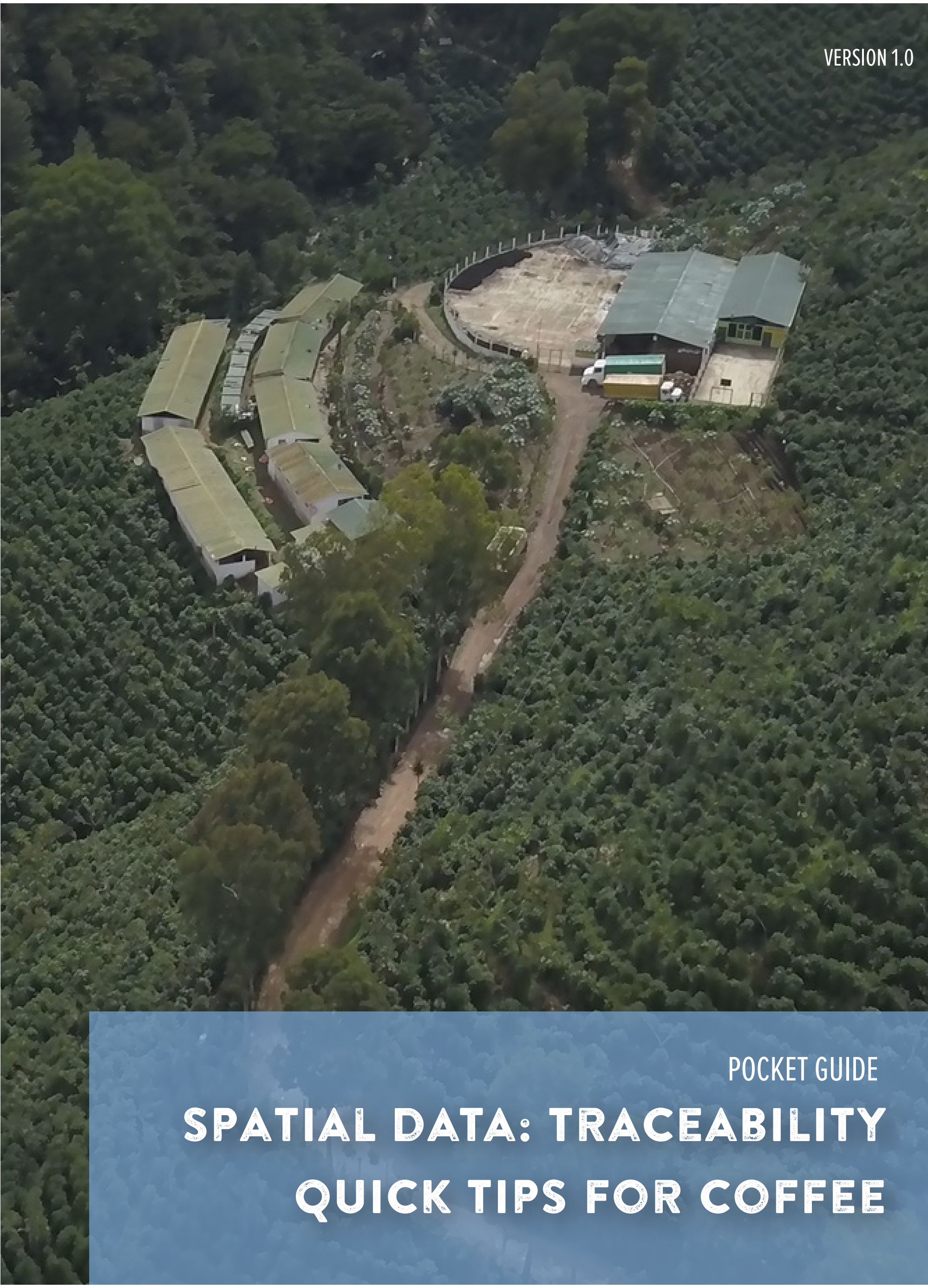
VERSION 1.0



POCKET GUIDE

SPATIAL DATA: TRACEABILITY QUICK TIPS FOR COFFEE

SUSTAINABLE COFFEE CHALLENGE

SPATIAL DATA QUICK TIPS FOR COFFEE

Context

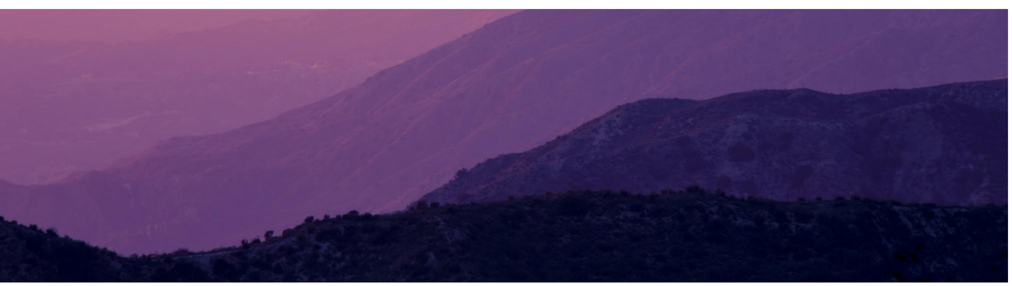
Coffee value chain actors are increasingly exploring innovative technological tools and solutions to enhance farm-level traceability. Having accurate location data is fundamental to understanding and managing a range of sustainability risks and opportunities within a company's supply chain. Especially with increased regulatory pressure – particularly the EU's Regulation on Deforestation-Free Products (EUDR) and Corporate Sustainability Due Diligence Directive (CSDDD) – there is an increasing need for accurate data collection.

To better ensure that farm data collection is conducted in a uniform and systematic manner, based on a culmination of over 20 years of experience in spatial data collection, Conservation International has developed the following guidance. While this guide is not comprehensive, it aims to prevent some of the most common problems that would otherwise require significant editing and cleaning to make spatial data useable, if at all.

The Pocket Guide is broken into three sections that provide guidance on 1) planning data collection, 2) collecting data in the field, and 3) how to manage the data once it's collected and troubleshoot common errors.

EUDR REQUIREMENTS RELATED TO SPATIAL DATA

- For each shipment, companies will need to provide "conclusive and verifiable information" that the product is deforestation free. One aspect of this is geolocational data, which is expected to be shared with the EU competent authorities and maintained by the company for 5 years.
- In addition to collecting the data for the purpose of reporting to the EU, it serves as a basis to monitor forest change as part of a company's broader due diligence approach.
- Point data will be required for plots <4 hectares. Polygons, which provide greater accuracy of an area's extent, are required per EUDR for any plot of land >=4 hectares.
- The required locational units for the EUDR's system are Decimal Degrees to six decimal places (i.e. dd.dddddd) using the WGS84 datum (EPSG4326).



