PROGRAM CYCLE

USAID
FROM THE AMERICAN PEOPLE

ADS 201 Additional Help

DATA DISAGGREGATION BY GEOGRAPHIC LOCATION

Introduction

This document provides supplemental guidance to ADS 201.3.5.7.G Indicator Disaggregation, and discusses concepts and methods needed to collect and analyze geographically disaggregated indicators for improved performance monitoring, learning, and adapting. This document will help managers and Contracting Officer's Representatives, Agreement Officer's Representatives, and Government Agreement Technical Representatives (CORs/AORs/GATRs) make decisions about disaggregating indicator data by geographic location, and provide information about how to collect and use the geographically disaggregated data. The content of this document has been organized into sections that address topics relevant to various roles at a Mission or Washington Operating Unit (OU).

Value-Added when Using Geographically Disaggregated Data

Disaggregating data by geographic location is important for two reasons. First, geographic data provide the foundation for investigating geographic variation in performance. This enables managers and CORs/AORs/GATRs to ask questions, such as: “Does an activity’s performance vary across the geographic area where it is implemented?” and “Where is an activity over- or under-performing?”

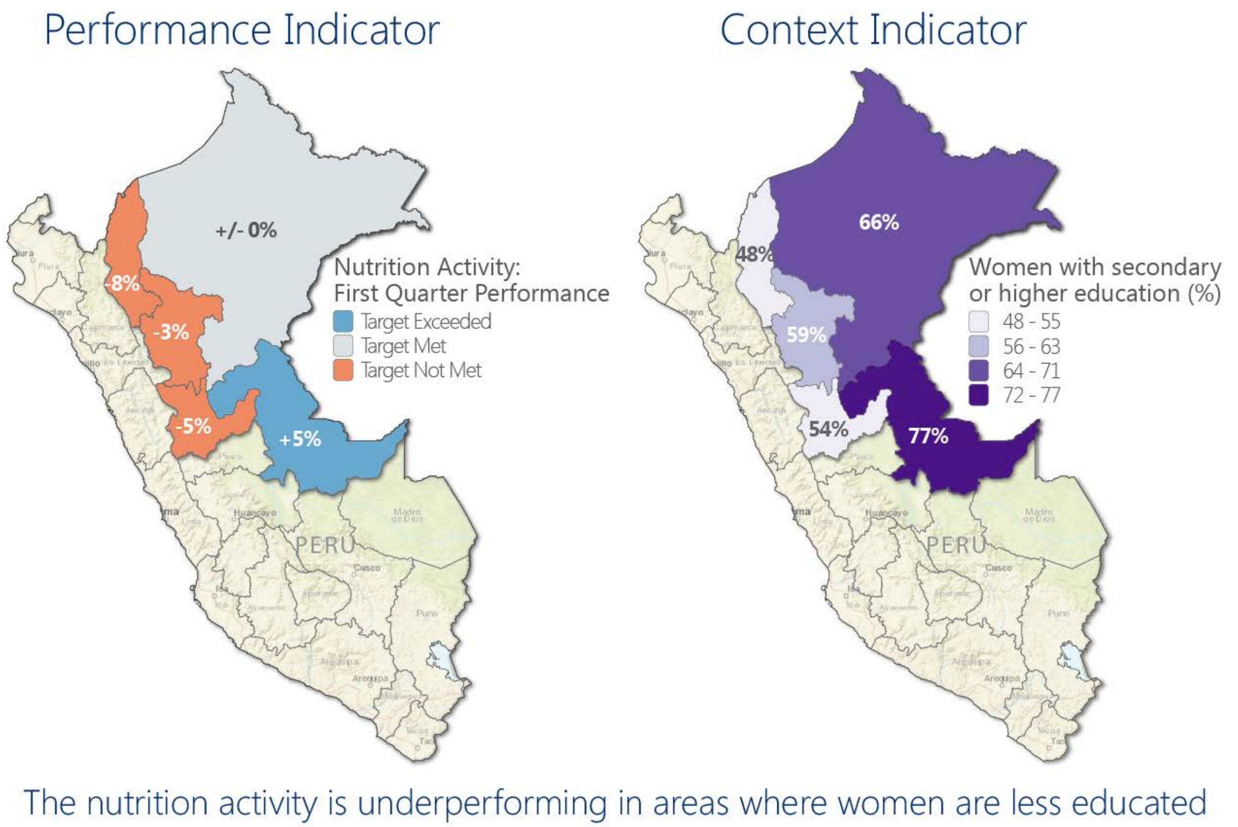
Second, disaggregating data by geographic location creates a set of unique identifiers (e.g., administrative units or populated places) which provide location information for each observation in the data. These same unique identifiers are also present in other geographically disaggregated datasets, such as performance or context indicators from other projects. By creating this common link across multiple datasets, geographically disaggregated data allow one to ask and answer additional questions, such as: “Does the geographic variability in an activity’s performance relate to geographic variation observed in a context indicator?” and “How does the pattern of geographic variation in one activity’s performance compare to the variation in another activity’s performance?”

