

MEL PRACTITIONER GUIDE:

ALTERNATIVES TO SURVEY MEASUREMENT FOR ACTIVITY AND CONTEXT MONITORING: USE CASES FROM LATIN AMERICA AND THE CARIBBEAN CITIZEN SECURITY PROGRAMMING

Gregory Haugan, McKinzie Davis, Miguel Albornoz, & Alejandra Mijares

This publication is made possible by the United States Agency for International Development and the generous support of the American People. The contents are the responsibility of NORC at the University of Chicago on behalf of Development Professionals, Inc.–Making Cents International LLC through the Analytical Services IV IDIQ Task Order Latin America and Caribbean Learning and Rapid Response (LACLEARN). This material does not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGEMENTS

In addition to the LACLEARN and USAID teams who provided inputs and reviewed this guide, the authors would like to specially recognize the following people for their contributions:

Douglas Farah, NORC Consultant

Michael Weintraub, PhD, NORC Consultant

Luis Camacho, PhD, NORC

Stephanie Schmidt, Making Cents International

Todd Anderson, USAID

Mikell Brough-Stevenson, USAID

LACLEARN CONTACTS

Contracting Officer's Representative Cara Thanassi, cthanassi@usaid.gov

Task Order Manager Gabriela Leva, gleva@developmentpi.com

SUGGESTED CITATION

Hagan, G., Davis, M., Albornoz, M., Mijares, A. (2023). MEL Practitioner Guide: Alternatives to Survey Measurement for Activity and Context Monitoring: Use Cases from Latin America and the Caribbean Citizen Security Programming. USAID: Washington, DC

TABLE OF CONTENTS

ACRONYMS	ii
INTRODUCTION	iii
ADMINISTRATIVE DATA	i
DOCUMENT TROVES DATA.....	5
REMOTE SENSING DATA.....	9
SOCIAL MEDIA DATA	13
TRAINED OBSERVER DATA.....	17
CROWD-SOURCED DATA	21

ACRONYMS

AOR	Agreement officer representatives
API	Application programming interface
CEV	<i>Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición/Colombian Truth, Coexistence, and Non-Recurrence Commission</i>
CNN	Convolution neural network
COR	Contract officer representatives
CSO	Civil society organization
DRG	Democracy, human rights, and governance
GBV	Gender-based violence
GESI	Gender, equity, and social inclusion
GIS	Geographic information system
GPS	Global positioning system
INL	International Narcotics and Law Enforcement Agency
INPEC	<i>Instituto Nacional Penitenciario y Carcelario/National Penitentiary and Prison Institute</i>
IP	Implementing partner
ISIS	Islamic State of Iraq and Syria
LfP	Land for Prosperity
MEL	Monitoring, evaluation, and learning
NASA	National Aeronautics and Space Administration
NGO	Non-governmental organization
NSA	National statistical agency
PBJ	Place-based justice
PBS	Place-based security
PII	Personal identifiable information
UNODC	United Nations Office on Drugs and Crime
USAID	United States Agency for International Development
VIIRS	Visible Infrared Imaging Radiometer Suite

INTRODUCTION

PURPOSE

This guide provides practical and useful guidance for Mission Monitoring, Evaluation, and Learning (MEL) staff; Democracy, Human Rights, and Governance (DRG) technical teams; Contracting/Agreement Officers Representatives (CORs/AORs); and Implementing Partners (IPs) on how to track activity performance and gauge hard-to-measure DRG indicators using alternative data sources to surveys. While these data sources may be used as alternatives to surveys in some situations, they may also be used to supplement or complement survey data. For example, remote sensing data might provide community-level measures of highly sensitive indicators that cannot safely be asked in a survey, such as presence of illicit crops. On the other hand, remote sensing data could be paired with survey data to measure indicators that remote sensing cannot, such as average size of landholdings, trust in government, or prevalence of land-related conflict. The guide also presents example use cases from the citizen security sector and builds upon USAID’s *Crime and Violence Prevention Field Guide* and USAID’s *Youth Violence Prevention Indicators and Interventions Resource Guide*. In addition, the content of the guide complements one of USAID Learning Lab’s monitoring toolkit resources, *Data Collection Methods and Tools for Performance Monitoring*.

It is not the intent of this guide to suggest that any of these alternative data sources are preferable to survey data or that they *should* be used in the place of a well-designed, representative survey. Rather, the guide is intended to provide an overview of alternative data sources that may work in situations where it is already determined that survey measurement is infeasible (e.g., due to cost or security concerns) or where it is desirable to consider the potential of existing data sources before committing to burdensome primary data collection.

STRUCTURE

The guide covers six data types (click the red links to proceed directly to each, if preferred):



The guide examines each of these using the following structure:

Overview

A brief description of the data type and illustrative examples.

Citizen security case study

An example of how the data type has been used (in a USAID citizen security program, where possible).

Data quality considerations

An analysis of the advantages and limitations of the data type using USAID's five data quality standards: validity,¹ reliability,² precision,³ integrity,⁴ and timeliness⁵ (*Data Quality | USAID Learning Lab*, n.d.). Here, we also include an analysis of the advantages and limitations of the data type in relation to gender, equity, and social inclusion (GESI). This analysis is included under data quality considerations due to the effects incomplete or unconsidered GESI considerations may have on data quality, e.g., representativeness, validity, etc.

Citizen security indicators

A brief list of examples of common citizen security indicators that can be measured with the data type (inspired by USAID's *Crime and Violence Prevention Field Guide* and *Youth Violence Prevention Indicators and Interventions Resource Guide*) and details of their use from the literature.

Guidance

A short guide on what to consider for the data type when developing evaluation timelines, thinking about staffing considerations, obtaining the data/conducting fieldwork/gathering primary data, preparing the data for analysis, and analyzing the data.

¹ Validity – Data should clearly and adequately represent the intended result.

² Reliability – Data should reflect stable and consistent data collection processes and analysis methods over time.

³ Precision – Data have a sufficient level of detail to permit management decision-making: e.g., the margin of error is less than the anticipated change.

⁴ Integrity – Data collection should have safeguards to minimize the risk of transcription error or data manipulation.

⁵ Timeliness – Data should be available at a useful frequency, should be current, and should be timely enough to influence management decision-making.



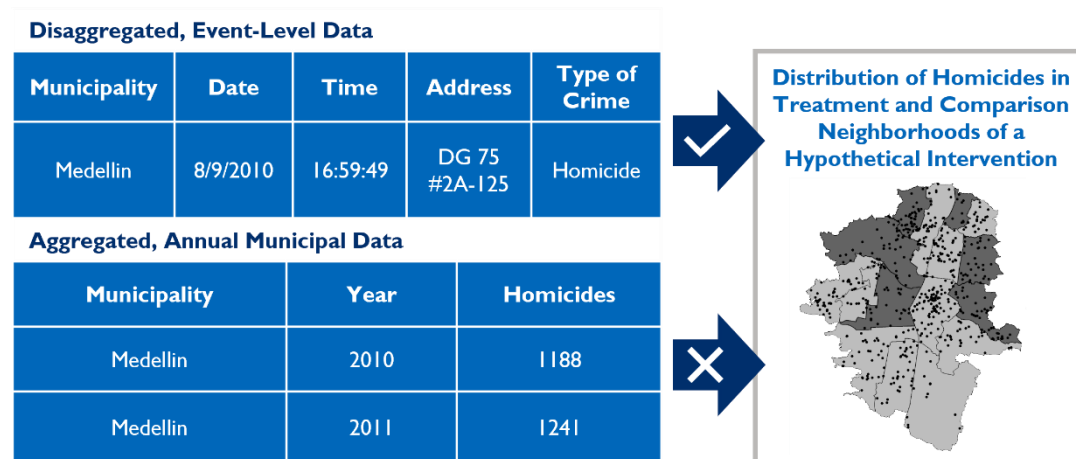
ADMINISTRATIVE DATA

OVERVIEW

DATA COLLECTED BY GOVERNMENT OR OTHER PUBLIC INSTITUTIONS VIA DAY-TO-DAY OPERATIONS.

