Geographic Data Collection and Submission Standards
An Additional Help for ADS 579

Introduction

ADS 579 USAID Development Data establishes the requirements governing USAID's development data lifecycle from collecting data to making it accessible. To fulfill the requirements of **ADS 579** and promote best practice geographic data collection and management, this document defines standards and procedures that apply to geographic data commonly associated with planning, managing, and implementing USAID development programming.

Activity Location Data

Data that indicate the location where an activity is implemented and the location of the intended beneficiaries (see **ADS 201.6**). Activity Location Data supports the Agency's objectives to improve the effectiveness and efficiency of development programming and fulfill accountability and transparency objectives.

Monitoring Data

Geographically disaggregated indicator data that are used to investigate the geographic variation in performance for improved monitoring, learning, and adapting (see **Monitoring Data Disaggregation by Geographic Location** for related resources).

Thematic Data

Geographic data related to thematic topics, such as demographic characteristics, environment, and infrastructure. Thematic data is commonly used by USAID Operating Units (OUs) and implementing partners to improve development program planning and management.

Overview

This document is organized in three sections:

I. Managing Data Risk

This section provides an overview of evaluating and mitigating data risk when planning for geographic data collection and managing geographic data that has been collected.

II. Geographic Data Collection Standards

This section defines four levels of geographic detail at which data is collected and provides data collection standards for each level of detail. The standards will be used to define and communicate geographic data collection requirements.

III. Geographic Data Asset Submission Standards

This section outlines the types of geographic data assets that should be submitted to USAID in accordance with **ADS 579** and how they should be submitted. The standards will be used to define and communicate geographic data asset submission requirements.

Audience

USAID should use this document to:

- o Define and communicate geographic data collection and submission requirements in procurement language; and

- o Evaluate and mitigate data risk when planning for data collection and managing geographic data that has been submitted to USAID by implementing partners.

I. Managing Data Risk

USAID development data are collected, managed, and shared for an intended benefit, such as improved development decision making and accountability and transparency. However, USAID development data may also present some level of data risk in addition to benefit. Therefore, balancing benefit and risk is a key component of responsible data collection and management.

There is a common misconception that USAID development data will be publicly released when it is submitted to USAID even when it may pose substantial data risk. This misconception may lead to unnecessarily limiting data collection due to potential data risk or result in overly redacting or aggregating data before it is submitted to USAID. In response, it is important to clarify the following:

1. When data that present risk are submitted to USAID data repositories, USAID can limit data access to: Public, Restricted Public, or Non-Public, as defined in **ADS 579.3.2.4**.

2. To maintain the spirit of openness and transparency, Activity Location Data that present risk redacted or aggregated versions of the dataset that mitigate data risk can be shared instead.

For example, when an implementing partner collects and submits data that presents risk, they can propose the appropriate data access level and recommend how the data should be redacted or aggregated before it is shared.

A. Overview

Defining Data Risk

Three interrelated parts define data risk:

$$\boxed{\text{Negative Outcome}} \rightarrow \boxed{\text{Likelihood}} \times \boxed{\text{Impact}}$$

Negative Outcome: A negative outcome that could result from unauthorized access to the data.
Likelihood: The likelihood that the negative outcome will actually occur.
Impact: The level of impact that would result from the negative outcome.

Negative outcomes include the Principled Exceptions outlined in **ADS 579.3.2.3**. When data include personally identifiable information (PII), safeguarding procedures should be applied to protect PII from unauthorized access (see **ADS 508, Privacy Program** and **Data Security Guidance: Protecting Beneficiaries**). In addition, certain country contexts may present insecure environments that result in safety concerns regarding geographic data collection in the field. When applicable, these data collection constraints should also be considered when planning for data collection.

Geographic Data Benefit and Risk

The location information included in a geographic dataset can be used to identify the locations associated with the other descriptive information contained in the dataset. Therefore, the location information is an additional data attribute to consider when evaluating and mitigating data risk. Collecting data at a high level of geographic detail often increases the benefit of collecting the data yet it may also increase the data risk in certain scenarios due to the detailed location information included in the data. As a result, the level of geographic detail in a dataset is often a determinant of its benefit but also its potential data risk.