Together, geographic data and analysis do not drastically alter the monitoring questions one seeks to answer, but add a geographic dimension to the questions that are asked and how they are answered. Such data analysis can strengthen the practice of monitoring and adaptive management.

Program Cycle Additional Help documents provide non-mandatory guidance intended to clarify ADS 201. Curated by the Bureau for Policy, Planning and Learning (PPL), these may include “how-to” guidelines, templates and examples of best practice.

Example: Peru – Comparing Performance and Context Indicators

With geographically disaggregated indicators, one can explore questions, such as: "Does the nutrition activity appear to perform better (performance indicator) in areas where female education rates (context indicator) are higher? Why might this be?" This type of analysis can support learning and adaptation of programs.



Key Management Decisions

Determining the level of geographic detail to disaggregate by is a key management decision. While collecting indicator data at a high level of geographic detail (e.g., latitude/longitude point locations) can provide a comprehensive amount of information, that level of geographic detail is not always necessary. It is useful to define the lowest common geographic disaggregation that will be used by the entire Mission in the Mission-wide Performance Management Plan (PMP). This allows for data comparison across projects and activities, providing insights into how the Mission's interventions may be affecting one another. A project team may decide to collect data disaggregated by a greater level of geographic detail to provide more precise analysis. The Mission's Mission Order on Performance Monitoring should document the roles and responsibilities involved in collecting and using geographically disaggregated monitoring data.

The first part of this section defines levels of geographic detail at which data can be disaggregated. The second part summarizes key considerations to determine the most appropriate level of geographic detail.

GEOGRAPHIC SCALE

The geographic scale at which indicator data are disaggregated affects the type of information that the data provide, as well as the flexibility of using the data for a range of analyses. For the purposes of this document, the term "geographic scale" is used to describe the level of geographic detail at which indicator data are disaggregated.

Example: Map and Table View of Geographic Disaggregation in Peru

Map View: The maps below display a range of geographic scales at which indicator data can be collected, from the lowest level of geographic detail on the left **(**administrative unit 1) to the highest level of geographic detail on the right (exact location using latitude/longitude coordinates).

