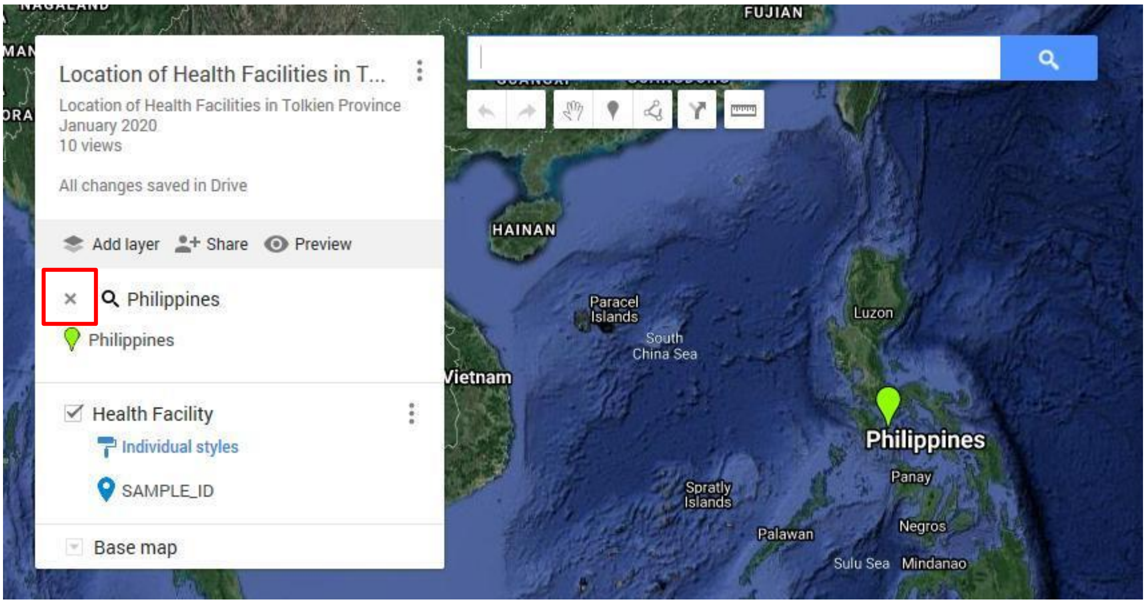


The table you created is now the attribute/data table of the layer that will contain the extracted geographic object.

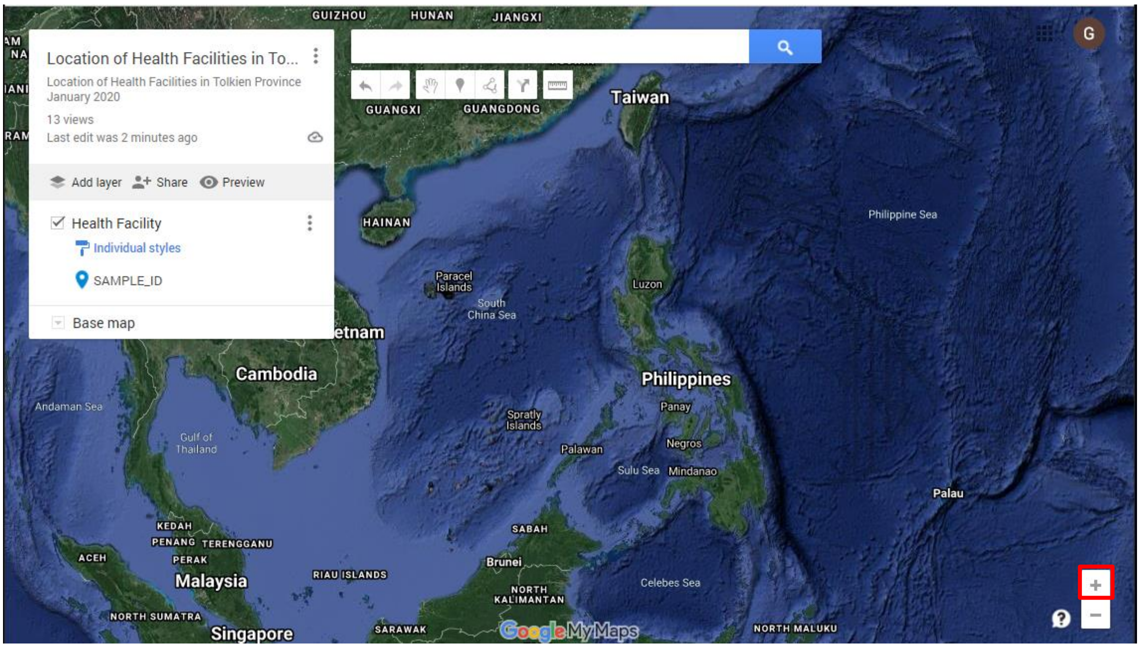
14. In case the basemap disappears, type your area of interest (or any location: country, city, town, etc.) in the search bar and press *Enter*.