B. Evaluating and Mitigating Geographic Data Risk

To maximize the benefit of collecting, using, and sharing geographic data while also accounting for risk, it is important to evaluate data risk at two different stages outlined below. In many data collection and management scenarios, the data risk will not require mitigation, yet it is still important to evaluate data risk and potential mitigation techniques.

I. Stage 1: Planning for Geographic Data Collection

Planning for geographic data collection commonly occurs during design and procurement. During this process, the USAID OU: 1) determines the geographic data that an implementing partner will collect and submit to USAID; and 2) communicates the data collection requirements in the reporting sections of procurement language.

1. **Evaluate Risk:** Will collecting data at the intended level of geographic detail pose substantial data risk? If yes, then:

   Mitigate Risk:

   a. Require that data security measures will be used to protect the data from unauthorized access while collecting, managing and submitting the data; or

   b. Collect the data at a lower level of geographic detail to mitigate the data risk.

2. **Evaluate Risk:** Will collecting data at the intended level of geographic detail require field data collection in insecure environments that may lead to safety concerns? If yes, then:

   Mitigate Risk:

   a. Collect the data at a lower level of geographic detail that will not result in safety concerns and therefore mitigate data risk.

II. Stage 2: Managing Geographic Data Submitted to USAID

Managing geographic data occurs after data is submitted to USAID according to the requirements of an award.

3. **Evaluate Risk:** Will publicly releasing the data at the current level of geographic detail pose substantial risk? If yes, then:

   Mitigate Risk:

   a. Protect the data from unauthorized access by proposing the access level at which data should be shared to sufficiently mitigate data risk: Public, Restricted Public, or Non-Public (see **ADS 579.3.2.4**).

   b. Aggregate the data to a lower level of geographic detail to create a mitigated version of the dataset that is appropriate to share.

III. Key Questions to Consider While Evaluating Data Risk

The following questions will help guide the process of evaluating data risk:

1. At which level of geographic detail should data be collected to maximize its benefit? Which other descriptive data will be collected in addition to the location information that may increase data risk when it's combined with geographic location data?

2. Will a dataset that combines location information at the intended level of geographic detail with the other descriptive information result in high, medium, or low data risk?

   a. Which negative outcome could occur if the dataset is accessed by unauthorized users? How negative is the outcome: high, medium, low?

   b. How likely is it that this outcome would occur if the dataset is accessed by unauthorized users: high, medium, low?

   c. How severe is the impact of the negative outcome: high, medium, or low?

   d. Will the negative outcome, its likelihood of occurring, and/or the severity of its impact change over time?

   e. Will the data include PII? (See **ADS 508** and **Data Security Guidance: Protecting Beneficiaries**).

3. At which level of geographic detail should the dataset be shared to maximize its benefit? If shared at this level of geographic detail, should the access level be Public, Restricted Public, or Non-Public? (See **ADS 579.3.2.4**).

4. If there is significant data risk, which mitigation technique will maximize the benefit of sharing the data while also addressing the data risk?

II. Geographic Data Collection Standards

This section defines four levels of geographic detail and provides data collection standards for each level of detail.

A. How to Use the Geographic Data Collection Standards

The data collection standards assist USAID OUs in defining geographic data collection requirements in procurement language.

Step 1: Identify the Level of Geographic Detail

The appropriate level of geographic detail at which data will be based on, but not necessarily limited to, depends on the following criteria:

- o The level of geographic detail necessary to support the intended use of the data;

- o The cost and capacity required to collect the data at a certain level of geographic detail;

- o Whether collecting data at a certain level of geographic detail will present data risk (see **Managing Data Risk** above); and

- o Whether collecting data at a certain level of geographic detail will not be permissible due to insecure environments.

Step 2: Identify the Appropriate Data Collection Method

Data collection standards are defined for each level of geographic detail and in some cases include standards for different data collection methods, one of which will be most appropriate based on the data collection scenario. For example, consider an activity that will collect several details about a district's health facilities: the facility name, latitude/longitude coordinates, and the number of health providers at the facility. When an existing high-quality dataset with the names and latitude/longitude coordinates of the facilities is accessible, it may not be necessary to capture the latitude/longitude coordinates of each facility using GPS-enabled device. Instead, the name and location data in the existing high quality health facility dataset can be used and data collection can focus on capturing the number of health providers present for each facility. In this scenario, using the existing high-quality health facility dataset rather than a GPS-enabled device is a more appropriate for collecting data at the Exact Site Location level of geographic detail.