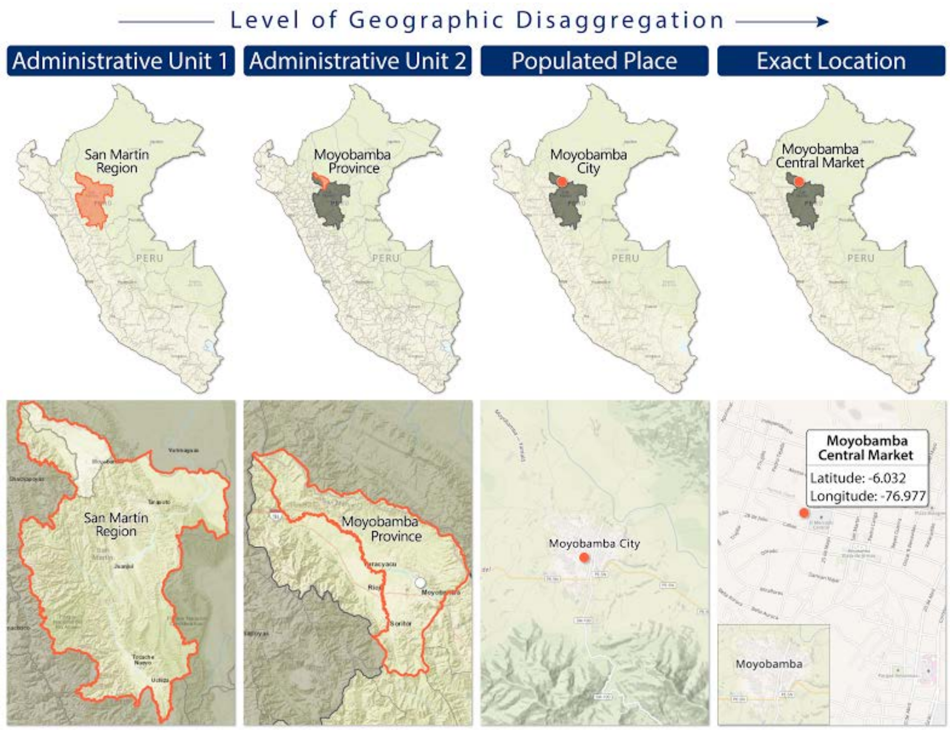


Table View: The table below displays the same data that appears on the maps above, where the first row corresponds to the map on the left and the last row corresponds to the map on the right. This illustrates that disaggregating at a higher level of geographic detail can provide greater flexibility to analyze and compare data.

Activity/IM Name	Administrative Unit 1	Administrative Unit 2	Populated Place	Latitude	Longitude
Improved Nutrition Now	San Martín	←	Can aggregate but not disaggregate		
Improved Nutrition Now	San Martín	Moyobamba	←		
Improved Nutrition Now	San Martín	Moyobamba	Moyobamba City	←	
Improved Nutrition Now	San Martín	Moyobamba	Moyobamba City	-6.032	-76.977

← Geographic Aggregation

As the table above illustrates, data that are collected at high levels of geographic detail can be aggregated to lower levels of geographic detail, but data that are collected at low levels of geographic detail (e.g., the administrative unit I) cannot be disaggregated to higher levels of geographic detail (e.g., a specific city). In other words, data containing only an administrative unit 1 level of geographic detail conceal more detailed geographic variation in the indicator data (e.g., the detail shown in the previous example of nutrition vs. education), and therefore do not provide the type of information that would support all analysis needs. In contrast, data that are collected at high levels of geographic detail often provide more useful information by depicting geographic variation in indicator performance in more detail while also being flexible for aggregation to lower levels of geographic detail, as needed.

Even when performance monitoring data are used at a low level of geographic detail (e.g., region or province), for routine performance management and reporting purposes, it may be beneficial to require collection and submission of data at higher levels of detail (e.g., latitude/longitude) to maintain the flexibility for more detailed investigation of the data. It is useful to discuss with partners these data needs, and the cost and feasibility of collecting such data. In some cases, partners are already collecting data at a high level of geographic detail and then aggregate the data to a lower level when reporting to the Mission.

DEFINING A MINIMUM GEOGRAPHIC SCALE

Defining a minimum geographic scale for indicator data in the Mission-wide PMP and within project teams in the Project MEL Plan, will result in a common geographic denominator among indicators, and will allow for the data to be used for a range of performance management needs throughout the Program Cycle.