15. Once the basemap reappears, you may close the search result by clicking the “X” mark.

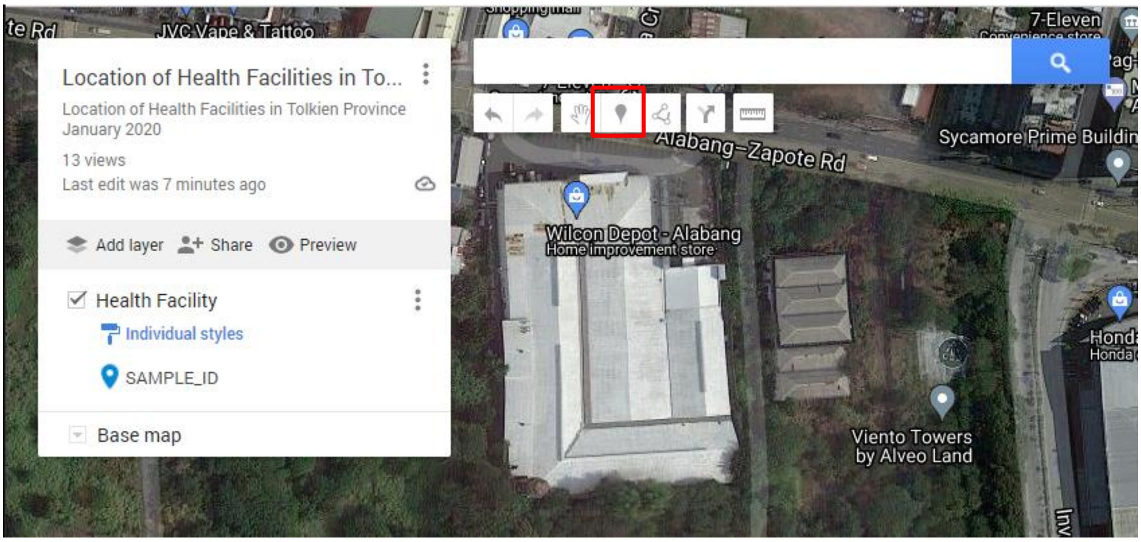


16. Zoom in on the first geographic object you want to extract by clicking the "+" button on the lower right corner of the window or using the mouse wheel. Use your mouse to pan the map by clicking, holding, and dragging the map to help center the object you are zooming in on the screen.



Note: The Google satellite map in this platform does not show the map scale. Zoom in until you are able to clearly identify the geographic object you would like to extract and distinguish it from the other structures beside and/or around it.

17. To start extracting the geospatial data, click the *Add marker* button  .

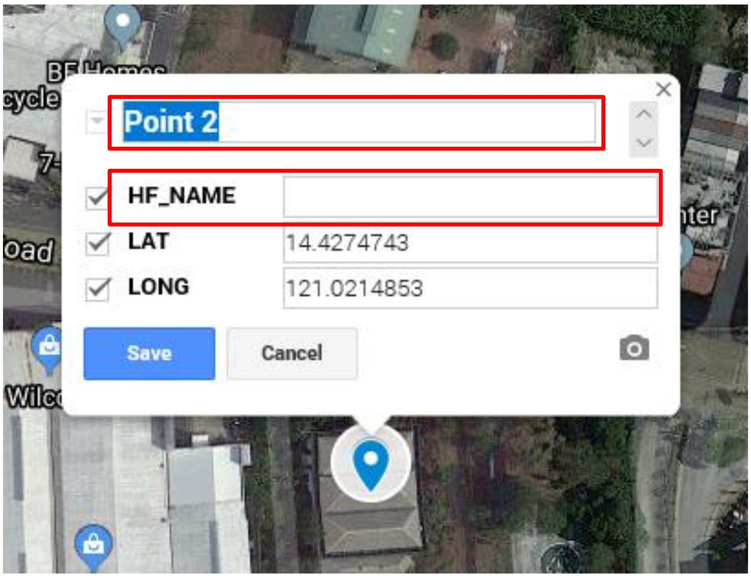


Note: When extracting line- or polygon-type object, click the *Draw a line* button  .