These data are available through National Statistical Agencies (NSAs), relevant government agencies, or non-governmental entities and may be publicly available or require a public information request or special contacts to obtain.

Understanding levels of aggregation: Administrative data are often published in an aggregated form, whether geographically (neighborhoods aggregated to a city) or temporally (months aggregated to a year). This limits the level of analysis that can be conducted with the data. Disaggregated datasets may exist but are often more difficult to obtain.



THESE DATA ARE USEFUL WHEN...

Data for indicator are routinely collected (e.g., police reports, court records)



Indicators measured at the level of common administrative units (e.g., municipality)



Data are reported at a frequency that aligns with analytical needs



CITIZEN SECURITY CASE STUDY

Program: USAID Place-Based Justice (PBJ) and USAID/INL Place-Based Security (PBS) Strategies

Evaluation details: NORC at the University of Chicago and Mendez England and Associates (ME&A) (2019) used administrative data to evaluate the impacts of justice sector reforms in El Salvador on crime, citizens' perceptions of security, and caseloads and capacities of justice sector institutions.

Data: Annual, municipal-level administrative data from the prosecutor's office, public defender's office, and National Civil Police:

- Reported crimes, arrests, warrants, indictments, convictions, and time-to-case resolution for various crimes.
- Number of prosecutorial personnel, police personnel, and judicial personnel, by sex.


Program interventions targeted specific municipalities, making municipal-level data appropriate for the evaluation.

DATA QUALITY CONSIDERATIONS

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> Administrative data are often publicly accessible. This may mean the data have already been publicly vetted. 	<ul style="list-style-type: none"> Administrative data that rely on citizen reports are likely underreported.
RELIABILITY	<ul style="list-style-type: none"> Data from NSAs often have their methods and indicator definitions published, facilitating monitoring of the source data's reliability. 	<ul style="list-style-type: none"> Indicator definitions may change over time. These changes could have political motives. Protocols, standards, and definitions for data vary across NSAs. Weaknesses in institutional capabilities can lead to scarce and unreliable data.
TIMELINESS	<ul style="list-style-type: none"> Administrative data are often published regularly for long windows of time, facilitating longitudinal analysis. 	<ul style="list-style-type: none"> Although published consistently, some administrative data may not be published frequently (e.g., monthly, yearly) and may be available only after a time lag (e.g., 2020 data may become available late in 2021). Researchers are beholden to the frequency reported by the institution, which may not align with analytical needs.
INTEGRITY	<ul style="list-style-type: none"> Data can be quite reliable, if there is citizen oversight of administrative data collection and/or data are harmonized across administrative agencies (e.g., police, attorney general's office, local authorities). 	<ul style="list-style-type: none"> Weaknesses in institutional capabilities, particularly at a local level, may lead to data storage methods susceptible to unauthorized changes or transcription errors.
PRECISION	<ul style="list-style-type: none"> Granular data can provide a high degree of spatial and temporal precision, sometimes down to specific addresses and minute of occurrence. Governments may have data collection advantages, including access to specialized technology or training. 	<ul style="list-style-type: none"> Administrative data may be aggregated at levels of space and time that are too broad for certain projects. Although more precise and detailed information may exist, it can be difficult to obtain.
GESI	<ul style="list-style-type: none"> In some cases, administrative data may be analyzed by different gender, ethnic, or socioeconomic groups. 	<ul style="list-style-type: none"> Administrative capacity for data collection can lead to representation gaps in data (e.g., low capacity in rural areas). Disaggregation is limited to variables already in the dataset. Many variables of interest may not be included (e.g., sexual orientation). Marginalized populations may be less likely to use or have access to government services, or have situations (e.g., lack of permanent housing) that means administrative data collection misses them. They may be underrepresented in the data.

CITIZEN SECURITY INDICATORS

INDICATOR	EXAMPLE
<p>Crime rates</p>	<p>Jaitman and Ajzenman (2016) used geocoded 2006–2015 police administrative records from Brazil, Colombia, Mexico, Uruguay, and Venezuela to evaluate geographic clustering of crime.</p> <hr/> <p>Blair and Weintraub (2021) used time-stamped, geocoded administrative data (including data on robberies, homicides, illegal possession of firearms, and human rights abuses) in conjunction with survey data and GPS data to evaluate military policing interventions in Colombia.</p>
<p>Recidivism rates</p>	<p>Tobón (2020) used administrative records from Colombia’s National Prison Institute (INPEC) to gauge recidivism rates by observing a prisoner’s initial incarceration and subsequent instances of reincarceration.</p>

 State capacity in different municipalities/departments means the process for recording data may be different in different geographies. Some may not have resources to digitize data and instead do hand-recording. This results in additional work when trying to compile and collect data.

GUIDANCE

OBTAINING DATA

- Determine whether data are publicly available or require special permission to obtain. This is key in establishing a timeframe to obtain the data.
- Identify the format the data are in and whether this will be suitable for your project or activity. If the data are not suitable because they are aggregated at a level that is too high (e.g., aggregated at the municipal level, but the program is implemented at the neighborhood level), obtain the raw data that were used to produce the aggregate dataset. This will be more flexible in its usage but may be more difficult to obtain.
- While determining the format of the data, establish whether the data contain any personally identifiable information (PII). While aggregated data are unlikely to contain PII, disaggregated data might include PII. Datasets containing PII may require a data use agreement with the owner of the data and special precautions for storing and using the data.

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- Administrative data are obtained from external sources and do not require any fieldwork or primary data collection. This provides significant time and cost savings compared to the collection of survey data and removes concerns around safety of staff in the field.

PREPARING DATA FOR ANALYSIS

- The time and skills required to prepare administrative data for analysis can vary widely depending on the specific situation.

- Where aggregated data are suitable for the project’s monitoring purposes, the level of effort required to prepare the data can be approximately one or two days, even if multiple indicators and time periods are involved.
- If working from disaggregated, event-level data, data preparation can be more time consuming. The amount of time can vary significantly, but it is not uncommon for data preparation to take several weeks.
- Complication arises when the data arrive in an unexpected format. Instead of a single Excel file with rows for each observation and columns for each variable, the data might arrive in several hundred text files with nonstandard formatting to distinguish between observations and variables. Or instead of a formatted database, images of thousands of forms that need to be coded using text mining or machine learning techniques are provided (as described in Document Trovers chapter). It can take a great deal of creativity and effort to resolve these issues, and the time required will vary enormously from case to case.

