***NOTE:*** *This document is in early draft form as of July 2025 and will be edited significantly over the coming months. It is provided here as an example and starting point. Please reach out to* *dataforimpact@mercycorps.org* *to request the latest copy of this guidance or to provide any comments or feedback, which are very welcome! When finalized, we will attempt to provide the full document link in another version of this report.*

**Draft** Spatial Data Collection Processes and Standards Guidance

* *When collecting data, it is recommended that a spatial/geographic component be included in all cases where data is collected unless there is some extenuating factor (like security, sensitivity, etc.). In contexts where collecting or storing geographic data may pose risks to participant safety, compromise sensitive program operations, or violate local data protection laws, spatial data should be omitted or anonymized. Exceptions should be clearly documented and justified ensuring that participant privacy and data security remain a top priority.*
* *What geographic component is captured will depend heavily on the use case and the level of geographic detail needed or desired:*
	+ *Level of geographic detail – administrative unit*
		- *Use a specific administrative boundary level (i.e. admin 1, admin 2). Wherever possible country team should harmonize geographic variables as admin 1, admin 2 to ensure consistency across datasets and programs.*
		- *See table below on choosing a dataset for administrative boundaries*
		- *Administrative unit names collected in data collection should exactly conform with (at least) one administrative boundary dataset chosen so that the data can be joined with the administrative boundary dataset. Failure to do so will result in significant additional data cleaning efforts or even inability to accurately map the collected data.*
	+ *Level of geographic detail – populated place*
		- *Use the coordinate of a specific populated place or collect the location with a GPS coordinate if it does not exist in an authoritative dataset (see row 1 in Table 2 below).*
		- *Populated names collected in data collection should exactly conform with (at least) one name in the administrative boundary dataset chosen so that the data can be joined with the administrative boundary dataset. Failure to do so will result in significant additional data cleaning efforts or even inability to accurately map the collected data.*
	+ *Level of geographic detail – exact location*
		- *For a general participant survey or activity location survey, this should be a GPS coordinate of the activity location at the time of intervention or a coordinate of a participant location.*
		- *For data collection on assets or structures (e.g. on engineering projects, WASH improvements) this would generally be a GPS coordinate (point), a line (i.e. water pipeline, road), or a polygon (location of a structure)*
		- *For data on market activity or information, this will typically be a GPS coordinate of the market site (or association with an existing market locations dataset)*
		- *For data collection on specific ecological areas (i.e. agricultural plots, restoration sites) this will typically be in the form of a polygon either collected on site (see methods in table below) or a point at the center of the area and later digitized with high-resolution imagery*

***Table 1: Standards for Collection of Spatial Data***

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Parameter*** | ***Standard*** | ***Description*** |
| *1* | *Accuracy Level* | *Less than 15 meters* | *This parameter defines the acceptable deviation in the measured location, ensuring precision in geospatial data. Accurate location data is crucial for mapping and analysis, as it affects the reliability of the results. For example, in disaster response, precise location data can help identify affected areas and allocate resources effectively.* |
| *2* | *Time Interval (****lines and polygons only****)* | *4 seconds per point walking, 2 seconds per point driving* | *This parameter specifies the frequency of data points that should be collected when measuring lines which is important for capturing adequately detailed paths and boundaries. Frequent data points ensure that the shape and course of features like roads, pipelines, and boundaries are accurately represented. This is essential for infrastructure planning and monitoring.* |
| *3* | *Number of Satellites Connected* | *Four or more*  | *Connecting to four or more satellites ensures reliable GPS data by providing multiple points of reference. This improves the accuracy and stability of location measurements, which is important for any geospatial analysis. Reliable GPS data is crucial for applications like navigation, asset tracking, and field data collection.* |
| *4* | *Navigation Units* | *Metric* | *Using metric units ensures consistency in data collection and analysis. Standard units of measurement are important for comparing and integrating data from different sources. This is essential for projects that involve international collaboration or require standardized reporting.* |
| *5* | *Coordinate System* | *World Geodetic System 1984 (WGS 84)* | *The World Geodetic System 1984 (WGS 84) is the reference framework for geographic data, ensuring compatibility and accuracy. Using a standard coordinate system allows for seamless integration of data from different sources and supports global analysis. This is important for applications like environmental monitoring and global positioning.* |
| *6* | *Coordinate Format* | *Decimal Degrees with at least 5 decimal places (eg. 42.04837,*  | *Decimal Degrees with at least 5 decimal places specifies the precision of location data, which is critical for detailed mapping. High precision is important for applications that require accurate positioning, such as land surveying, asset management, and navigation.* |
| *7* | *North Reference (if Applicable)* | *True North* | *True North defines the directional reference for orientation in mapping. Using a consistent north reference ensures that maps and spatial data are aligned correctly, which is important for navigation, planning, and analysis.* |
| *8* | *Metadata* | * *Date collected*
* *Time collected*
* *Unique ID of device or enumerator*
* *GPS precision at time of collection*
 | *Metadata includes additional information about the data collection, such as date, time, unique ID, GPS precision, and location. This information is crucial for context and validation, as it helps verify the accuracy and reliability of the data. Metadata is important for tracking data provenance and supporting reproducibility in analysis.* |

*Table adapted from USAID Geographic Data Collection and Submission Standards (579saa)*