18. Trace from the basemap the object you want to extract. As you are trying to capture the location of health facilities in this example, a point-type object, just click once on top of the object in the basemap that corresponds to the structure of the health facility.

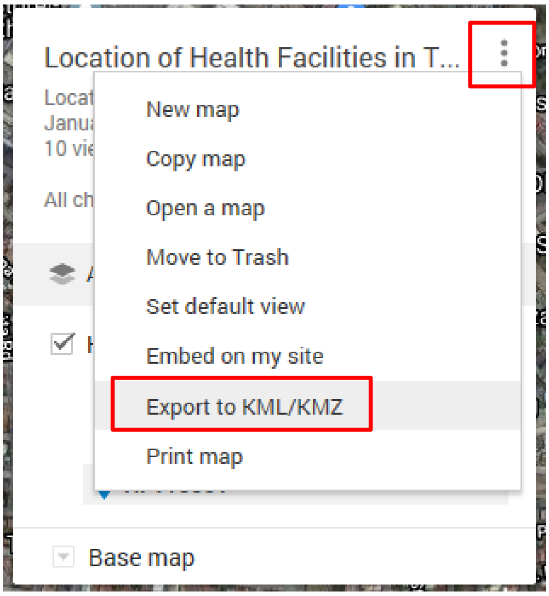
Note: For line-type objects (such as roads, rivers, boat route, etc.), you will have to trace along the length of the object you are extracting, usually along the center line. For polygon-type objects (such as administrative boundaries, operational division boundaries, building footprints, etc.), you will have to trace the boundary of the object you are extracting. You can learn more about creating line-and polygon-type objects from this link: https://support.google.com/mymaps/answer/3433053?co=GENIE.Platform%3DDesktop&hl=en

19. Once you've clicked on top of a geographic object, a pop-up window will appear. This corresponds to the fields in the attribute/data table. Enter the unique ID in the top field (if you chose the HF_ID to serve as the title for each individual point) and the official health facility name in the HF_NAME field as they appear in the health facility master list.



20. Note that the LAT and LONG fields are automatically populated. Do not change these values. Click *Save*.
21. Navigate to the next geographic object then repeat steps 17-20.
22. Once all geographic objects are extracted, you may acquire the data for use: either download it as KML/KMZ (Steps 23-25) or copy the content of the data table (Steps 26-32).

23. To download as KML/KMZ, click the 3 dots at the end of the map title then choose *Export to KML/KMZ*.

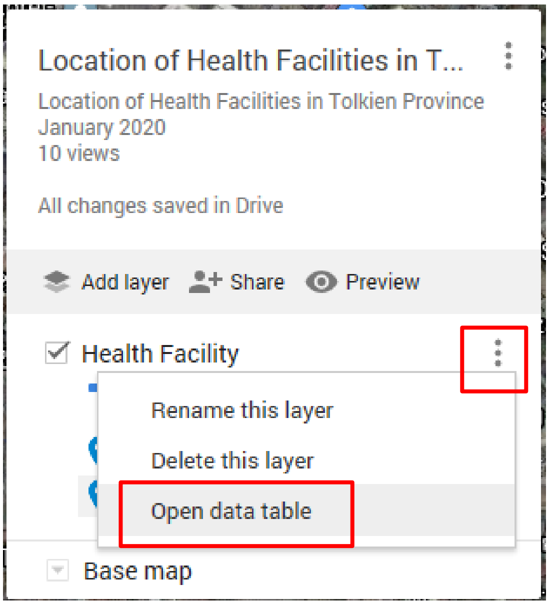


24. In the window that opens, choose the layer containing the extracted objects from the dropdown list (instead of “Entire map”). Click *Download*.