1 The "plot of land" – the subject of geolocation under the Regulation – is defined in Article 2 as "land within a single real estate property, as recognized by the law of the country of production, which possesses sufficiently homogeneous conditions to allow an evaluation of the aggregate level of risk of deforestation and forest degradation associated with relevant commodities produced on that land."...Polygons are to be used to describe the perimeter of the plots of land where the commodity has been produced. Each polygon should indicate a single plot of land, whether contiguous or not. A polygon cannot be used to trace the perimeter of a random land area that might include plots of land only in some of its parts. FAQ - Deforestation Regulation_1.pdf (europa.eu)

STEP 1: PLANNING DATA COLLECTION

Location data can be collected in several ways: in the field using GPS-enabled devices, created in geospatial software on a device (e.g. tracing boundaries with satellite imagery as a reference), extracted from existing spatial data (e.g. extracting a jurisdiction from a national dataset), or using Artificial intelligence (AI). Some of the factors that affect choosing an approach are time, cost, safety, scale, and accuracy/precision. Field measurements are best when the site is relatively small, and measurements with great accuracy are required. Larger geographic areas like an entire jurisdiction or watershed, or areas that can be easily distinguished using available satellite imagery, are better candidates for using GIS and satellite imagery. Efforts to map coffee using AI is an emerging field, especially for agroforestry and shade-grown systems.

1.1 Determine who will collect the data

When determining the approach to data collection, and whether field measurement is required, you should also consider who will be responsible for collecting location data. This could be someone from your own organization or a third party, such as a supply chain partner, not-for-profit or a reputable service provider. In any of these cases, it is best to develop standard operating procedures for data collection so that surveyors are collecting data across places using a standardized approach and format.