Step 3: Define and Communicate the Geographic Data Collection Standards

After determining the level of geographic detail and appropriate data collection method, the standards can be used to define and communicate the data collection requirements in solicitation or post-award guidance documents. The standards language is provided in a generic form so that it can be applied across a range of data collection scenarios. When it is used to communicate data collection requirements, additional context should be provided about the data collection scenario so that the standards are more easily interpreted. For example, when collecting health facility locations at the Exact Site Location level of geographic detail, replace "exact site location" with "health facility" so the standard language specifically reflects the data collection scenario.
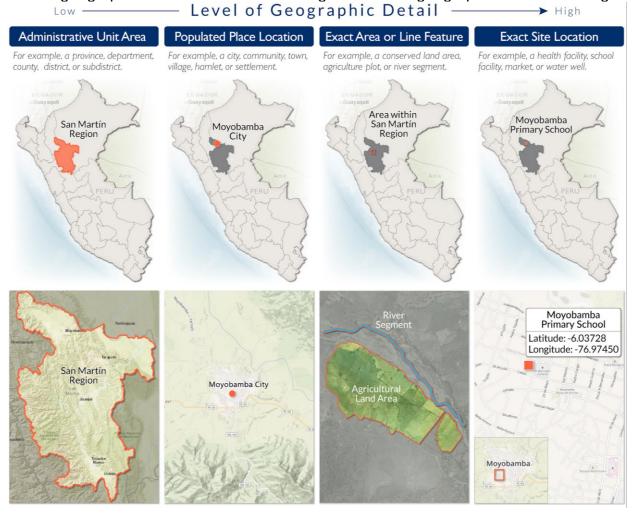
B. The Four Levels of Geographic Detail

Table 1 outlines the four level of geographic detail. An example is also provided to illustrate each level of detail. When determining the appropriate level of geographic detail for data collection, it is important to consider that data collected at a higher level of geographic detail, such as the Exact Site Location of a facility, can be aggregated to a lower level of geographic detail, such as the Administrative Unit Area, e.g., a district. However, data collected at a lower level of geographic detail cannot be disaggregated to a higher level of geographic detail. Therefore, data collected at a higher level of geographic detail is more flexible for a broader range of analysis and visualization purposes.

Table 1 - Levels of Geographic Detail

Level of Geographic Detail	Definition	Example
Administrative Unit Area	Data is collected by administrative unit, such as district, and visually represented by the entire area of the administrative unit.	The number of community of health workers by district.
Populated Place Location	Data is collected by populated place, such as a city or town, and visually represented as the point location of the populated place using the latitude/longitude coordinates.	The location of supply chain warehouses experiencing chronic stock outs by city or town.
Exact Area or Line Feature	Data is collected by area or line features and visually represented as an area or line.	The agriculture land areas within a community; or river segments that are located in conservation areas.
Exact Site Locations	Data is collected by exact site location and visually represented as the point location of the exact site using the latitude/longitude coordinates.	The exact location of cell towers within a district.

Example: Levels of Geographic Detail in Peru

The maps below display the different level of geographic detail in Peru, from the lowest level of geographic detail on the left to the highest level of geographic detail on the right.

Administrative Unit Definition

Administrative Units define the hierarchy of administrative divisions within a country. The units are delineated by geographic boundaries, starting with the country boundary that is referred to as administrative unit 0. The first level of subdivision within a country is referred to as the administrative unit 1. Administrative 1 units are subdivided into administrative 2 units and this logic continues until the lowest level of subdivision is reached. Terms such as "Admin 1" or "Adm1" are commonly used to refer to administrative units and frequently appear in column headings or filenames for datasets. Each country will refer to the administrative units with different terms for each level, such as region, province, and district.

As an illustrative example, Table 2 displays the names and number of units for each administrative level in Peru in 2018.