DATA ANALYSIS

- To prepare and analyze administrative data, strong statistical programming skills in Stata or R are recommended.

REFERENCES

- Blair, R., & Weintraub, M. (2021). Little Evidence that Military Policing Reduces Crime or Improves Human Security. *Nature Human Behavior*. <https://doi.org/10.2139/ssrn.3925245>
- Jaitman, L., & Ajzenman, N. (2016). Crime concentration and hot spot dynamics in Latin America. In *IDB Working Paper Series, No. IDB-WP-699, Inter-American Development Bank (IDB)*. <https://publications.iadb.org/en/crime-concentration-and-hot-spot-dynamics-latin-america>
- ME&A & NORC. (2019). *Impact evaluation baseline report of USAID PBJ and USAID/INL PBS strategies in El Salvador*. USAID. https://pdf.usaid.gov/pdf_docs/PA00W6QB.pdf
- Tobón, S. (2020). Do better prisons reduce recidivism? Evidence from a prison construction program. *The Review of Economics and Statistics*, 1–47. https://doi.org/10.1162/rest_a_01007

RETURN TO INTRODUCTION



DOCUMENT TROVES DATA


OVERVIEW

DOCUMENT TROVES REFERS TO DATA IN THE FORM OF TEXT.



Textual data can derive from emails, document leaks (e.g., the Panama Papers), news stories, narrative administrative records, government reports, etc. Because these troves typically contain many documents, information of interest is extracted through computer programs in a process known as text mining.

Text mining: An application of machine learning. Tasks include categorization, sentiment analysis, summarization, and event coding.

➔ **Example of text mining:** Researchers generated a trove of news stories by scraping Google News for stories related to criminal groups in some country of interest and applied text mining to extract data on the criminal group(s) named, the municipality where they operate, and the date of the story.

 **National Police Make Drug Bust in Santa Maria**

Operations today in the northern Santa Maria province resulted in the capture of 15 tons of illicit drugs, and the arrest of 20 members of the *Clan del Bahia* mafia.

2:53 pm - August 8, 2022  



Date	Province	Criminal Group
8/8/22	Santa Maria	Clan del Bahia

CITIZEN SECURITY CASE STUDY

Program: Colombia's Truth, Coexistence, and Non-Recurrence Commission (CEV)

Evaluation details: Acosta et al. (2021) provide an example of using document troves for monitoring in the context of the CEV, which requested a performance evaluation through USAID/Colombia's MEL activity.

Data: Researchers conducted social listening on the CEV using Google News stories.

- Scraped Google News for relevant stories, searching specific terms, using the search term "Comisión de la Verdad," generating a trove of 1,760 relevant stories.
- Applied text mining algorithms to categorize these stories as positive, negative, or neutral while also identifying the date of the story and region of the publisher.

Data analysis: Researchers analyzed the percentage of stories that were positive, examined evolution of sentiment over time, identified words most associated with positive and negative stories, and the most important news sources in each region of the country. A key finding was that online news stories link the Commission to Special Jurisdictions for Peace when the subject matter is conflict, peace, and memory.

THESE DATA ARE USEFUL WHEN...

Troves provide a reasonably full or representative picture of the indicator of interest



Geographic precision is less important, and you are measuring indicators at the level of common administrative units



You are measuring indicators that are not typically reported in administrative data



DATA QUALITY CONSIDERATIONS

Given the variety of potential sources of document troves, all sources should be assessed for data quality on a case-by-case basis, as there is wide variation in advantages and limitations across each source.

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> In limited contexts, a trove of documents may provide a complete record. For example, court records may completely account for all court activity. 	<ul style="list-style-type: none"> All data quality considerations for document troves, including validity, depend on the source text. For example, if using document troves to identify police abuses, whether the data clearly represent the intended result may depend on whether the source text comes from internal police records or news stories, as well as the nature of those police records or quality of the news sources included.
RELIABILITY	<ul style="list-style-type: none"> Reliability considerations will vary across sources, and more reliable data sources would include standardized language and be produced by trained professionals, such as administrative documents. 	<ul style="list-style-type: none"> Text mining may not pick up subtleties in language, such as variations in language over time, geographies, or authors. Standards for what is documented may vary and be highly subjective. For instance, news may be more likely to document events in high-profile areas than in marginalized neighborhoods.
TIMELINESS	<ul style="list-style-type: none"> In some cases (e.g., scraping news stories from the internet), data may be available in real time, for a wide time range, and at high frequency. 	<ul style="list-style-type: none"> For text from official sources (e.g., government records), there can be long time lags before data are made available.
INTEGRITY	<ul style="list-style-type: none"> Researchers should consider author biases, accuracy of automated text classification programs, and odds of misreporting or transcription errors. Some sources may have built-in integrity and quality assurance processes. 	<ul style="list-style-type: none"> Some sources may have few integrity or quality assurance processes. For example, many newspapers have cut down on fact-checking mechanisms or may not have them in place at all.
PRECISION	<ul style="list-style-type: none"> Documents may contain precise levels of geographic information and be available at high frequency. Some document types may contain very detailed information. This lends itself to precision in detecting changes in the data. 	<ul style="list-style-type: none"> Some data sources, such as news stories, are inconsistent in how they report the indicator being measured. Researchers may need to coarsen the indicator to compensate. For example, stories may not always report the number of gang members, so an indicator may be coarsened to identify the binary presence of a gang in a geography.

- Document troves may provide insight into how institutions treat specific groups that are not found in other data. Emails or call transcripts may reveal how officials or institutions talk about or treat members of marginalized populations.

- Users must consider potential biases of the author (and publisher). This may influence what is reported and how.
- Users must consider potential blind spots or biases in the text mining algorithms. Is there nonstandard speech, slang, foreign languages, or poor spelling/grammar that may result in biased text mining results?

CITIZEN SECURITY INDICATORS

INDICATOR	EXAMPLE
Human rights violations	Osorio et al. (2019) used natural language processing to extract data from a published collection of human rights violations narratives and map the presence of armed actors during Colombia's civil war.
Legislative accessibility for voters	Correa et al. (2018) used text mining on the peace agreements of three Latin American countries in order to classify the difficulty (i.e., reading level) of the texts, as a measure of their accessibility to citizens who may be voting on their approval in popular referenda.
Discrimination in court sentencing	Pina-Sanchez et al. (2019) used text mining of court sentence records to analyze and compare the sentence lengths of Muslim-named offenders to non-Muslim-named offenders.
Domestic violence	Karystianis et al. (2022) used text mining of police reports from domestic violence events in Australia to analyze victim injuries, abuse types, and mental health details of suspects and victims.

GUIDANCE

OBTAINING DATA

- Determine where the data are stored, the format it is in, and how it can be obtained. These will be key parameters for estimating the level of effort required.
- If the data are obtained via a web search (e.g., online news stories), the COR should understand and agree with the search parameters, such as date range, search terms, etc.
- Obtaining primary documents from government agencies (e.g., police reports, court records, etc.), can be time and labor intensive. In some cases, government agencies may be unwilling or unable to provide the information.
- If the document trove does not come from a publicly available source (such as a trove of news articles), establish whether the data contain any PII. Datasets containing PII may require a data use agreement with the owner of the data and special precautions for storing and using the data.
- Carefully consider the quality of the source text. Does the source text come from news stories? If so, are the news sources included known to be reliable or might they include sources that contain deliberate misinformation or insufficient fact-checking? Does the source text come

from police records? If so, what is the nature of those records, and is there reason to believe certain events or details might be deliberately omitted or misrepresented?