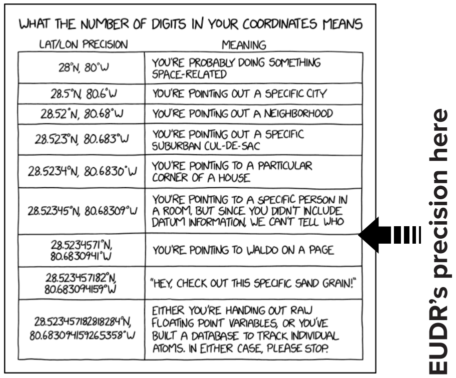
Once you've established who is collecting spatial data in the field, you can turn to how it's collected. The following section walks you through considerations for equipment and apps. Before determining a final approach, it is recommended that you consult with the colleagues or the organization who will collect the data to understand their current systems, equipment, and preferences.

ACCURACY VS PRECISION WITH GPS MEASUREMENTS IN THE FIELD

While standard GPS units and mobile devices will display coordinates with a precision of many decimal places, the accuracy of this number may vary widely due to several factors:

- Terrain: just as being in a valley reduces your ability to see in the distance, steep terrain can also limit the ability to receive GPS satellite signals.
- Trees and other canopy cover will interfere with GPS satellite reception.
- Number and angle of GPS satellites overhead at time of collection: Signals from a minimum of 4 satellites, ideally spread across the sky, are required for an accurate location measurement. Websites like GNSS Mission Planning can be used before going to the field to identify the best times of day to collect GPS data for a given date and location.

EUDR's required 6 decimal places for coordinates is a precision of ~11 cm. The chart below helps to illustrate what this precision means in real life.


WHAT THE NUMBER OF DIGITS IN YOUR COORDINATES MEANS

LAT/LON PRECISION	MEANING
28°N, 80°W	YOU'RE PROBABLY DOING SOMETHING SPACE-RELATED
28.5°N, 80.6°W	YOU'RE POINTING OUT A SPECIFIC CITY
28.52°N, 80.68°W	YOU'RE POINTING OUT A NEIGHBORHOOD
28.523°N, 80.683°W	YOU'RE POINTING OUT A SPECIFIC SUBURBAN CUL-DE-SAC
28.5234°N, 80.6830°W	YOU'RE POINTING TO A PARTICULAR CORNER OF A HOUSE
28.52345°N, 80.68309°W	YOU'RE POINTING TO A SPECIFIC PERSON IN A ROOM, BUT SINCE YOU DIDN'T INCLUDE DATUM INFORMATION, WE CAN'T TELL WHO
28.5234571°N, 80.6830941°W	YOU'RE POINTING TO WALDO ON A PAGE
28.523457182°N, 80.683094159°W	"HEY, CHECK OUT THIS SPECIFIC SAND GRAIN!"
28.523457182818284°N, 80.683094159265358°W	EITHER YOU'RE HANDING OUT RAW FLOATING POINT VARIABLES, OR YOU'VE BUILT A DATABASE TO TRACK INDIVIDUAL ATOMS. IN EITHER CASE, PLEASE STOP.

Source: https://xkcd.com/2170/

1.2 Establish what equipment is needed

Most smartphones and tablets are GPS-enabled and can collect data in the field using a variety of mobile apps, but their native GPS receivers will have limited locational accuracy—an important consideration as explained in the "Accuracy vs. Precision" callout box. In addition to a smartphone or tablet, a separate Bluetooth-paired GPS receiver is recommended for each mobile device (smartphone or tablet) to ensure faster and more accurate location data. Power banks will also be needed to charge equipment in the field. If the data collectors already own and prefer to use their handheld GPS units that can collect locational data (e.g. Garmin), it will need to be complemented by other efforts to collect information, like volumes. The GPX file the devices create may also need to be converted to another spatial data format (e.g. shapefile).

1.3 Define the desired data and how to structure it

Beyond defining what data needs to be collected, giving forethought to how it should be structured helps ensure its usability, reduces subsequent data cleaning, and may inform which data collection apps to use (step 1.4).