Often, projects may require that indicator data are collected at a higher level of geographic detail than the minimum geographic scale set for the PMP, and activities may require that indicator data are collected at a higher level of geographic detail than the minimum geographic scale set for the project. This allows for project indicator data to be aggregated to meet the minimum geographic scale set for the PMP, and for the activity indicator data to be aggregated to the minimum geographic scale set for the project. In this way, the level of geographic detail required for performance analysis at the project and activity level is maintained while also meeting the minimum requirement for Mission-wide analysis.

The following questions are useful when defining a minimum geographic scale for geographically disaggregated data for the Mission in the PMP and the project team in the Project MEL Plan:

Mission-wide (documented in the PMP)

Unit of Analysis: Which geographic unit of analysis will provide the most useful performance monitoring information across the PMP?

Country Context: Which factors shape the country context?

- Is the administrative unit 1 or administrative unit 2 very large? Will data provide the level of detail required for analysis if collected at this geographic scale?
- Is the Mission's portfolio focused on one area of the country that is represented by only a few administrative units at the first level? Will it be most useful to collect data at the administrative unit 2 level or higher?

Mission Programming: Which type of programming does the Mission primarily implement?

- Does the Mission mostly focus on programming that supports specific locations (e.g., health or education facilities, populated places, etc.) and should data disaggregation reflect this?

Project Team (documented in the Project MEL Plan)

Unit of Analysis: Which geographic unit of analysis will provide the most useful performance monitoring information across the project?

Project Context: Which factors shape the project context?

- Are the project activities all implemented in one area of the country, represented by a select number of administrative unit 1 or administrative unit 2 regions? Is a higher level of geographic detail needed to understand any performance variation within these regions?
- What is the minimum level of geographic detail required by the activity level indicators? Note: The minimum geographic scale for project level indicators cannot be a higher level of detail than what is collected by the activity.

Project Programming**:** Which type of programming does the project primarily implement?

- Is the project mostly focused on activities that support specific locations (e.g., populated places, health or education facilities, etc.) and should data disaggregation reflect this?
- Is the project collecting data on vulnerable populations, about politically or socially sensitive topics, or in less populated areas where risks could be associated with disaggregated data?

Example: Defining a minimum geographic scale
A Mission defines the minimum geographic scale at administrative unit 2 for the PMP. A project focused on food security defines the minimum geographic scale at administrative unit 3. An activity within the project collects data on the "number of hectares under improved management practices," by delineating individual geographic areas in hectares. For project level analysis data are aggregated to administrative unit 3, and for PMP level analysis the project level data are aggregated to administrative unit 2. Aggregating to these lower levels of geographic detail, data are comparable to other indicator data and can be used in a range of performance monitoring analyses, while maintaining the level of geographic detail required for the project and activity level analyses.