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- Data from document troves are obtained from external sources and do not require any fieldwork or primary data collection. This provides cost savings compared to the collection of survey data and removes concerns around safety of staff in the field. Sampling methods are typically not required. However, while still less expensive than a traditional household survey, some of these savings may be offset by a much more intense data preparation phase.

PREPARING DATA FOR ANALYSIS

- Converting documents into analyzable data requires text mining or machine learning techniques, which can be time and labor intensive and will vary in intensity. These tasks typically require high-level technical skills.

DATA ANALYSIS

- To prepare and analyze document troves data, advanced skills in R or Python are needed. Certain applications may require skills in other programming languages.

REFERENCES

- Acosta, P., Llanes, L., Velásquez, M., Rudling, A., & Agudelo, J. (2021). *Truth, Coexistence, and Non-Recurrence Commission performance evaluation: Final report*. USAID. https://pdf.usaid.gov/pdf_docs/PA00XMIT.pdf
- Correa, J. C., García-Chitiva, M. del P., & García-Vargas, G. R. (2018). A text mining approach to the text difficulty of Latin American Peace Agreement. *Revista Latinoamericana de Psicología*, 50(1), 61–70.
- Karystianis, G., Adily, A., Schofield, P. W., Wand, H., Lukmanjaya, W., Buchan, I., Nenadic, G., & Butler, T. (2022). Surveillance of domestic violence using text mining outputs from Australian police records. *Frontiers in Psychiatry*, 12, 787792. <https://doi.org/10.3389/fpsy.2021.787792>
- Osorio, J., Mohamed, M., Pavon, V., & Brewer-Osorio, S. (2019). Mapping violent presence of armed actors in Colombia. *Advances in Cartography and GIScience of the ICA*, 1, 1–9. <https://doi.org/10.5194/ica-adv-1-16-2019>
- Pina-Sánchez, J., Roberts, J. V., & Sferopoulos, D. (2019). Does the Crown Court discriminate against Muslim-named offenders? A novel investigation based on text mining techniques. *The British Journal of Criminology*, 59(3), 718–736. <https://doi.org/10.1093/bjc/azy062>

RETURN TO INTRODUCTION



REMOTE SENSING DATA

OVERVIEW

REMOTE SENSING DATA TYPICALLY ORIGINATES FROM SATELLITE SCANS OF THE EARTH.

The information from these scans is processed into a dataset that divides space into pixels, assigning a value to each. Data are then stored in “raster data” file format.

➔ **Examples of raster data:** Measures of nighttime light intensity, annual precipitation, or a categorical variable for the type of land use (e.g., farmland).

GPS tracking data: A related but distinct data type, GPS tracking data derive from cellphones or GPS tracking devices. Data may be stored in raster or standard dataset format.

➔ **Examples of GPS tracking data:** Data from police equipment that record the precise location of patrols at regular time intervals.



Using surveys to measure illicit crop cultivation faces the obvious problem of respondents failing to honestly report their activities, due to fear of legal consequences, retaliation by illegal armed groups, or wanting to give socially desirable answers. Remote sensing data, by contrast, do not face these problems. Remote sensing also has the added advantage of avoiding security risks to survey enumerators.

CITIZEN SECURITY CASE STUDY

Program: Land for Prosperity (LfP) Activity in Colombia

Evaluation Details: Haugan et al. (2022) conducted the baseline for an impact evaluation of LfP. They sought to measure the impacts of LfP on agricultural productivity, women’s empowerment, conflict, and coca production.

Data: The study used United Nations Office on Drugs and Crime (UNODC) raster data from high-resolution imagery to measure area under illicit crop cultivation in communities in program municipalities:

- Program municipalities were matched to comparison communities with similar levels of illicit crop cultivation.
- At endline, the researchers will use the UNODC data to measure the impact of the program on illicit crop cultivation.

Highly sensitive, geographically precise data, such as UNODC illicit crop monitoring data, usually require special permissions to obtain.



THESE DATA ARE USEFUL WHEN...

Geographically precise information is available about where a program is operating or expects to have an impact



Measuring indicators that can easily be observed from space or aerial photography



Tracking the movement of individuals



Data are needed for units of space not typically reported in administrative data



DATA QUALITY CONSIDERATIONS

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> • Often publicly accessible, which may mean the data have already been through a public validation process. • Provides objective measures and generally does not face validity concerns of survey data, such as social desirability bias or recall problems. • Data usually completely cover the study area, avoiding challenges with obtaining a representative sample. 	<ul style="list-style-type: none"> • In rare cases where researchers produce new datasets by creating new algorithms to analyze satellite imagery, extra care is needed to ensure data fall within expected ranges and do not systematically over- or under-count. • Remote sensing data are available for relatively few indicators of interest to citizen security studies. In some contexts, this might lead to using the available data as “proxies” in ways that may lack credibility.
RELIABILITY	<ul style="list-style-type: none"> • Public datasets are often well documented and transparent in their methodology. • In most cases, remote sensing datasets use consistent methods over time, providing reliable time series data. 	<ul style="list-style-type: none"> • Researchers must review documentation to ensure methods are consistent across time. Improvements in technology or changes to classification algorithms sometimes result in changes in methods over time.
TIMELINESS	<ul style="list-style-type: none"> • Many raster datasets are available at regular, high-frequency intervals, given that they are based on satellite images taken near-continuously. 	<ul style="list-style-type: none"> • Some data may not be available at high-frequency intervals or made public in a timely manner. Users will need to ensure data are available according to the needs of the activity or project.
INTEGRITY	<ul style="list-style-type: none"> • Raster data from remote sensing are classified using computer algorithms, removing some of the biases present in other data collection methods. • Data on movement are collected through phones or GPS tracking devices, which automatically capture location at set intervals and send the data to a central server. 	<ul style="list-style-type: none"> • Machine learning algorithms for converting satellite imagery into raster data may be less accurate in locations with perpetually dense cloud cover. In extreme cases, this may result in missing or inaccurate data. Researchers should see dataset documentation for more information.
PRECISION	<ul style="list-style-type: none"> • Spatial data can be highly precise. Many remote sensing raster datasets are available at fine levels of spatial detail, such as 10- or 100-meter grid squares. Even the coarsest raster datasets are often more spatially precise than a typical municipal-level dataset. 	<ul style="list-style-type: none"> • Satellite imagery converted to raster format generally relies on machine learning classification algorithms. These algorithms are not always 100% accurate. Users should understand how the data are generated and, in some cases, review existing literature to assess the accuracy.
GESI	<ul style="list-style-type: none"> • Remote sensing datasets provide a spatial census and may avoid challenges such as underreporting, sample representativity, and data manipulation. 	<ul style="list-style-type: none"> • It is rare to be able to break down remote sensing data for reporting by subpopulations. While different gender and ethnic groups should be represented in the data, they usually cannot be distinguished.