Spatial Data

Importing coordinates from a table into a GIS is easier when collected as follows:

- Collect locations in decimal degrees (DD) to six decimal places using WGS84 as the coordinate system.
- If collecting a boundary, collect any additional features (e.g. mill) separately and distinctly to avoid it accidentally being included in the boundary polygon. If possible, make these collections separate files. Per EUDR, this is only for plots >4 ha.
- If a table of coordinates is sent for a polygon instead of GIS data file, the longitude (X) and latitude (Y) symbols should be indicated, and points should include N, E ,S ,W with "-" (minus sign) for west and south coordinates. For example:
    - Coordinate in Mexico: 104.341°W, 22.371° N (should be collected: -104.341, 22.371)
    - Coordinate in Madagascar: 50.604° E, 18.663° S (should be collected: 50.604, -18.663)
    - Coordinate in Brazil: 52.424° W, 18.196° S (should be collected: -52.424, -18.196)

Tabular Data

Tabular information, complementary plot aspects that are associated with a boundary, can vary considerably and be

a combination of information collected on-site via survey or observation, or information that can be added independently of the fieldwork (e.g. country, project ID). Suggested columns (typically called "fields" in the geospatial industry) include: unique ID that can be used to connect the plot's location with other relevant non-spatial information (aka "key field"), date visited, unique farm ID, and area classification (i.e. plot boundary). The Linux Foundation's AgStack Project is one effort worth noting for generating potential unique farm/plot IDs.

Further considerations regarding tabular data include:

- If the data will be shared[2], it is important to flag any potentially sensitive or Personally Identifiable Information (PII) such as a farmer's name. It is advised to only share a derivative copy of the data with any non-essential information removed instead of the full, original data.
- When creating the schema for the tabular information, it is advised to follow the rigor required for the shapefile format to ensure interoperability with other geospatial formats and minimize loss or difficulty combining with other data. For example, no text field can be longer than 256 characters. The rules for shapefile column (field) names include:
    - Column names must start with a letter.
    - Column names must contain only letters, numbers, and underscores (i.e. no spaces or special characters).
    - Column names cannot start with the following invalid characters: (`~@#$%^&*()-+=|\\,<>?/{}.!'[]:;_0123456789)
    - Column names must not exceed 10 characters.
    - "Date" is a reserved name and cannot be used in column names
- Data types: software will often try to recognize if the data in a given column is numeric only vs text; that data type is fixed once defined in some platforms. This matters because of potential issues when combining data from different sources that used different data types or moving data from one platform to another (e.g. opening an Excel file in GIS software).

1.4 Select a mobile field data collection App

There are numerous apps for mobile field data collection. For this work, the most important factors are that it is compatible with your existing device, allows for the collected data to be accessed elsewhere (either through synching to the cloud or sharing an export file electronically), and preferably collects polygons in a standard geospatial format rather than requiring the extra steps of conversion from lines or to an accepted data format using additional software.

Regardless of which app you plan to use, it is recommended to have more than one app available on a device, and to log in and test using each of them before heading into the field and going offline to ensure all necessary schema and context files are downloaded, that the login credentials work, and to become familiar with the interface.

Other considerations when determining the best data collection app include:

- ☑ Cost of using the app, particularly if there are multiple users needed
- ☑ User-friendliness and your team's comfort level navigating the app
- ☑ Ability to connect to other software or platforms that are compatible with your existing systems
- ☑ Whether it supports collecting user-identified points or continuous streaming (or both)

2 No personal information is required from the farmers (unless they are direct suppliers of the operators or operators themselves). The geolocation of the land they cultivate is sufficient. FAQ - Deforestation Regulation_1.pdf (europa.eu)

- ☑ Ability to collect tabular attributes, capture pictures or other media
- ☑ Ability to create custom templates and the time and skills needed to create one
- ☑ How quickly the app drains your device's battery

For more information on field apps, a chart comparing several options is available at: [Mobile Data Field Collection Comparison v2023]

1.5 Establish a protocol for file formats

Point and polygon data can be collected in a variety of standard formats, based on the software or tool used and its defaults. Note that while some apps allow you to select the file format upon exporting, others only provide it in a single format. Below is a list of the most common vector spatial data formats. While it is possible to convert between these formats using geospatial software and the EU has not yet specified which data formats their system will accept, Conservation International recommends using either a Shapefile, GeoJSON, or GeoPackage format for greatest interoperability.