Table 2 - Levels of Geographic Detail

Administrative Unit Level	Name	Number of Units
1	Region	26
2	Province	196
3	District	1,869

C. Geographic Accuracy

Geographic accuracy refers to how closely the location data captured in a geographic dataset indicates the actual location of a feature on the ground, such as the location of a school facility. When collecting latitude/longitude coordinates with GPS-enabled devices, the geographic accuracy required will depend on how the data will be used. For example, if school location data will be collected for general planning purposes, collecting latitude/longitude coordinates that are within 15m of the actual school locations will provide sufficient geographic accuracy for the intended use of the data. The standards defined below that apply to collecting latitude/longitude coordinates with GPS-enabled devices will fulfill a majority of USAID data collection scenarios yet certain data collection scenarios will require that data are collected at a higher level of geographic accuracy. In these scenarios, the data collection standards can be adjusted to require a higher level of geographic detail accuracy. Example data collection scenarios that may require higher levels of geographic accuracy include, but are not limited, to the following: land tenure and resource management; forestry and biodiversity; programming that occurs within urban areas; infrastructure; agriculture and food security; and countering violent extremism.

D. Geographic Data Collection Standards by Level of Geographic Detail

Data collection standards are defined below for each level of geographic detail and Table 3 lists the different data collection methods for each level of detail. In some cases, the standards include more than one data collection method to ensure that the standards are applicable to a range of data collection scenarios.

Table 3 - Data Collection Methods

Level of Geographic Detail	Data Collection Methods
I. Administrative Unit Areas	○ Existing Administrative Unit Dataset
II. Populated Place Locations	○ Existing Populated Place Dataset ○ Geocoding Tool
III. Exact Area or Line Features	○ Existing Exact Area or Line Dataset ○ GPS-enabled Device ○ Digitizing High Resolution Satellite/Aerial Imagery
IV. Exact Site Locations	○ Existing Exact Site Location Dataset ○ GPS-enabled Device ○ Digitizing High Resolution Satellite/Aerial Imagery

I. Level of Geographic Detail: Administrative Unit Areas

Example: Number of health facilities providing basic maternal and child health services by district.

Data Collection Method: Administrative Unit Dataset

Description: When accessible, it is best practice for a USAID OU to provide either an authoritative or high quality administrative unit dataset to the implementing partner, such as a district or sub-district dataset. An authoritative administrative unit dataset comes from a host-country government agency, such as the statistics or planning agency, and contains administrative unit names and codes. If the administrative unit dataset is available in a GIS format it will also contain authoritative geographic boundaries. Using the same authoritative administrative unit names, codes, and boundaries among multiple activities will standardize the data collected at this level of geographic detail and increase its interoperability with data created by the host-country government. When an authoritative dataset does not exist or contains known errors, then a separate dataset that is considered high quality should be used.

Data Collection Standards:

1. When an authoritative administrative unit dataset is accessible, USAID may require the use of authoritative unit names, codes, and boundaries, as geographic location identifiers for data that is collected about the administrative units. For example, if USAID requires data collection on the number of schools disaggregated by district, the authoritative district names, codes, and boundaries

will be used as the geographic location identifiers for the district level data collected. If variations of administrative unit names are commonly used, the name variations can be included in addition to the authoritative unit names.

2. When an authoritative administrative unit dataset is not accessible, USAID may require the use of administrative unit names, codes, and boundaries from the best available non-authoritative administrative unit dataset as geographic location identifiers for data that is collected about the administrative units.

3. Supplementary information that may be necessary to effectively use the data in the future should be documented during data collection, including but not limited to the following:

   o The date and source of the existing administrative unit dataset;

   o A description of when and how the data was collected;

   o The level of confidence in data quality; and

   o If there is any potential data risk associated with the data.

   When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

4. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

II. Level of Geographic Detail: Populated Place Locations
Example: The populated places where technical and vocational education facilities are located.

Data Collection Method 1: Populated Place Dataset
Description: When accessible, it is best practice for a USAID OU to provide either an authoritative or high quality populated place dataset to the implementing partner. An authoritative populated place dataset comes from a host-country government agency, such as the statistics or planning agency, and contains the names and latitude/longitude coordinates for populated places within a country. Using the authoritative populated place names and latitude/longitude coordinates among multiple activities will standardize the data collected at this level of geographic detail and increase its interoperability with data created by the host-country government. When an authoritative dataset does not exist or contains known errors, then a separate dataset that is considered high quality should be used.