CITIZEN SECURITY INDICATORS

INDICATOR	EXAMPLE
Perception of community safety	Bullock and Pellegrino (2021) used a mobility index produced by data firm GRANDATA. This dataset originated from cell phone GPS tracking data and shows how decreased mobility correlates with crime levels in Brazil.
Illicit crop cultivation	Sviatschi (2022) used UNODC data from annual satellite images, which UNODC verifies by aerial flyovers, to identify the density of coca production in Peruvian districts.
Extractive industries in conflict-affected areas	Do et al. (2018) used satellite imagery to construct a census of oil production in areas controlled by the terrorist group ISIS. The researchers use this to measure variation in oil production over space and time in areas under ISIS control to provide insight into how the group used oil as a revenue source.

GUIDANCE

OBTAINING DATA

- Carefully assess whether the proposed data are a valid measure of the indicator being monitored. Remote sensing data are available for relatively few indicators of interest to citizen security monitoring.
- Remote sensing data are often freely and publicly available and can be obtained relatively fast with exceptions:
 - Special permission may be needed to obtain sensitive datasets. This can lead to delays in obtaining data.
 - Analyzing remote sensing data usually requires shapefiles (computer files that contain the borders for spatial units of analysis). Obtaining shapefiles for common administrative units (e.g., municipalities) is generally easy and requires little to no time. However, shapefiles for less common units (e.g., neighborhoods, census blocks) can be difficult to obtain and may require special permission.
- Some examples of free, publicly available raster datasets include: VIIRS nighttime lights, Hansen et al. Global Forest Change, WorldPop.org population estimates, NASA Earth Data Land Use Categories.
- If the data involve GPS tracking of individuals, determine if the dataset contains PII. For remote sensing data, the most commonly used datasets are publicly available and would not be considered to include PII; however, in uncommon situations where the data are not publicly available and considered highly sensitive, the data could contain PII (e.g., raster data on illicit crops at the 100m resolution level could be used to identify individuals involved in illicit coca cultivation). Data containing PII may require a data use agreement with the owner of the data, and special precautions for storing and using the data.

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- In nearly all cases, remote sensing and GPS tracking data are obtained from external sources and do not require any fieldwork or primary data collection. This provides time and cost savings compared to the collection of survey data and removes concerns around safety of staff in the field. Further, sampling methods are typically not required, providing further cost and time savings.

PREPARING DATA FOR ANALYSIS

- The time and skills required to prepare spatial data for analysis can vary widely depending on the situation.
 - In straightforward applications, where the team has obtained remote sensing data in raster format, is conducting analysis for a single point in time, and has been provided with shapefiles for the spatial units that indicators will be measured at, preparation of the data might take a day or two of effort. The time required should increase only modestly for conducting measurements of multiple outcome indicators and for multiple periods of time.
 - A common complication arises when no shapefiles exist or can be obtained for the unit of spatial analysis. These situations can require substantial effort to resolve, and it can be difficult to predict how long it will take. In general, however, this usually does not require more effort than the data cleaning and preparation stage for a two- to three-hour household survey.
 - In rare cases, raster data may not be available, and instead raw satellite images are used. This can require machine learning techniques to convert the images into an analyzable dataset and involves considerable time and effort and highly specialized skills.

DATA ANALYSIS

- To analyze remote sensing and other spatial data types, contractors will need strong GIS skills in a platform such as R, Python, ArcGIS, or QGIS. In some cases, skills in multiple platforms (e.g., R or Python and ArcGIS) may be needed.

REFERENCE

- Bullock, J., & Pellegrino, A. P. (2021). How do Covid-19 stay-at-home restrictions affect crime? Evidence from Rio de Janeiro, Brazil. *Economia*, 22(3), 147–163. <https://doi.org/10.1016/j.econ.2021.11.002>
- Do, Q.-T., Shapiro, J. N., Elvidge, C. D., Abdel-Jelil, M., Ahn, D. P., Baugh, K., Hansen-Lewis, J., Zhizhin, M., & Bazilian, M. D. (2018). Terrorism, geopolitics, and oil security: Using remote sensing to estimate oil production of the Islamic State. *Energy Research & Social Science*, 44, 411–418. <https://doi.org/10.1016/j.erss.2018.03.013>
- Haugan, G., Protik, A., & Persha, L. (2022). *Evaluation of the Land for Prosperity (LfP) activity in Colombia: Baseline Report*. USAID Communications, Evidence and Learning (CEL) Project. https://land-links.org/wp-content/uploads/2022/07/NORC_CEL_LandForProsperityImpactEvaluation_BaselineReport.pdf
- Sviatschi, M. M. (2022). Making a NARCO: Childhood exposure to illegal labor markets and criminal life paths. *Econometrica*, 90(4), 1835–1878. <https://doi.org/10.3982/ECTA17082>

RETURN TO INTRODUCTION



SOCIAL MEDIA DATA

OVERVIEW

SOCIAL MEDIA DATA ARE MESSAGES, IMAGES, OR VIDEOS POSTED TO ANY SOCIAL MEDIA PLATFORM.

This type of data often requires an application programming interface (API) or permission from the social media company to obtain the data. Social media data will also often require text mining or machine learning to extract information from the posts and convert the information into variables in a dataset.

Machine Learning: Typical tasks include categorization, sentiment analysis, and event coding.

- ➔ **Example of machine learning for classifying images:** Researchers with a trove of Twitter posts apply an algorithm to identify posts with images of protests. The algorithm reviews each image to determine whether it shows state violence, protestor violence, and percentage of protestors by sex.



Date	6/21/21
City	San Cristobal
State Violence	No
Male	0.6

THESE DATA ARE USEFUL WHEN...

Measuring indicators related to public sentiment or perceptions



The representativeness of the data with respect to the general population is not a key concern



Using the data as a form of crowd-sourced data, and no crowd-sourcing mechanism for collecting the data currently exists or can be developed



Measuring movement of social media users



CITIZEN SECURITY CASE STUDY

Program: ResiliencyCambodia

Evaluation details: Starosta et al. (2020) provide an example of using social media for monitoring in the context of USAID programming in Cambodia.

Data: Researchers conducted social listening on Cambodian civil society organizations (CSOs), scraping posts from organizations' Facebook pages.

Data analysis: The researchers applied text mining algorithms to categorize the content of posts and record the number of likes, comments, and shares. The Facebook activity of CSOs was analyzed by CSO activity sector. Researchers plan to repeat this data collection at endline to examine how the ResiliencyCambodia program improved organizational capacity of these organizations and expanded their networks and reach.