Format	Details
Shapefile	This is one of the most common vector data formats. A single shapefile is actually comprised of several file extensions: • Mandatory to be usable: .shp, .shx, and .dbf • Strongly encouraged: .prj, .xml, .sbn, and .sbx This format was developed by Esri, but can be read by other geospatial software. Shapefile attributes cannot store null values, they round up numbers, they do not allow field names longer than 10 characters, and they cannot store both a date and time in a field. Shapefiles do not work well for storing information in languages other than English (i.e. poor support for Unicode characters).
GeoJSON	A subset of JSON, and an open standard format for simple geographic features and their non-spatial attributes; often used in online environments. It only uses the WGS84 coordinate reference system and units of decimal degrees.
GeoPackage	A relatively new open format from the Open Geospatial Consortium. It is compatible with QGIS and other open data software, as well as ArcGIS Pro. It supports a variety of coordinate systems, as well as vector, raster, tables, metadata formats. Like Esri's File Geodatabase format, a single GeoPackage can contain multiple spatial data files, or a single one. Geopackages also supports a higher volume of data, as well as longer field names and text lengths compared to Shapefiles.
Feature class in a file geodatabase (*.gdb)	A file geodatabase will appear as a folder outside geospatial software and may contain multiple feature classes; these can only be filtered using geospatial software. This format was developed by Esri but can be read by other geospatial software. It was developed to address the shapefile limitations listed above.
GPX	This is a common format for GPS units and collects waypoints, tracks, and routes (i.e. additional processing would be needed to convert collected data to polygons). Non-spatial attributes would need to be collected separately.
KML/KMZ	This format was developed for Google Earth, and only uses the WGS84 coordinate reference system and units of decimal degrees. The structure for its tabular attributes is less rigid than Shapefiles (e.g. higher limit for field length and column names), which risk losing information in format conversion.

STEP 2: COLLECTING DATA IN THE FIELD

The most critical part of mapping in the field happens before it begins through thorough preparation. This will include:

- ☑ Installing and testing the apps on the mobile devices
- ☑ Charging all equipment, including power banks as both devices and GPS receivers will need to be charged in the field
- ☑ Wearing appropriate clothing for the conditions (social and environmental)
- ☑ Checking weather conditions
- ☑ Planning a potential route and backup options using input from local guides/ experts, etc.
- ☑ Using mission planning software to understand the ideal time for GPS signal accuracy
- ☑ Ensuring the landowner has granted permission prior to any mapping efforts and remembering to leave no trace (i.e. no littering).

Once in the field, it may be helpful to note that GPS receivers can be held up higher than mobile devices when collecting data for better GPS reception. Additionally, unless a polygon boundary is very complex (e.g. along a winding riverbank), opt to collect distinct boundary points at all significant corners of the boundary rather than collecting the locational data as a continuous stream or trail of points. This not only saves battery power in the field, but also has the benefit of smaller file sizes. Lastly, ensure the GPS is set to latitude/longitude in decimal degrees and datum WGS84 or, if collecting in another projection, e.g. UTM, make sure to include this in the metadata.

STEP 3: MANAGING THE COLLECTED DATA

• • • • • • • • • • •

"PROVIDING INCORRECT GEOLOCATION DETAILS WOULD CONSTITUTE A BREACH OF THE OBLIGATIONS UNDER THE REGULATION.[3]"

• • • • • • • • • • •

Plot data needs to be reviewed and cleaned, if necessary, using a desktop or online geospatial resources before it is ready to be shared with any third parties or checked against forest cover. Performing this step is recommended before compiling the data with any existing plot data. Basic things that would be reviewed at this step are that the plot is in the expected location, and that a polygon's boundary appears to be correct. This is also the phase where the data could be converted to a different data format.

It is possible to convert measurements using Excel or geospatial software to other systems, but it is essential to know the coordinate system used either when collecting coordinate points with a GPS unit on the ground, or using a desktop or web-based GIS/geospatial software (e.g. UTM zone and associated datum). The following sections explain how to convert coordinate systems and units.