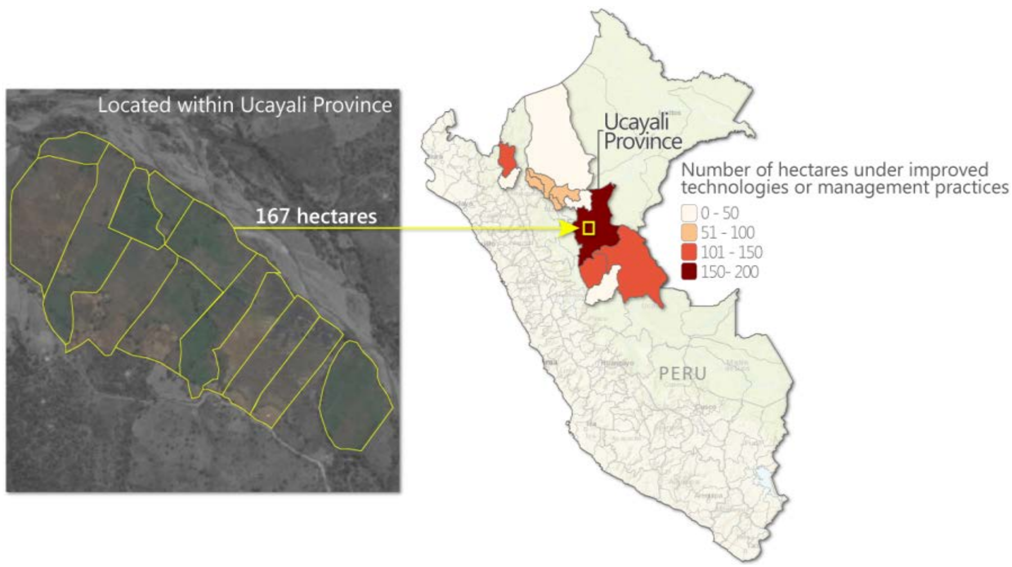
Identify Potential Privacy and Security Implications
Similar to collecting, storing, or using any data when determining a minimum geographic scale for the Mission or project team, it is important to consider potential privacy or security implications. Even when personal data are anonymized, data with high levels of geographic detail could be used to identify an individual, especially in combination with other data sets. One way to mitigate security concerns is to aggregate data to a geographic scale that would not single-out the beneficiary. To ensure adequate consideration of privacy and security concerns when collecting geographic data, work with a sector specialist who is aware of the country's context, a Mission's security specialist, or a GIS specialist. For more information about securely collecting, storing, reporting, and using monitoring data, please see the Data Security guidance document in the Monitoring Toolkit.

USE CAUTION WHEN ANALYZING AGGREGATED GEOGRAPHIC DATA

Aggregating geographically disaggregated data can influence the results of an analysis and the way that the data will be interpreted. Aggregating data can create artificial geographic patterns and resulting analysis can differ at each level of aggregation (this is known as the [Modifiable Areal Unit Problem](#)). While collecting data at high levels of geographic detail provides data that are the most flexible for use in a range of performance monitoring analyses, it is important to understand how the process of aggregating data will affect analysis and the conclusions drawn from the analysis, illustrated in the following example.

Example: Peru Food Security and Nutrition Example

A food security activity collects data on the "number of hectares under improved management practices." The activity collects indicator data that show 167 hectares under improved management practices, all 167 hectares are located in an isolated area of Ucayali province. Later, the indicator data are aggregated to the administrative unit 2 level and are analyzed in the map on the right. When using this visual analysis to compare these results, Ucayali seems to outperform all other provinces. However, this analysis is misleading and conceals the fact that all 167 hectares are actually in an isolated location and not equally distributed throughout the Ucayali province. In this example, aggregating the indicator data results in misinterpretation of the data and leads to management decisions that lack understanding of the activity's true performance.



The Process of Disaggregating Data by Geographic Location

At this stage, a minimum geographic scale has been defined and documented in the PMP and Project MEL Plans, setting basic parameters for all indicator data disaggregated by geographic location. The next steps are to decide: (1) which indicators will provide useful information if they are disaggregated by geographic location; (2) the level of geographic detail at which data are collected to be useful and flexible; and (3) the resources needed to collect and use the data effectively.

SELECTING INDICATORS TO DISAGGREGATE BY GEOGRAPHIC LOCATION

To assist the process of determining which indicators are most suitable for geographic disaggregation, it is helpful to consider indicators as falling into one of three categories: Type I, Type II, and Type III.

Type I
Geographic Measurement: Not required
Geographic Analysis: Limited value

Type I indicators do not require geographic disaggregation and offer limited value in geographic analysis.