DATA QUALITY CONSIDERATIONS

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> • Social media has the capacity to cover a large variety of indicators of interest for citizen security programming. • Useful for measuring indicators related to thought leaders or other influencers that shape public opinion. 	<ul style="list-style-type: none"> • When producing new datasets (e.g., by creating new algorithms to analyze social media data), care must be taken to avoid systematically over- or under-counting. • Social media users and their posts do not constitute a representative sample.
RELIABILITY	<ul style="list-style-type: none"> • There are no obvious advantages to social media data with respect to reliability. 	<ul style="list-style-type: none"> • What is posted to social media can be highly subjective. How people define and perceive specific situations mentioned in a post (e.g., abuse, gang activity, etc.) will vary across geographies and time.
TIMELINESS	<ul style="list-style-type: none"> • Social media data provide real-time, high-frequency data. 	<ul style="list-style-type: none"> • There are no obvious limitations for social media data regarding timeliness.
INTEGRITY	<ul style="list-style-type: none"> • There are no obvious advantages to social media data with respect to integrity. 	<ul style="list-style-type: none"> • It may be difficult or impossible to distinguish between real and false posts. Some countries operate troll farms on behalf of authoritarian-leaning regimes that may skew results. • Governments may suppress what is posted or visible, impose internet or social media blackouts, or censor certain types of posts. The political and/or social environment may also lead to self-censorship.
PRECISION	<ul style="list-style-type: none"> • Social media data are often geographically precise, are high frequency, and contain many observations. Some posts may contain very detailed information. This lends itself to precision in detecting changes in the data. 	<ul style="list-style-type: none"> • Because there is no consistency for how data are reported, details may be lost or not presented consistently.

- For some marginalized groups, social media may provide a unique space where they feel safe.

- Issues around representativeness of social media data may be particularly acute for highly marginalized groups with little to no internet access, or for groups who cannot access or use social media.
- Other marginalized groups may be underrepresented because they fear posting without negative repercussions.

CITIZEN SECURITY INDICATORS

INDICATOR	EXAMPLE
Gang structure and operations	Wijeratne (2015) manually collected tweets and used spatio-temporal thematic analysis, people-content-network analysis, and emotion-sentiment analysis to analyze the structure, function, and operation of gangs in Chicago.
Protest size and violence	Using a convolution neural network (CNN), Steinert-Threlkeld and Joo (2022) identified protest pictures from millions of geolocated Tweets and generated estimates of protest size, demographic characteristics of protesters, and type and severity of violence.
Public sentiment	Acosta et al. (2021) conducted social listening on the CEV using Twitter. Applying text mining to categorize tweets as positive, negative, or neutral in sentiment, researchers measured the evolution of sentiment over time.



EMERGING DATA SOURCES AND ANALYTICAL METHODS

The field for using social media data for monitoring is ever-expanding. New data sources – like Meta’s Data for Good platform – are growing, and researchers continue to pioneer new techniques to analyze social media data. While current literature has few examples of these methods being used for monitoring purposes, this is an area with new possibilities.

GUIDANCE

OBTAINING DATA

- Identify the social media platform to be used and the process for accessing data from that platform. Often, social media platforms will charge to access user data, though the associated costs are generally well below the cost for a traditional household survey.
- Consider the geographic area, time period, and search terms that will be used for scraping posts.
- The process for obtaining data often requires use of an API and moderately specialized skills, though usually does not require much time. Determining permissions, setting up an API, establishing the terms of the search, and running the program for scraping the data likely requires 5 to 10 days of effort for an experienced researcher, which could be spread across multiple weeks.

- Determine whether the dataset contains PII related to the social media account owner. Datasets containing PII may require a data use agreement with the owner of the data (i.e., the social media company) and special precautions for storing and using the data.

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- Social media data are obtained from external sources and do not require any fieldwork or sampling methods. This provides additional time and cost savings compared to the collection of survey data and removes concerns around safety of staff in the field. However, some of these savings may be offset by more intense data preparation.

PREPARING DATA FOR ANALYSIS

- Converting social media posts into analyzable data requires text mining and machine learning techniques, which can be time and labor intensive and will vary in intensity depending on the specific situation. These tasks require high-level technical skills.

DATA ANALYSIS

- To prepare and analyze social media data, advanced skills in R or Python are recommended. Certain applications may require skills in other programming languages.

REFERENCES

- Acosta, P., Llanes, L., Velásquez, M., Rudling, A., & Agudelo, J. (2021). *Truth, Coexistence, and Non-Recurrence Commission Performance Evaluation: Final Report*. USAID.
https://pdf.usaid.gov/pdf_docs/PA00XMIT.pdf
- Starosta, A., Springman, J., Hatano, R., Wibbel, E., & Right, L. (2020). *Resiliency Cambodia Baseline Findings Report: Baseline Findings from a Civil Society RCT in Cambodia*. USAID.
https://pdf.usaid.gov/pdf_docs/PA00X47C.pdf
- Steinert-Threlkeld, Z., & Joo, J. (2022). MMCHIVED: Multimodal Chile and Venezuela Protest Event Data. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1), 1332-41.,
<https://ojs.aaai.org/index.php/ICWSM/article/view/19385>
- Wijeratne, S. (2015). *Analyzing the Social Media Footprint of Street Gangs*. 2015 IEEE International Conference on Intelligence and Security Informatics (ISI), Baltimore, MD.
<https://doi.org/10.13140/RG.2.1.2878.2244>.

RETURN TO INTRODUCTION



TRAINED OBSERVER DATA

OVERVIEW

TRAINED OBSERVER DATA ARE COLLECTED VIA OBSERVATIONS BY TRAINED INDIVIDUALS.

Training focuses on providing people with the knowledge and tools needed to generate standardized ratings, both across observers and across time. Observations may focus on the presence or absence of specific characteristics of interest, rating perceptions of certain conditions, or frequency counts.

- ➔ **Example:** Classroom observations present a common example outside the citizen security area, where observations may determine the presence or absence of certain teaching practices over the course of a lesson, provide a quality rating for the state of the school's physical infrastructure, and/or count the number of students who went to the chalkboard to solve a problem during the lesson. In each case, the key point is that training on the use of a well-designed instrument should result in ratings that are approximately equal when different observers at different points in time observe similar conditions.

CITIZEN SECURITY CASE STUDY

Program: Greater Transparency and Government Responsibility Program

Evaluation details: Civil society organization representatives and municipal employees used trained observer ratings to monitor solid waste collection and road maintenance in cities in Honduras (Mark, 2008).

Data: Observers graded street conditions and street quality.

- Each street's ranking was determined by the average rating from nine observers.
- Four grades: "good condition," "regular condition," "bad," and "worst."

Data use: To facilitate a smooth transition during local elections, all mayoral candidates and the incumbent mayor agreed on joint goals, monitored by citizens, to ensure that municipal services continued during the transition period. Representatives from civil society organizations were trained to rate solid waste collection and road maintenance. Ratings were scheduled to occur before elections and six months after elections and served to identify locations for intervention by the city council.

THESE DATA ARE USEFUL WHEN...

Measuring an indicator that can be observed objectively



Repeated measures of an indicator are needed over time, at frequencies that may not be found in administrative data



Data for precise geographies or units are needed that may not be found in administrative data



Researchers generally set up primary data collection activities for trained observer ratings, meaning that the data collection's geographic level (e.g., neighborhood, street block, etc.) can be customized to suit the needs of the evaluation or research. This can be a solution for contexts where the study requires data for a specific geographic level that is not reported in administrative data.