3 FAQ - Deforestation Regulation_1.pdf (europa.eu)

3.1 Ensure Consistent Coordinate Systems Across Collected Data

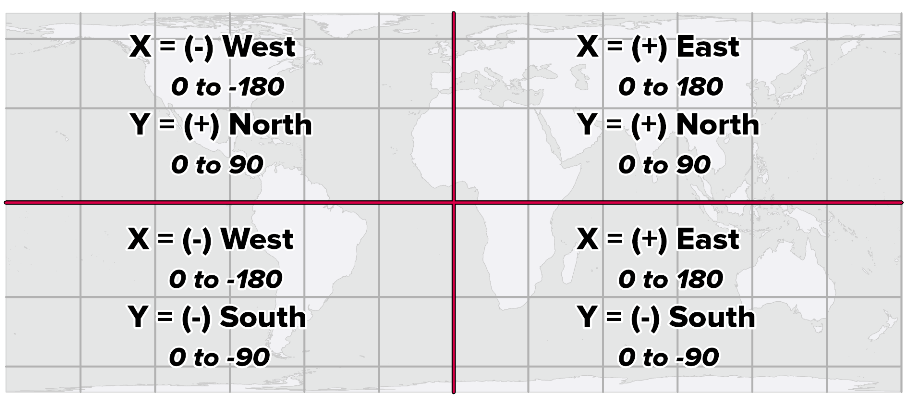
A **coordinate system** is a framework to define a location on the earth. The GPS satellite constellation uses a mathematical model of the Earth's shape called WGS84, and data collected by handheld GPS units are typically x and y in decimal degrees using WGS84 by default. However, many regions prefer to collect data using their local UTM (Universal Transverse Mercator) projection, and potentially using a national or regional datum (e.g. Clarke 1880, India 1960, NAD83, SAD69). The units of UTM projections are meters, instead of decimal degrees. **All spatial data for EUDR will need to be in WGS84, with units in decimal degrees.** If geospatial software such as Esri's ArcGIS or QGIS is unavailable, an online coordinate system converter like [Lat/Lon and UTM Conversion] or [epsg.io] can be used to convert coordinates one at a time to WGS84 in decimal degrees.

3.2 Ensure Consistent Units Across Collected Data

GIS data formats can use location units of degrees or meters, and geospatial software can convert between them. Coordinates using meters as the unit can be easily identified by the numbers involved—instead of degrees as the unit with a maximum value of 180, meters are usually in the hundreds of thousands or millions. Degrees can be expressed in several formats, such as Degrees Minutes Seconds (i.e. DD° MM' SS") which uses a base-60 grid. To manually convert this format to the required Decimal Degrees (base-10) format, use this formula for both latitude and longitude:

Decimal Degrees = Degrees + (Minutes / 60) + (Seconds / 3600).

Instead of cardinal direction, Decimal Degrees uses a negative value for south and west. It is important to note if these coordinates are north or south of the equator, and east or west of the Prime Meridian (Greenwich). If it is a south or west coordinate, multiply the result by -1.



Cardinal directions and associated range of values for decimal degrees. Note that X is Longitude; Y is Latitude.

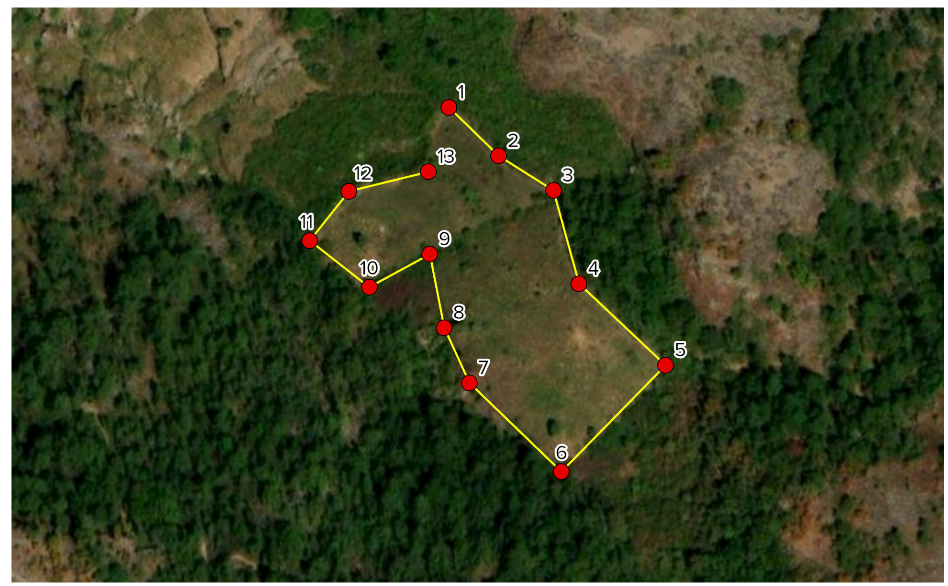
3.3 Troubleshooting Common Errors

Over the years, we've run into many data quality issues that could have been easily avoided in the field. While some can be corrected using geospatial software and a familiarity with the mapped area, it can add significant (and unnecessary) costs and time. If not corrected, these errors would require polygons or points to be removed and lost, or even worse, threaten the integrity of the data set.