Data Collection Standards:

1. When an authoritative populated place dataset is accessible, USAID may require the use of authoritative populated place names and latitude/longitude coordinates as geographic location identifiers for data that is collected about the populated places. For example, if USAID requires data collection on the number of health facilities disaggregated by town, the authoritative town names and latitude/longitude locations will be used as the geographic location identifiers for the data collected. If variations of populated place names are unofficially used, the name variations can be included in addition to the authoritative populated place name.

2. When an authoritative populated place dataset is not accessible, USAID may require the use of populated place names, codes, and boundaries from the best available administrative unit dataset as geographic location identifiers for data that is collected about the populated places.

3. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

   o The date and source of the existing populated place dataset;

   o A description of when and how the data was collected;

   o The level of confidence in data quality; and

   o If there is any potential data risk associated with the data.

   When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

4. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

Data Collection Method 2: Geocoding Tool
Description: Web-based geocoding tools are used to retrieve the latitude/longitude coordinates of populated places by searching for the name of the populated place.

Data Collection Standards:

1. If authoritative latitude/longitude coordinates for populated places is not accessible but a list of authoritative names is accessible, USAID may require the

use of authoritative names to retrieve latitude/longitude coordinates for the populated places using a geocoding tool. The authoritative names and the latitude/longitude coordinates retrieved from the geocoding tool will be used as the geographic location identifiers for data collected about the populated places.

2. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

   o The name of the geocoding tool;

   o A description of when and how the data was collected;

   o The level of confidence in data quality; and

   o If there is any potential data risk associated with the data.

   When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

3. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

III. Level of Geographic Detail: Exact Area or Line Features
Example: Exact boundaries of agricultural areas or the exact line segments of primary rivers.

Data Collection Method 1: Exact Line or Area Feature Dataset
Description: When accessible, it is best practice for a USAID OU to provide either an authoritative or high quality exact area or line feature dataset to the implementing partner. An authoritative exact line or area feature dataset comes from a host-country government agency and contains the authoritative names, line segments or geographic boundaries, and any other relevant information; for example, a roads dataset from the transportation agency or protected area boundaries from the conservation agency. Using the authoritative names and line segment or geographic boundary data among multiple activities that all collect data using the same exact area or line feature data and its interoperability with data created by the host-country government. When an authoritative dataset does not exist or contains known errors, then a separate dataset that is considered high quality should be used.

Data Collection Standards:

1. When an authoritative exact line or area feature dataset is accessible, USAID may require the use of names, line segments or geographic boundaries as geographic

location identifiers for data that is collected about the line or area features. For example, if USAID requires data collection on the travel speed of different road types within a district, USAID may require the use of authoritative name and line segments as the geographic location identifiers for the data collected about the roads.

2. When an authoritative exact line or area feature dataset is not accessible, USAID may require the use of names and line segments or geographic boundaries from the best available exact line or area feature dataset as geographic location identifiers for data that is collected about the line or area features.

3. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

   o The date and source of the existing exact line or area feature dataset used;

   o A description of when and how the data was collected;

   o The level of confidence in data quality; and

   o If there is any potential data risk associated with the data.

   When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

4. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

Data Collection Method 2: GPS-Enabled Device
Description: GPS-enabled devices are used to manually trace a line feature segment or the boundaries of an area feature. This is commonly done by recording latitude/longitude coordinates at a set time or distance interval while walking or driving with the GPS-enabled device along the path of the area boundary or line segment. The latitude/longitude coordinates create point locations that become the vertices of the line or area segment. To standardize data collection with GPS-enabled devices, requirements for several key device settings should be established. For example, if two implementing partners collect the same data with a GPS-enabled device yet use a different coordinate system/map datum setting, the data will not be comparable.

Data Collection Standards:

1. At a minimum, USAID may require that the following settings are used on a GPS-enabled device to collect the area feature boundaries or line feature segments. An implementing partner may enhance the standard settings to increase geographic accuracy, when required.

Parameter Name	Description
Accuracy Level	Less than 15 meters
Time Interval	Two seconds for driving; five seconds for walking
Number of Satellites	Four or more
Navigation Units	Metric
Coordinate System/Map Datum	World Geodetic System 1984 (WGS 84)
Coordinate Format	Decimal Degrees (±ddd.ddddd) with at least five decimal places
North Reference	True North
Latitude Coordinates	Between the minimum and maximum latitude coordinates of the implementation area
Longitude Coordinates	Between the minimum and maximum longitude coordinates of the implementation area

2. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

- o The name of the GPS-enabled device and applications used;

- o A description of when and how the data was collected;

- o The level of confidence in data quality; and

- o If there is any potential data risk associated with the data.