Example: Peru - Type I Indicator

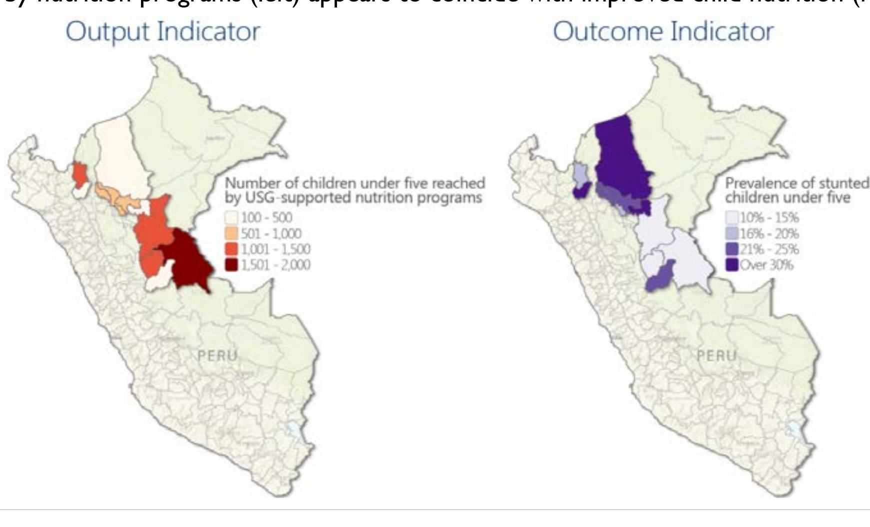
An indicator is designed to measure the number of sectors represented in a national government nutrition working group that meets in the country's capital city. The data can be geographically disaggregated to display that five sectors are represented in the working group, but the location of the capital city is the only geographic information to associate with the performance indicator. As a result, there is limited value in conducting geographic analysis of the data because it won't provide any additional information for performance management.

Type II
Geographic Measurement: Not required but possible
Geographic Analysis: Valuable

Type II indicators do not require geographic measurement but they can be geographically disaggregated, and there is value in geographic analysis of the data.

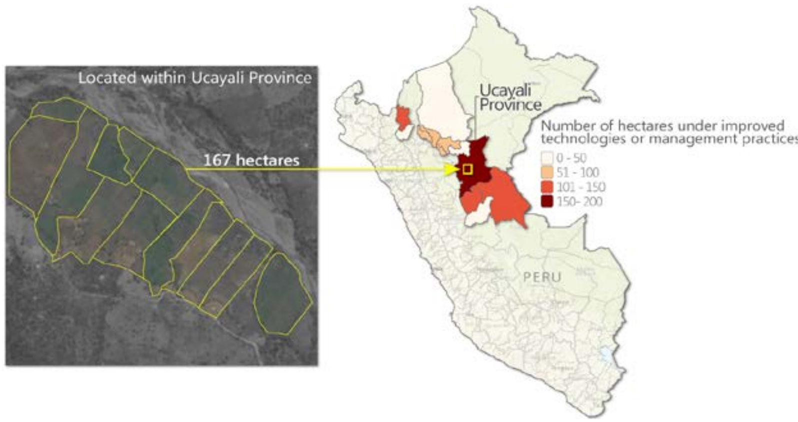
Example: Peru - Type II Indicator

The indicator "number of children under five reached by USG-supported nutrition programs," is used to monitor performance. The data can be geographically disaggregated to include information regarding locations of children (e.g., at a populated place, sub-district, or district). Using geographic analysis to identify and monitor changes in locations where more or fewer children have been reached provides useful performance management information. Additionally, the data depicting where children have been reached can be compared with other geographically disaggregated indicator data, such as the prevalence of stunted children under five, to determine whether a higher number of children being reached coincides with improved child nutrition. In this example, a higher number of children reached by nutrition programs (left) appears to coincide with improved child nutrition (right).



Type III
Geographic Measurement: Required
Geographic Analysis: Valuable
Type III indicators require geographic measurement to collect the data and there is inherent value in geographic analysis of the data.