Here are a few of the most common errors we've found and some tips on how to avoid them.

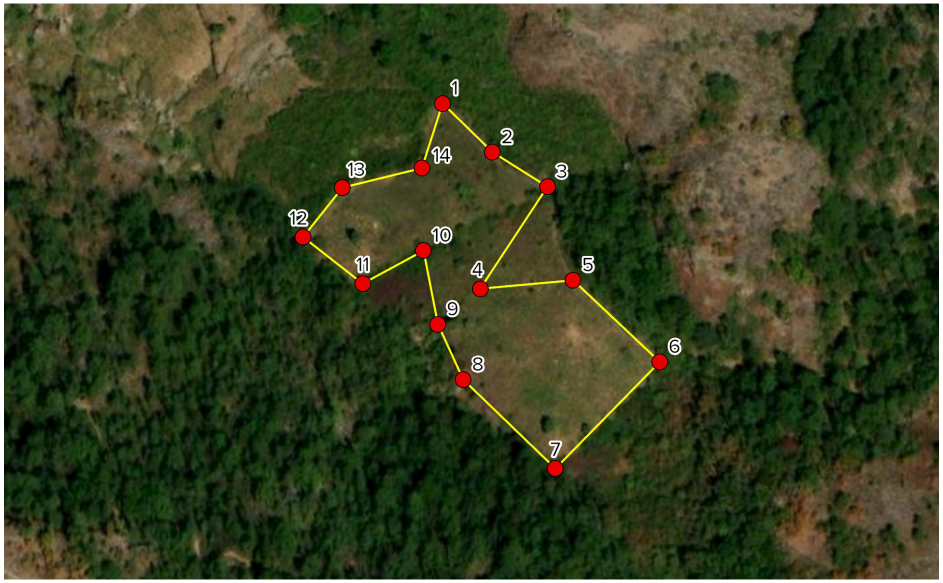
Common errors with field-collected polygons:

1. Collecting a line that doesn't close, causing a failure to convert it to a polygon boundary. This can happen when the technician doesn't return to the starting point, or access is physically limited by ponds, creeks, or other physical features.
   - *Solution: Review the boundary using satellite imagery and correct using geospatial software.*



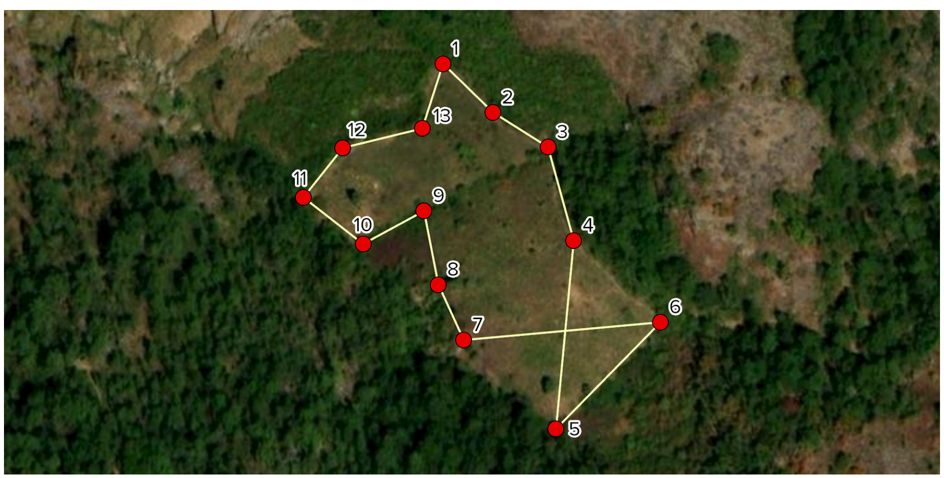
In this example, a line created by a series of points was collected as part of the boundary data collection, but was not completed.