When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

3. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

Data Collection Method 3: Digitizing High Resolution Satellite/Aerial Imagery

Description: High resolution satellite/aerial imagery is used to visually identify and trace area feature boundaries or line feature segments to create geographic data. This process involves using imagery and a GIS tool or using a Web-based tool that includes the functionality to view imagery and trace area or line features. The geographic accuracy of the data created depends on several factors, including the digitizer's ability to accurately identify the contours of the area or line features and the positional accuracy of the imagery. Therefore, it is important that information describing how the data was created is provided as this information will be required to appropriately use the data.

Data Collection Standards:

1. At a minimum, USAID may require that the following information is recorded when digitizing high resolution satellite/aerial imagery to collect data on the boundaries of area features or segments of line features. When additional information is relevant, it may also be provided.

Name	Description	
Image Source	The source of the satellite/aerial imagery	
Image Date	The date of the satellite/aerial imagery	
Image ID	The satellite/aerial imagery identification code, when available	
Precision Code	Exact Location	The latitude/longitude coordinates for the vertices of the exact area or line feature are accurate within 15m
	Near Exact Location	The latitude/longitude coordinates for the vertices of the exact area or line feature are estimated to be accurate within 15m

2. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

- o The tools used to digitize the data;

- o A description of when and how the data was collected;

- o The level of confidence in data quality; and

- o If there is any potential data risk associated with the data.

When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

3. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

IV. Level of Geographic Detail: Exact Site Locations

Example: The exact site locations of primary school facilities.

Data Collection Method 1: Exact Site Location Dataset

Description: When accessible, it is best practice for a USAID OU to provide either an authoritative or high quality exact site location dataset to the implementing partner, such as a health facility dataset. An authoritative exact site location dataset comes from a host-country government agency and contains the authoritative names, latitude/longitude coordinates, and any other relevant information. Using the authoritative exact site location names and latitude/longitude coordinates will standardize data collection among multiple activities that all collect data using the same exact site locations and its interoperability with data created by the host-country government. When an authoritative dataset does not exist or contains known errors, then a separate dataset that is considered high quality should be used.

Data Collection Standards:

1. When an authoritative exact site location dataset is accessible, USAID may require the use of authoritative names and latitude/longitude coordinates as geographic location identifiers for data that is collected about the site locations. For example, if USAID requires data collection on the number of providers present at a health facility, the implementing partner will use the authoritative name and latitude/longitude coordinates as the geographic location identifiers for the data collected about the health facilities.

2. When an authoritative exact site location dataset is not accessible, USAID may require the use of names and latitude/longitude coordinates from the best available exact site location dataset as geographic location identifiers for data that is collected about the site locations.

3. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

- o The date and source of the existing exact site location dataset of the GPS-enabled device and applications used;

- o A description of when and how the data was collected;

- o The level of confidence in data quality; and

     ○ If there is any potential data risk associated with the data.

When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

4. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

Data Collection Method 2: GPS-enabled Device

Description: GPS-enabled devices are used to manually collect the latitude/longitude coordinates of exact site locations. To standardize data collection with GPS-enabled devices, requirements for several key device settings should be established. For example, if two implementing partners collect the same data with a GPS-enabled device yet use a different coordinate system/map datum setting, the data will not be comparable.

Data Collection Standards:

1. At a minimum, USAID may require that the following settings are used on a GPS-enabled device to collect the boundaries of an exact area feature or the segment of a line feature. An implementing partner may enhance the standard settings to increase geographic accuracy, when required.

Parameter Name	Description
Accuracy Level	Less than 15 meters
Number of Satellites	Four or more
Navigation Units	Metric
Coordinate System/Map Datum	World Geodetic System 1984 (WGS 84)
Coordinate Format	Decimal Degrees (±ddd.ddddd) with at least five decimal places
North Reference	True North
Latitude Coordinates	Between the minimum and maximum latitude coordinates of the implementation area
Longitude Coordinates	Between the minimum and maximum longitude coordinates of the implementation area

2. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

o The name of the GPS-enabled device and applications used;

o A description of when and how the data was collected;

o The level of confidence in data quality; and

o If there is any potential data risk associated with the data.