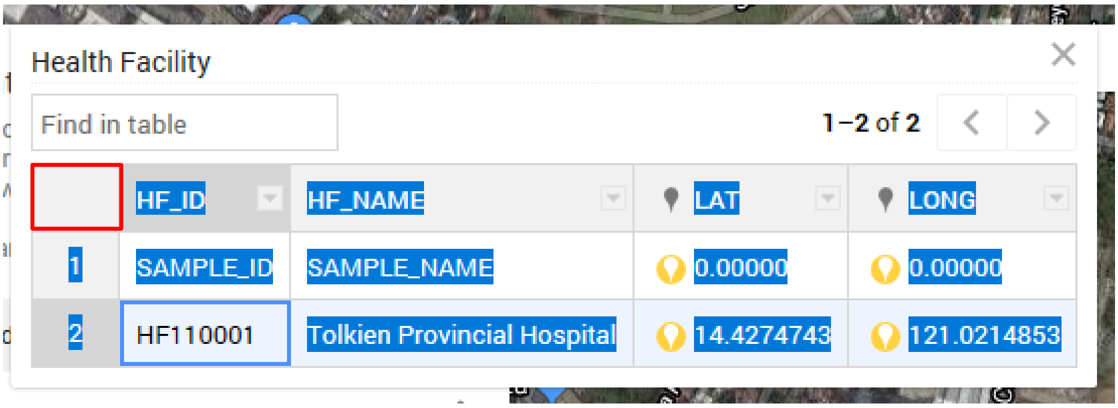
25. You may now use the downloaded KML/KMZ file in GIS software to convert it to the required file format as defined in the data specification.

26. To copy the contents of the data table, first open a new Microsoft Excel sheet (or your preferred spreadsheet program) on your computer.

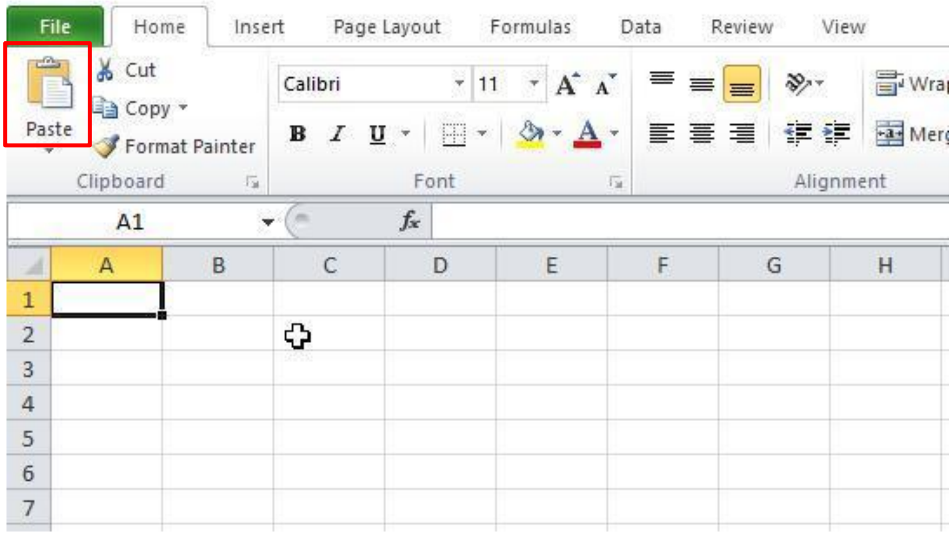
27. In Google My Maps, click the 3 dots at the end of the layer name then choose *Open data table*.



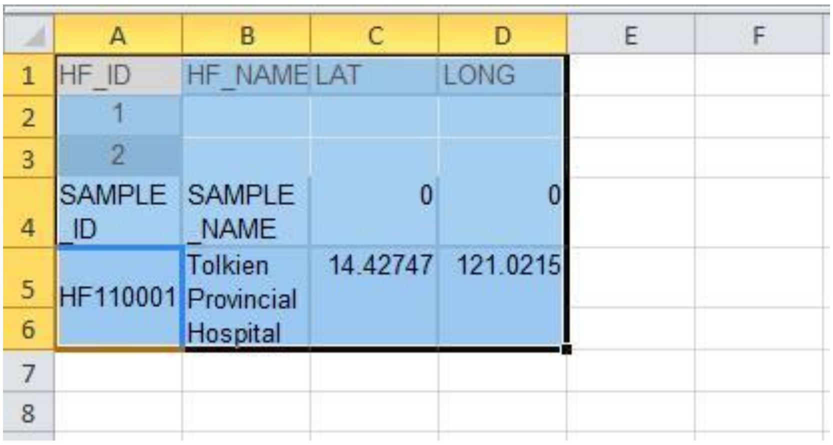
28. When the data table opens, starting from the empty cell beside HF_ID, click, hold, and drag to the last entry in the LONG field. Make sure that all entries are highlighted. Press *CTRL+C* on your keyboard.