Example: Peru - Type III Indicator
The indicator "number of hectares under improved management practices" is used to monitor performance. The unit of measure for this indicator is hectares and is therefore inherently geographically disaggregated. The data can be used in a geographic analysis to identify areas where more or less agricultural area is under improved management and provide meaningful information for performance management (left). Additionally, these data can be aggregated to be compared with other performance or context indicators. For example, the data can be aggregated to the province scale (right) to compare it to other data at the province scale, such as improved child nutrition outcomes (e.g., "Outcome Indicator" maps above) to determine whether or not a larger amount of agricultural land area under improved management in a province coincides with improved child nutrition in the same province.



DETERMINING THE APPROPRIATE LEVEL OF DISAGGREGATION

When determining the appropriate level of geographic disaggregation for an indicator, consider the level of geographic detail needed to provide useful information and the flexibility to aggregate the data so that the data can be used in a range of analyses. The same guiding questions provided in the section on defining a minimum geographic scale will be helpful in this process.

DETERMINING DATA COLLECTION AND ANALYSIS REQUIREMENTS

The process of designing indicators for geographic disaggregation is a balance between identifying data that will provide meaningful performance management information and the resources that are required to collect and analyze the data. This section outlines requirements for a Mission and its partners to collect and analyze geographically disaggregated data.

Data Collection & Analysis Requirements
In general, the complexity of an indicator and how frequently the data are collected will determine the level of effort and expertise that are required to collect high quality data. For geographically disaggregated indicators, collecting data at a higher level of geographic detail increases the required level of expertise and technology. Considering the range of geographic scales at which data can be collected, and the varying

level of effort, expertise, and technology required at each scale is an important step in finding a balance between collecting useful information and resource availability.

Table 1: Summary of Requirements for Data Collection and Data Analysis by Geographic Scale

Scale	Description	Data Collection Requirements		Data Analysis Requirements		Level of Effort
		Skill	Technology	Skill	Technology	
Admin. Unit	A subnational, political boundary (e.g., region, district, etc.)	Low	MS Excel, etc.	Low	MS Excel, etc.	Low
Populated Place	The location of a populated place (e.g., a city or settlement)	Low	MS Excel, etc.; populated place location dataset	Low-Medium	MS Excel, etc.; GIS, or other mapping tool	Low
Point Location	A discrete point location (e.g., a health facility, a household, etc.)	Medium	GPS-enabled device	Medium-High	MS Excel, etc.; GIS or other mapping tool	Medium
Area/Line Features	A discrete line or area (e.g., an agricultural area, a road, etc.)	High	GPS-enabled device; Satellite Imagery; Manual Methods	Medium-High	GIS or other mapping tool	High

ASSESSING EXISTING EXPERTISE AND RESOURCES

Often, a Mission and its partners will have existing expertise and access to the resources required to collect and analyze geographically disaggregated data. Several guiding questions will assist in assessing existing expertise and resources:

- What level of expertise to analyze geographically disaggregated data already exists at the Mission? Does the Mission have a GIS Specialist? Which tools are required and are they already accessible? Which other applicable tools are available at USAID?
- What level of expertise to collect and analyze geographically disaggregated data already commonly exists among partners? Could methods be shared from partners with higher capacity to partners with lower capacity?
- Is geographically disaggregated data routinely collected by secondary data sources, such as the host-country government or other organizations? At which geographic scale is the data disaggregated? Could these data serve the Mission's performance management needs?
- Is geographically disaggregated data already collected by partners? At which geographic scale is the data collected even if the data are aggregated to fulfill performance indicator requirements that do not specify geographic disaggregation?

Geographic Data Analysis and Visualization

Geographic analysis and visualization add a new dimension to common performance monitoring methods. Identifying how trends, patterns, and relationships change over geographic space, in addition to time, increases the ability to understand and adapt to the complex environments in which development programming commonly occurs. Exploring data and analytical results in a visual form helps to identify and communicate conclusions in ways that are easily understood, shared, and acted upon.