***Table 2: Process Guidance for Collection of Spatial Data***

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Parameter*** | ***Protocol*** | ***Description*** |
| *1* | *Location (coordinate)* | *This will vary by application.** *For buildings, homes, and structures (i.e. markets, stores): In front of the entrance (taking a point at the center of a building inside or covered by a roof will likely compromise the connection with satellites and thus the accuracy of the measurement)*
* *For implementation areas (i.e. a field or a camp): as close to the center point of the location as possible*
* *For an asset, natural feature or non-building structure (e.g. water pump, cistern, tree, cellular tower, river access point, reservoir): over the top if possible or nearest the location without interfering with the satellite connection to the device (so not inside a structure or under an awning)*
 | *This parameter specifies point data for various applications, ensuring accurate placement. For example, taking a point at the entrance of a building ensures that the location is accessible and visible. Accurate point data is important for applications like asset management, infrastructure planning, and field data collection.* |
| *2* | *Location (line)* | *This will vary by application but generally:** *Follow the course of the thing being measured as closely and accurately as possible.*
* *Example: for a water pipeline, walk next to the pipeline or safely under if possible. If an impediment (like a bush, tree, boulder, or building necessitates deviating from the path, note the location and fix during data cleaning.*
* *Example: for a road, try to collect in a dedicated vehicle if possible (or consider digitizing with recent high-resolution imagery)*
* *Example: river – collect on a boat where possible, digitization of high-resolution imagery highly preferable*
 | *Path data for linear features, such as pipelines, roads, and rivers, is important for infrastructure mapping. Following the course of the feature ensures that the data accurately represents its shape and location. This is essential for planning, monitoring, and maintenance of linear infrastructure.* |
| *3* | *Location (polygon)* | *This will vary by application:** *For agricultural plot boundaries generally less than 10 hectares:*
	+ *walk the precise boundaries of the defined plot* ***OR***
	+ *Take coordinates of the corners of the plot (ensuring ALL corners are collected so that a boundary can be created)* ***OR***
	+ *(if access to recent high-resolution imagery for the area and not in a highly tropical place with trees and cover crops interspersed) take a GPS coordinate close to the middle of the plot and digitize.*
		- *NOTE: this method is much less certain – issues with imagery including cloud cover, presence of other vegetation and unclear boundaries, and lack of available imagery at the right time in the season can make this method very challenging and error prone, though it can be more cost effective than the field data collection methods above. However, one should check the imagery and test the digitization process before fully committing to this method*
* *For agricultural plot boundaries larger than 10 hectares:*
	+ *Manual collection of boundaries will likely not be practical*
	+ *Collect point location on/near the farm/plot to digitize*
* *For buildings and other structures:*
	+ *Many AI-generated building footprint datasets exist and are accessible for free. This would be the preference to field data collection for buildings and structures.*
	+ *For ephemeral buildings (like tents) high resolution imagery + digitization would still be preferable to field collection of boundaries*
 | *Boundary data for areas, such as agricultural plots and property boundaries, is essential for land use and property mapping. Walking the boundaries or taking corner points ensures that the data accurately represents the area. Accurate polygon data is important for applications like land management, zoning, and environmental monitoring.* |
| *4* | *Location (Populated place [i.e. town, village, settlement])*  | * *Mercy Corps has curated populated place datasets for each country that we work in. These can be used as a starting point.*
 |  |
| *5* | *Location (Administrative Area Unit)* | * *Mercy Corps recommends the use of the UN OCHA dataset for mapping administrative boundaries due to its spatial contiguity across countries and its ability to layer administrative levels from admin0 (country) to admin2 (county/district)*
* *Names in this dataset may not exactly match what is officially designated by the government or what is commonly used. Names are in Latin script and may deviate especially for countries using an Arabic script (Arabic, Farsi, etc.)*
 | *Data for political or administrative boundaries, such as counties and districts, is important for governance and resource allocation. Using standardized datasets ensures that the boundaries are consistent and accurate. This is essential for applications like policy making, public administration, and regional planning.* |

***What Constitutes Personally Identifiable Information (PII) with Spatial Data?***

Generally, spatial data is personally identifiable when it can be used to identify an individual directly or indirectly with another dataset. With the exception of household locations, most anonymized spatial data (i.e. no ownership information attached) is not PII, but becomes so when it is owned by an individual and includes that ownership information.

There are a lot of nuances to PII, but generally, here are some key and relevant examples of what would and would not be considered PII in program data collection.

|  |  |  |
| --- | --- | --- |
| Data item | PII? | Explanation |
| GPS coordinate of a household location (regardless of information on who is living there) | Yes | Can be used to directly identify people participating in a program or activity |
| GPS coordinate or polygon boundary of a farm plot (with ownership information) | Yes | Can identify an individual who owns or works on the farm |
| GPS coordinate or polygon boundary of a farm plot (without ownership information or information coded in a protected lookup table) | No | The location is anonymized (fully, or encoded with information that cannot be reasonably traced to an individual) so it is not PII. The lookup table would be PII, however |
| GPS coordinate of a water tap, public park, or public utility | No | These are public assets that are not identifiable to individuals even when combined with other datasets. |
| GPS coordinate of a business location (with or without ownership information, whether or not a sole proprietorship) | No | As long as a business is not at a place of residence, this is not generally considered PII because it is identifying an entity (business) rather than an individual.  |
| Land ownership boundaries or coordinates with ownership information | Yes | Unless fully anonymized, this information is PII (even if it is a business) because it can directly identify a person through the ownership information |
| Coordinates of project sites, offices, IDP camps, health centers, etc.  | No | Project information is not PII unless it is at an individual’s residence or is tracking their movements |