DATA QUALITY CONSIDERATIONS

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> Trained observer ratings may avoid validity limitations that can occur in survey data, including problems with recall and social desirability. Raters could access dangerous areas for short, discreet observations, particularly when they are recruited directly from the location being observed. 	<ul style="list-style-type: none"> There may be limitations in what can be rated objectively. Perceptions of security in a neighborhood, for example, are unlikely to be objective, even if raters are trained. This may require the use of proxies (e.g., presence of graffiti, individuals who are visibly drunk or on drugs, etc.).
RELIABILITY	<ul style="list-style-type: none"> Carefully calibrated raters with carefully designed tools can produce stable and consistent data over time. Ratings can be conducted at greater frequency than other data collection methods (e.g., household surveys). 	<ul style="list-style-type: none"> Rater calibration and proper instrument design is the result of training and instrument validation but cannot be guaranteed a priori. Therefore, reliability of trained observer ratings is less assured than a data collection method like remote sensing. Risk of violence in certain locations may mean observers are unable to spend sufficient time in certain locations, causing variation in reliability.
TIMELINESS	<ul style="list-style-type: none"> Trained observer ratings have the ability to be collected at a high frequency. Trained observer ratings are available in near-real time. 	<ul style="list-style-type: none"> There are no obvious limitations with respect to timeliness.
INTEGRITY	<ul style="list-style-type: none"> Raters from a program’s target communities may have incentives not to over- or underrate their observations (overrating might provide evidence for continuing costly but ineffective programs, while underrating might provide evidence for discontinuing effective programs). The integrity of collected data will depend on the procedures established for the specific data collection effort. 	<ul style="list-style-type: none"> Integrity of trained observer ratings is comparable to household survey data collection. Well-designed data quality checks can catch improper manipulation, transcription errors, or falsification, but the integrity of the data is not guaranteed a priori.
PRECISION	<ul style="list-style-type: none"> Researchers and evaluators can design the methodology to conduct observations at geographic and frequency levels with the precision required for their specific evaluation. 	<ul style="list-style-type: none"> Researchers and evaluators may coarsen their instruments (i.e., make them less precise) to improve objectivity. Rating the level of a graffiti problem on the block could be less objective but more precise. While simply observing presence of graffiti is more objective but less precise.

GESI	<ul style="list-style-type: none"> The key assumption of this method is that any rater, with proper instrument design and training, will rate an observation the same way, regardless of their individual traits (e.g., race, gender, wealth). 	<ul style="list-style-type: none"> Without careful design and training, the key assumption may not hold. Ratings may then vary by individual traits, introducing biases into the ratings.
-------------	---	--

CITIZEN SECURITY INDICATORS

The academic literature using trained observer ratings to measure indicators of citizen security is limited. However, the existing literature does suggest trained observer ratings could be a useful data collection methodology for citizen security indicators.

INDICATOR	EXAMPLE
Crime prevalence, perceived risk of crime	Blattman et al. (2021) randomly assigned 1,919 street segments in Bogota to treatment and control groups, with treatment streets receiving doubled police patrols and greater municipal services. Enumerators sent to each street took a daytime photo and rated the presence of graffiti, garbage, visibly broken streetlights, and run-down buildings on a scale of 0–5.
Local government services	In Albania, Mark (2014) presented several examples of using trained observer ratings to rate the condition of schools, street cleanliness, and road conditions. They showed experiences where raters were city staff, teachers, parents, students, NGO staff, and citizens.
Gang presence and activities	In San Pedro Sula, Honduras, Farah and Babineau (2018) conducted observations of gang-controlled neighborhoods. Observations marked the presence of gang “rapid judicial sentencing” to resolve neighborhood conflicts, small-scale social programs run by gangs, and other neighborhood services.

GUIDANCE

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- In nearly all cases, trained observer ratings require fieldwork and primary data collection. The cost of implementing this approach will depend on the design of the fieldwork. In cases where observers are volunteers drawn from the units they are observing (e.g., residents observing their communities), there could be many cost savings (e.g., enumerator wages, lodging and meals in the field).
- CORs should keep in mind that the key assumption of this data collection modality is that any observer will give the same rating. This is important to consider when:
 - Reviewing observation instruments:* Are all items in the instrument possible to rate objectively? What are the cognitive requirements for the observation, and is it reasonable to expect raters to pay close attention to the number of indicators being observed over the anticipated time the observation will require?
 - Considering plans for training:* Will training be in-person or remote? How will that impact training quality? Is the duration of training adequate for the amount of material presented? How will trainers calibrate observers’ ratings to ensure observations have been effectively standardized? What actions will be taken if results from calibration are below standard?

- While it may be easier for trained observers to access dangerous neighborhoods or conflict hotspots than a team of survey enumerators, it is critical to prioritize *safety and ethics*:
 - What is the level of risk for observers? How can observers be kept safe?
 - If data collection identifies incursions by gangs or illegal armed groups, human rights abuses, or other serious violations, is there an obligation to report this to authorities, and what are the potential implications for affected communities of doing so (e.g., could reporting to authorities put the communities at risk)?

PREPARING DATA FOR ANALYSIS

- Preparing data from trained observer ratings requires a similar level of skill as the preparation of household survey data. However, observation instruments tend to be shorter than household survey instruments. Thus, the amount of time and effort required to prepare trained observer rating data is commonly much less than for a household survey, though the trade-off is that this is because less information is usually collected.

DATA ANALYSIS

- To prepare and analyze data from trained observer ratings, strong skills in a statistical program, such as Stata or R, is recommended.

REFERENCES

- Blattman, C., Green, D. P., Ortega, D., & Tobón, S. (2021). Place-based interventions at scale: The direct and spillover effects of policing and city services on crime. *Journal of the European Economic Association*, 19(4), 2022–2051. <https://doi.org/10.1093/jeea/jvab002>
- Farah, D., & Babineau, K. (2018). *The rapid evolution of the MS 13 in El Salvador and Honduras from gang to tier-one threat to Central America and U.S. security interests* [Perry Center Occasional Paper]. National Defense University. https://www.ibiconsultants.net/_upload/mediaandpublications/document/the-rapid-evolution-of-the-ms-13-march-23-2018-final.pdf
- Mark, K. (2008). Experience with Trained Observers in Transition and Developing Countries: Citizen Engagement in Monitoring Results. In Patria de Lancer Julnes, Francis Stokes Berry, Maria P. Aristigueta, & K. Yang (Eds.), *International Handbook of Practice-Based Performance Management* (pp. 248–249). Sage Publications.
- Mark, K. (2014). Performance Management Can Improve Local Services in Developing Countries: The Service Improvement Action Plan. *The World Bank Poverty Reduction and Economic Management Network Notes*, 29(84371).

RETURN TO INTRODUCTION



CROWD-SOURCED DATA

OVERVIEW

CROWD-SOURCED DATA ARE CREATED BY ALLOWING THE GENERAL PUBLIC TO MAKE REPORTS THAT CONTRIBUTE TO A DATABASE.

Methods include hotlines and reports made through mobile apps or online platforms.

Crowd-Sourced Data Hypothetical Examples



A well-known domestic violence hotline fields calls from victims of domestic abuse, providing them with psychological support and referrals to legal resources. Each call generates a report that feeds into a database that can be analyzed by evaluators.



A municipal government creates and popularizes an app to allow businesses to anonymously report extortion attempts. Each report feeds into a database that can be used by evaluators.

CITIZEN SECURITY CASE STUDY

Note: There are limited examples of USAID programs using crowd-sourced data for monitoring activities. The case study below examines a study conducted by academic researchers using crowd-sourced data resources in Latin America.

Research question: How does territorial control by organized crime groups affect the enforcement of public policies?