When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

3. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

Data Collection Method 3: Digitizing High Resolution Satellite/Aerial Imagery
Description: High resolution satellite/aerial imagery is used to visually identify and create latitude/longitude coordinate locations for exact site locations. This process involves using imagery and a GIS tool or using a Web-based tool that includes the functionality to view imagery and create latitude/longitude point locations. The geographic accuracy of the data created depends on several factors, including the digitizer's ability to accurately identify the site locations and the positional accuracy of the imagery. Therefore, appropriately using the data requires an understanding of how it was created, and this information should be included with the data.

Data Collection Standards:

1. At a minimum, USAID may require that the following information is recorded when digitizing high resolution satellite/aerial imagery to collect data on exact site locations. When additional information is relevant, it may also be provided.

Name	Description	
Image Source	The source of the satellite/aerial imagery	
Image Date	The date of the satellite/aerial imagery	
Image ID	The satellite/aerial imagery identification code, when available	
Precision Code	Exact Location	The latitude/longitude coordinates for the vertices of the exact area or line feature are accurate within 15m
	Near Exact Location	The latitude/longitude coordinates for the vertices of the exact area or line feature are estimated to be accurate within 15m

2. Supplementary information that may be necessary to use the data effectively in the future should be documented during data collection, including but not limited to the following:

   o The tools used to digitize the data;

   o A description of when and how the data was collected;

   o The level of confidence in data quality; and

   o If there is any potential data risk associated with the data.

   When data are stored in Shapefile (.shp) formats, this information will be included as part of metadata using the ISO 19139 XML implementation schema that fulfills ISO 19115.

3. The data submission standards will be referenced prior to data collection so that the collection and management procedures lead to fulfilling the data submission standards.

III. Geographic Data Asset Submission Standards

Standard provisions are included in USAID awards to support the implementation of **ADS 579**. In accordance with the award requirements, implementing partners should submit geographic data to USAID. The geographic data asset submission standards below define the types of geographic data assets that should be submitted and how.

A. Types of Geographic Data Assets

1. Geographic Data Files

o Geographic Data Files identify the geographic location and characteristics of natural, demographic, cultural, political, or constructed features on the Earth; for example, a school location dataset or a literacy rate by district dataset.

o Geographic Data Files include datasets that are stored in file formats traditionally associated with geographic data, such as a Shapefile (.shp) or GeoTIFF (.tif), as well as other file formats that may not be as commonly associated with geographic data, such as a CSV (.csv). For example, a CSV (.csv) file that includes descriptive information and geographic location identifiers, such as total population by district listed in a table.

o When derived Geographic Data Files are generated by processing or analyzing primary Geographic Data Files, both the derived and primary Geographic Data Files will be submitted.[1]

2. Geographic Data Processing and Analysis Files

o Geographic Data Processing and Analysis Files execute data processing and analysis methods or describe a workflow and are commonly referred to as scripts, tools, or models:

a. Data processing: Existing data is manipulated through a processing transformation.

b. Data analysis: New data is derived based on input data.

c. Workflow Description: A written description of the steps performed in a data processing and/or analysis workflow.

o Geographic Data Processing and Analysis Files will be submitted in addition to the Geographic Data Files so that an end user can understand and replicate the data processing and/or analysis that was performed (see **ADS 579.3.2.6 c**). For example, a script written in the Python programming language and used to manipulate or analyze data is considered a Geographic Analysis File and should be submitted.[2]

3. Geographic Data Visualization Files

o Geographic Data Visualization Files are used to generate static or interactive data visualization products presented in hard copy, digital, or Web-based formats, for example:

a. Static: Files used to create a hard copy visualization product and stored in an Esri Map Document (.mxd) or QGIS Project Document (.qgs).

b. Interactive: Files used to create a Web-based data visualization product that is written in programming languages, such as HTML, CSS, or Javascript.

o Multiple files are often required to generate a geographic data visualization product and each relevant file should be submitted.