ASKING AND ANSWERING GEOGRAPHIC QUESTIONS

Together, geographic analysis and visualization strengthen monitoring and adaptive management. Incorporating these methods into monitoring does not drastically alter the questions that one seeks to answer, but adds a geographic dimension to the questions that are asked and how they are answered.

For example, "To what extent is the project or activity meeting targets, and which contextual factors may be influencing performance?" remains a fundamental question, but by analyzing and visualizing geographically disaggregated data one can pose a geographically-explicit question, "in which locations is the project or activity under- or over-performing? Which contextual factors may be related to the variation of performance in these specific locations?" Using geographic analysis and visualization to answer these questions uncovers a new level of performance monitoring insight, and enhances the way one understands and responds to place-based factors affecting development programming.

Although geographic analysis and visualization is most commonly associated with maps, a range of other analyses and visualization methods can be used to analyze geographic data, including dot plots, bar charts, and line graphs, among others. Analyzing and visualizing data in these ways draws comparisons, identifies trends or patterns, and generates new questions or ideas about relationships in the data that can be further explored with additional methods, as needed. To review an example of how a project team might determine and what visual analysis is useful, and what data are needed to create the visual analysis, please see the Country Example: Analyzing Geographic Monitoring Data available in the Monitoring Toolkit. The Monitoring Toolkit also has a one-page Geographic Data Disaggregation Checklist to help managers and CORs/AORs/GATRs make decisions about disaggregating indicator data by geographic location.

Key Terms and Definitions

Administrative Units: The hierarchy of political divisions within a country is comprised of administrative units delineated by geographic boundaries. The highest level administrative unit is the country and is referred to as the administrative unit 0. The first level of subdivision within the country is referred to as the administrative unit 1. First level administrative units are subdivided by second administrative units and so on until the lowest level of subdivision is reached. Shorthand terms such as "Admin 1" or "Adm1" may be used to refer to administrative units and commonly appear in column headings or filenames for datasets. The most important feature in these terms is the number that denotes the level of subdivision. Additionally, the terms used to refer to the levels of administrative units may vary by country. For example, an administrative unit 1 in the United States is referred to as "state" whereas in Peru an administrative unit 1 is referred to as "region."

Administrative Unit Hierarchy - Peru Example

Administrative Level	Number of Units	Geographic Size
0 - Country	1	Largest
1 - Region	25	---
2 - Province	196	---
3 - District	1,838	Smallest

Modifiable Areal Unit Problem (MAUP): A challenge occurring during geographic analysis of aggregated data, where results differ when the same analysis but different aggregation schemes are applied to the same data. MAUP takes two forms: scale effect and zone effect. Scale effect creates different results when the same analysis is applied to the same data, but changes the scale of the aggregation units. For example, analysis using data aggregated by county will differ from analysis using data aggregated by census tract. Often this difference in results is valid: each analysis asks a different question because each evaluates the data from a different perspective (different scale). Zone effect is observed when the scale of analysis is fixed, but the shape of the aggregation units is changed. For example, analysis using data aggregated into one-mile grid cells will differ from analysis using one-mile hexagon cells. The zone effect is a problem because it is an analysis, at least in part, of the aggregation scheme rather than the data itself. (http://gispopsci.org/maup/)

Geographic Information Systems (GIS): A computer system designed to capture, store, manipulate, analyze, manage, and present all types of geographic data.

Global Positioning System (GPS): A system of radio-emitting and -receiving satellites used to determine a position on the earth. Orbiting satellites transmit signals that enable a GPS receiver anywhere on earth to calculate its own location through trilateration.

GPS-enabled Device: A GPS-enabled device refers to GPS units explicitly intended for the collection of geographic data or other devices, such as smart phones, that include GPS functionality. Today, a broad range of GPS-enabled devices exist to collect geographic location information yet there are differences in the level of geographic accuracy and precision among different devices. Therefore, it is important to first determine the data collection requirements and second select a device that fulfills the determined requirements.