2. Collecting additional points (e.g. farm mill) as part of the boundary set, causing the boundary's outline to plunge into the intended area. It is important the boundary list only contains those outermost points.
   - *Solution: Use geospatial software to convert the polygon to its vertices, delete the excess/incorrect points, ensure the remaining points have a correct order in their attribute table, and use the software to convert the vertices into a polygon again. Going forward, be sure to collect boundary points as separate files from other farm features.*



In this example, a point in the middle of the farm was collected as part of the boundary point data collection.

3. Not having the boundary points in order, causing shapes resembling a bow tie.
   - *Solution: If the boundary coordinates are from a table, correct the order there and ensure it has a column indicating the numeric order before converting them to a polygon. Or, use geospatial software to convert the polygon to its vertices, delete any incorrect points and edit their attributes to ensure the correct order before using the software to convert the vertices into a polygon again. Tools such as "Repair Geometry" in Esri's ArcGIS software can also often resolve this. Going forward, be sure to organize the boundary coordinates in order, and collect a boundary in a single session, if possible.*



4. A projection or coordinate system is undefined.
   - *Solution: The most accurate way to resolve this is to ask the data provider to clarify the projection or coordinate system. If this isn't possible, loading it into geospatial software and observing the associated numbers can help to identify whether the associated coordinate system is using decimal degrees (0-180) or meters (millions). Based on this and a knowledge of what country the polygon should be in would then allow you to look up any official national coordinate systems using those units.*

Common errors with field-collected point data:

1. Point data is often provided in tables. Review the collected points to check for any clearly incorrect locations. For example, the following set of points for Mexico are all between 15-16 degrees North latitude and 92-93 degrees West longitude, however the first point has a -1 degree longitude and latitude. This point is clearly incorrect:

Latitude	Longitude
1 1 1.0	**-1 1 1.0**
16 31 43.0	-93 25 38.0
15 59 0.0	-93 20 9.0
15 59 0.0	-93 18 17.0
15 58 15.0	-93 28 30.0
15 55 47.0	-93 22 38.0
15 32 27.0	-92 3 12.0

2. Points collected with the ° (degree) ' (minute) " (second) symbol must be edited before they can be imported into a GIS. For example, the list of points from Mexico on the left needs editing before importing. The table on the right shows the same set of points, correctly edited to be imported into GIS (° ' " symbols removed, W symbol replaced with "-" (minus sign)). *Note that these are still in DD MM SS format, and would need to be converted to the dd.dddddd format before submitting to a system.*

Latitude	Longitude
17°2'6.0"N	97°49'2.0"W
17°2'13.0"N	97°48'45.0"W
17°2'0.0"N	97°49'5.0"W
17°2'14.0"N	97°48'41.0"W
17°2'4.0"N	97°47'7.0"W
17°1'37.0"N	97°98'16.0"W
17°4'52.0"N	97°50'14.0"W

Latitude	Longitude
17 2 6.0	-97 49 2.0
17 2 13.0	-97 48 45.0
17 2 0.0	-97 49 5.0
17 2 14.0	-97 48 41.0
17 2 4.0	-97 47 7.0
17 1 39.0	-97 48 16.0
17 1 21.0	-97 48 20.0

Additional errors commonly found with point data include:

- Reversing X and Y, resulting in farms in polar regions
    - X = Longitude, Y = Latitude
- Omitting a negative sign for southern and western locations

3.4 Consider Open Data

Making project location information publicly available is increasingly required by donors, and can help to identify potential collaborators, avoid potential duplication of effort, and coordinate to maximize the impact of investment. Sharing data with other platforms can also help to ensure the collected data can be found and accessed after the project's investment has ended. It is advised to use standard geospatial formats as described above, removing PII or any sensitive attributes, and to use a standard metadata format (e.g. ISO 19115) to ensure usability and attribution.

• • • • • •

In sum, the steps above offer a starting framework for effectively collecting and managing spatial data and avoiding common pitfalls as coffee sector stakeholders work to enhance farm-level traceability and comply with evolving regulations.

SUSTAINABLE COFFEE CHALLENGE

www.sustaincoffee.org

CONSERVATION INTERNATIONAL