---

[1] See the definition of "Complete" under the Open Data in section I. Definitions of OMB M-13-13: **https://project-open-data.cio.gov/policy-memo/#i-definitions**

[2] See the definition of "Described" under the Open Data in section I. Definitions of OMB M-13-13: **https://project-open-data.cio.gov/policy-memo/#i-definitions**

B. Geographic Data Asset Submission Requirements

1. Geographic Data Files

    a. Both primary and derived Geographic Data Files should be submitted.[3]

    b. To fulfill the submission requirements, Geographic Data Files should be submitted in Shapefile (.shp), GeoTIFF (.tif), CSV (.csv), or GeoJSON (.geojson) file formats.

    c. In addition to fulfilling the submission requirements noted above, Geographic Data Files that are stored in other file formats that may be useful to end users can be voluntarily submitted in those file formats. For example, Geographic Data Files stored in a File Geodatabase (.gdb) or as a Web Map Service (WMS) can also be submitted.

    d. Geographic Data Files that are stored in Shapefile (.shp) or GeoTIFF (.tif) formats will include metadata that follows ISO 19115 using the ISO 19139 XML implementation schema.[4]

    e. Geographic Data Files that are stored in a CSV format and contain latitude/longitude coordinates will include supplementary information regarding:

    o The method used to create the latitude/longitude coordinates;

    o The spatial reference used to generate the latitude/longitude coordinates; and

    o The geographical precision of the latitude/longitude coordinates using the International Aid Transparency Initiative standards.[5]

    For example, a latitude/longitude coordinate may represent an exact location, the center of a populated place, or the center of subnational administrative unit.[6] This information should be included as an attachment to the actual CSV dataset.

    f. Geographic Data Files that are stored in Shapefile (.shp), GeoTIFF (.tif), or GeoJSON (.geojson), JSON (.json) file formats will use the Geographic Coordinate System World Geodetic System 1984 (GCS WGS 1984) spatial reference. If the use of a different spatial reference was required for an

---

[3] See the definition of "Complete" under the Open Data in section I. Definitions of OMB M-13-13: https://project-open-data.cio.gov/policy-memo/#i-definitions

[4] For example, see the instructions on changing the metadata style in Esri ArcGIS Desktop: http://desktop.arcgis.com/en/desktop/latest/manage-data/metadata/choosing-a-metadata-style.htm

[5] For example, see the International Aid Transparency Initiative Geographical Precision coding system: http://iatistandard.org/201/codelists/GeographicalPrecision/

[6] See the definition of "Described" under the Open Data in section I. Definitions of OMB M-13-13: https://project-open-data.cio.gov/policy-memo/#i-definitions

analysis, the spatial reference that was used should be indicated elsewhere in the metadata.

2. Geographic Data Processing and Analysis Files

a. Geographic Data Processing and Analysis Files that are written in programming languages will include a description of the language they are written in (for example, a Python file would be .py, a Javascript file would be .js, etc.), the intended purpose of the file, use limitations, and execution instructions.

b. Geographic Data Processing and Analysis Files will include descriptive comments and relative pathnames to enable efficient and effective use by an end user.

c. Geographic Data Processing and Analysis Files are stored in an online repository, such as GitHub. These files can also be submitted as a clone of the repository.

3. Geographic Data Visualization Files

a. Geographic Data Visualization Files created using desktop GIS software will be submitted in Esri Map Document (.mxd) or QGIS Project (.qgs) file formats.

b. Geographic data visualization products that are finalized in separate desktop software or stored in a different file format can also be submitted if they may be of use to an end user. For example, when a data visualization product is stored as a Map Package or is finalized in Adobe Illustrator, the Map Package (.mpk), or Adobe Illustrator (.ai) file can be additionally submitted.

c. Geographic Data Visualization Files written in a programming language and used to generate Web-based data visualization products will include an indication of the programming language (for example, a Python file would be .py, a Javascript file would be .js, etc.).

d. Geographic Data Visualization Files will reference relative pathnames to the source of all Geographic Data Files that are displayed in the data visualization product.

e. When products created with the Geographic Data Visualization Files are included in documents that are submitted to the USAID Development Experience Clearinghouse (DEC), such as a report that includes a map created with a Geographic Data Visualization File, the URL that links to the location of the document in the DEC should be provided.

579saa_073118