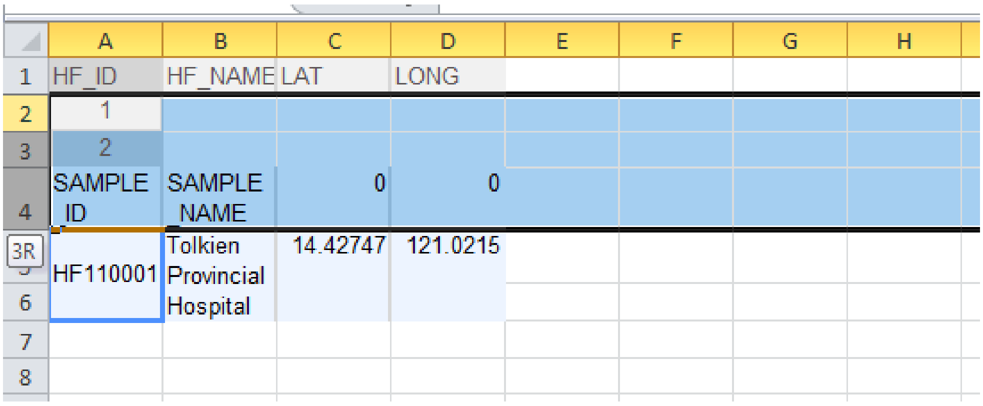
29. On your Excel sheet, click on Cell A1 then click the *Paste* button  . (Or press *CTRL+V* on your keyboard.)



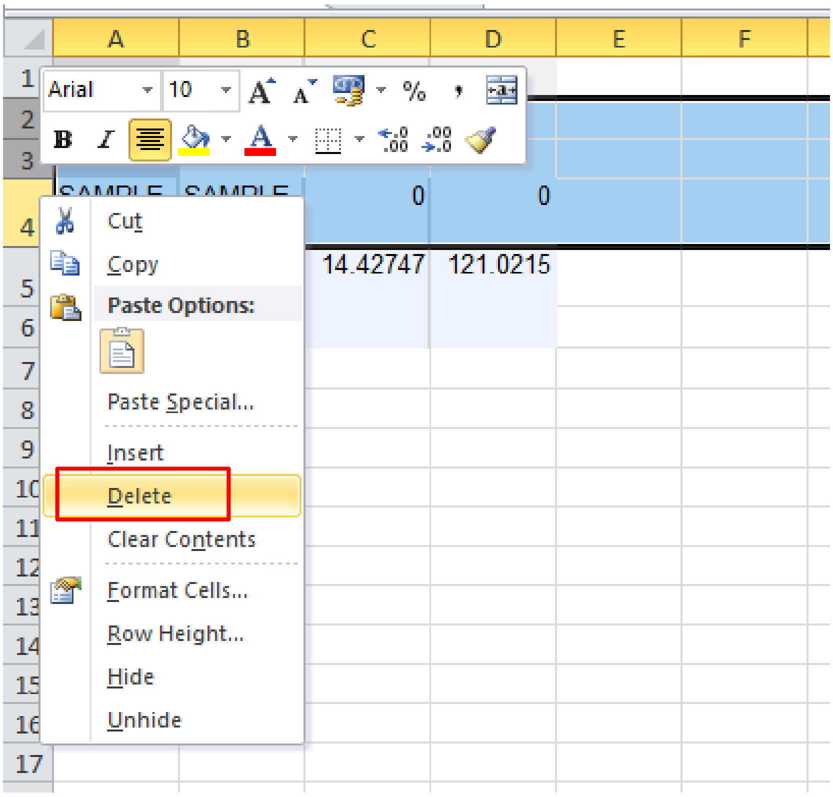
The pasted table appears on the Excel sheet. The field names and corresponding content do not line up in this pasted table and it will have to be cleaned before use.



30. Delete the entries “1” and “2” under HF_ID and the sample entry you created in Part 1. To do so, click and drag from Row 2-4 to select them.



31. Right click on the selected row number and choose *Delete*.



32. Save your Excel file.

33. The extracted data is now ready for final validation.