Research details: Bruce et al. (2021) used a crowd-sourced dataset from *Disque Denúncia* to estimate the impact of territorial control by drug trafficking organizations and paramilitary groups on the enforcement of the government's COVID-19 social-distancing policies in Rio de Janeiro.

Data: To determine where each criminal group held territorial control, researchers used a dataset that combined reports from *Disque Denúncia* and neighborhood maps from Pista News.

- *Disque Denúncia:* A well-known crime reporting hotline in Brazil, it served as an existing crowd-sourcing data collection mechanism for this project. Researchers combed the dataset using text-mining techniques to analyze over 38,000 transcribed calls to the hotline. The analysis identified neighborhoods with presence of armed groups and classified them according to dominance.
- The data sources were used to generate a map of territorial control by organized criminal groups in Rio de Janeiro.

THESE DATA ARE USEFUL WHEN...

The crowd-sourcing mechanism provides a reasonably full picture of the indicator to be measured



Repeated measures of an indicator are needed over time, at frequencies that may not be found in administrative data



Data for precise geographies are needed, and administrative data are not available



DATA QUALITY CONSIDERATIONS

	ADVANTAGES	LIMITATIONS
VALIDITY	<ul style="list-style-type: none"> Potential for covering a large variety of indicators of use to citizen security. 	<ul style="list-style-type: none"> Those who report do not constitute a representative sample. It may be difficult to identify when multiple reports refer to the same incident. Likelihood of reporting may be influenced by geography or other factors, so systematic over- or under-reporting may occur.
RELIABILITY	<ul style="list-style-type: none"> There are no obvious advantages, but a carefully designed instrument may help limit reliability concerns. 	<ul style="list-style-type: none"> What is reported can be highly subjective. There are limited tools for standardizing what is reported.
TIMELINESS	<ul style="list-style-type: none"> Crowd-sourced data produce high-frequency data. In some contexts, the data may be available in real time. 	<ul style="list-style-type: none"> Data being accessed from existing crowd-sourcing platforms run by external sources may only be available after a time lag based on data sharing agreements.
INTEGRITY	<ul style="list-style-type: none"> Integrity of crowd-sourcing platforms may vary considerably. 	<ul style="list-style-type: none"> These data are vulnerable to spoilers. It may be difficult or impossible to distinguish between real, false, or exaggerated reports.
PRECISION	<ul style="list-style-type: none"> Crowd-sourced data can be geographically precise and provide high-frequency data. Some reports may contain very detailed information, especially with a well-designed instrument. 	<ul style="list-style-type: none"> Some crowd-sourcing mechanisms may be inconsistent in how they report data, and granularity may be lost.
GESI	<ul style="list-style-type: none"> In some contexts, marginalized populations may feel more comfortable reporting via an anonymous crowd-sourcing platform than making a police report or taking an in-person survey. 	<ul style="list-style-type: none"> GESI considerations depend on the technology required for reporting, how well known the crowd-sourcing platform is, effort required to report, and groups' perceptions of the risks of reporting and who manages the platform.

CITIZEN SECURITY INDICATORS

INDICATOR	EXAMPLE
Calls to domestic violence hotlines	Perez-Vincent and Carreras (2022) used data from domestic violence hotlines (in addition to administrative records from police) in six Latin American countries to analyze the frequency of domestic violence rates at the beginning of the COVID-19 pandemic.
Gun violence among drug gangs	Monteiro and Rocha (2017) used data from the <i>Disque Denúncia</i> in Brazil, combined with education administrative data and geospatial data, to examine the effects of gang violence on educational outcomes.
Reports of gender-based violence (GBV)	Reactor, an independent research organization in Skopje, North Macedonia, conducted a scoping study regarding GBV in Skopje's public spaces (<i>Scoping Study on Gender Based Violence and Discrimination Against Women and Girls in Urban Public Spaces of the City of Skopje, 2020</i>). Their methodology included crowd-sourced data from the platform Ushahidi.

GUIDANCE

OBTAINING DATA

- Determine whether the data have been used in research or evaluations. This may provide insight of how feasible the data are to obtain. The amount of time it may take to obtain the data will vary substantially from source to source, as each source will have its own rules and processes for sharing data.
- If data from an existing crowd-sourcing platform can be obtained, there will likely be large cost and time savings over a traditional household survey. However, identifying an existing platform may be difficult – often they simply don't exist.

FIELDWORK AND PRIMARY DATA COLLECTION CONSIDERATIONS

- In rare cases, evaluators may decide to set up their own data collection mechanism for crowd sourcing. This may require a **significant amount of effort**, with work broken down into multiple stages:
 1. Developing an app, setting up a hotline, or establishing another mechanism for reporting.
 2. Promoting the data collection mechanism so the public knows what it is and how to use it.
 3. Running the mechanism and monitoring the incoming data.
- The costs of conducting the data collection will vary depending on the geographic scale, duration, amount of incoming data, and crowd-sourcing mechanism. However, in cases where researchers are setting up their own crowd-sourcing mechanism, there may be limited to no cost savings over a traditional household survey.
- It can be difficult to predict how widely the reporting method will be adopted by the public and may take time before the mechanism is widely socialized. This increases the risk the data collection could be unsuccessful.
- While it may be easier for crowd-sourcing methods to access dangerous neighborhoods or conflict hotspots than a team of survey enumerators, it is critical to prioritize *safety and ethics*:
 - What is the level of risk for someone who uses the crowd-sourcing platform?
 - If data collection identifies incursions by human rights abuses or other serious violations, is there an obligation to report this to authorities, and what are the implications for affected communities of doing so?

PREPARING DATA FOR ANALYSIS

- Preparation of crowd-sourcing data requires a similar level of skill as the preparation of household survey data. However, instruments for crowdsourcing are shorter than a household survey instrument. As a result, the amount of time and effort required to prepare crowd-sourced data is much less than for a household survey.

DATA ANALYSIS

- To prepare and analyze crowd-sourcing data, strong statistical programming skills in Stata or R are recommended.

REFERENCES

- Bruce, R., Cavgias, A., & Meloni, L. (2021). *Policy Enforcement in the Presence Of Organized Crime: Evidence from Rio de Janeiro*. Available at SSRN: <https://doi.org/10.2139/ssrn.3678840>
- Monteiro, J., & Rocha, R. (2017). Drug battles and school achievement: Evidence from Rio de Janeiro's favelas. *The Review of Economics and Statistics*, 99(2), 213–228. https://doi.org/10.1162/REST_a_00628
- Perez-Vincent, S. M., & Carreras, E. (2022). Domestic violence reporting during the COVID-19 pandemic: Evidence from Latin America. *Review of Economics of the Household*, 20, 799–830. <https://doi.org/10.1007/s11150-022-09607-9>
- Scoping Study on Gender Based Violence and Discrimination Against Women and Girls in Urban Public Spaces of the City of Skopje*. (2020). Reactor - Research in Action. <https://reactor.org.mk/en/wp-content/uploads/sites/4/2020/11/Reaktor-Scoping-Study-ANG-web.pdf>

RETURN TO INTRODUCTION

U.S. Agency for International Development

1300 Pennsylvania Avenue, NW

Washington, DC 20523

Tel: (202) 712-0000

Fax: (202) 216-3524

www.usaid.gov