

Violent Presence of Armed Actors in Colombia: Methodological Appendix

Javier Osorio

School of Government and Public Policy
University of Arizona

Version 1.1

January, 2019

Contents

1	Introduction	2
2	Funding	2
3	Violent Presence of Armed Actors	3
4	Information Source	4
5	Computerized Textual Annotation	6
6	Technical Aspects of the Map	11
7	Research Team	13
8	Data Access	14
9	How to Cite	14
10	Related Publications	14

1. Introduction

This project contributes to ongoing efforts of Colombia and international community to support the prospects and sustainability of peace by providing a systematic overview of the violent presence of armed actors in Colombia over time and across space. We hope the academic, policy, and civil society sectors will find this application useful to identify trends in the dynamics of conflict in Colombia.

This visualization presents an interactive heat map of the violent presence of armed actors in Colombia between 1988 and 2017. The data presents an array of state and non-state armed actors clustered in four main types: Government forces, Insurgent organizations, Paramilitary groups, and Criminal organizations. Each of these types of actors is further disaggregated in a multitude of specific organizations.

The database behind this interactive map is generated using Eventus ID, a protocol for supervised event coding capable of geo-referencing actors that appear mentioned in text written in Spanish (Osorio and Reyes 2016). The information used as an input for the computerized coding came from *Noche y Niebla*, a collection of narratives of human rights violations published by the Centro de Investigación y Educación Popular, CINEP (2016).

2. Funding

The Violent Presence of Armed Actors in Colombia interactive map is possible thanks to the generous support of:

- The United States Department of Defense - Minerva Research Initiative, "A Nested Mixed-Methods Approach to Armed Non-State Actor Governance and the Rule of Law," Desmond Arias (PI), Javier Osorio (Co-PI), Harold Trinkunas (Co-PI), and Vanda Felbab-Brown (Co-PI), Award No. 71623-LS-MRI.

- The University of Arizona, Research, Discovery & Innovation, Technology and Research Initiative Fund (TRIF).

3. Violent Presence of Armed Actors

The term *Armed Actor* refers to state and non-state armed entities that exercise the organized use of violence in a specific territory to achieve political or economic goals. For this particular research, the main types of armed actors include: government forces, insurgent organizations, paramilitary groups, and criminal organizations as indicated in Section 5. These main actors are further disaggregated in a variety of specific organizations. State armed groups refer to the formal institutions of the state in charge of providing internal and external security such as the Police or the Army. Non-state armed actors refer to a broad range of armed organizations outside the direct control of government authorities who are willing and able to use force. For a more technical definition of armed actors see International Committee of the Red Cross (2010, Part 1, Art. 1, p. 84), and for a more nuanced theoretical discussion see Berti (2016) .

The data indicates the *Violent Presence* of armed actors at the municipality-year level. Given the nature of the information source, this data reflects the location of armed actors involved in violent incidents which can be lethal (e.g. assassinations) or non-lethal (e.g. threats, displacement, etc.). However, it does not reflect cases in which armed actors are present in a territory but exercise no violence (Arjona 2011). Such cases would correspond to areas of dominant or monopolistic control of armed actors in which the use of violence might not be unnecessary (Kalyvas 2006). However, the information source might not report such cases, thus inhibiting the software from detecting non-violent presence of armed groups. In consequence, inferences drawn from this data should consider a narrow approach focused on the violent presence of armed actors.

4. Information Source

The information used to develop this data base comes from *Noche y Niebla*, a publication created by Centro de Investigación y Educación Popular (2016) comprising a collection of human rights violations narratives. CINEP has been gathering information related to human rights incidents taking place within the context of the Colombian conflict since 1987. For more than three decades, CINEP has been systematically reviewing, gathering, systematizing, and categorizing dispersed information from dozens of national and local newspapers on a daily manner, as well as from verbal and written testimonies issued by victims of human rights violations, their relatives, community leaders, other civil society organizations, and other relevant stakeholders. Being closely affiliated with the Catholic church, CINEP counts with a vast network of information nodes throughout Colombia, which provides CINEP a broad territorial network reaching even the most remote areas of the country.

As in all its different tasks, CINEP abides to the highest ethical and methodological standards for developing the collection of narratives comprised in *Noche y Niebla*. In this way, CINEP tries to be as encompassing as possible in their survey of documentation providing information about each incident, while committing to publishing only verified events based on cross-reference of sources. The high quality information produced by CINEP over the course of three decades contributed to consolidate this organization as one of the most prestigious civil society organizations in Colombia. Their contributions have been crucial for advancing the respect for human rights, increasing accountability and reparation, documenting historic memory, and promoting peace in Colombia.

We owe deep gratitude to CINEP, its leadership, and researchers, for making their information publicly available. This research would not have been possible without such a valuable public good.

The corpus used in this research comprises 45,375 daily reports of human rights

violations between 1988 and 2017 directly extracted from *Noche y Niebla* data bank.¹ After downloading all the narratives from the website, we integrated all the stories into a plain text file (.txt) that constitutes the corpus of this study. The corpus contains a series of short and systematic descriptions of violent incidents related to the Colombian conflict as reported in *Noche y Niebla*.

After a simple pre-processing procedure of the information, each line of the corpus contains four key elements:

- The date of occurrence in YYYYMMDD format
- A unique identifier for each narrative
- A brief description of the incident as curated by CINEP experts based on multiple accounts or information sources
- The location identified in the narrative

The following narratives present an example of information contained in the corpus:

20060812 20060812_NYN_P1_P30376 | Presuntos insurgentes del ELN asesinaron al joven Alfredo Torres , el hecho se presento en la vereda Patanguejo , corregimiento de Cajamarca . Segun la fuente , hace pocos dias la victima habia terminado de prestar servicio militar , Cauca , Mercaderes , Cajamarca

20060812 20060812_NYN_P1_P30377 | Unidades de la Policia Nacional y tropas del Ejercito Nacional , adscritas a la Brigada Movil 8 detuvieron arbitrariamente , sin orden judicial a las ocho personas citadas , las cuales con enganos fueron llevadas hasta la guarnicion militar donde se encuentra la Brigada Movil 8 , en ese lugar les informaron que ellos hacian parte de una lista de mas de once personas que presentaban orden de captura emitida por el Fiscal 40 de Ibague . Bajo esos argumentos y sin ellos esgrimir la orden , los enviaron a la sede de la Sijin . Tirsa estuvo detenida dos dias en ese lugar y fue puesta en libertad , Tolima , Planadas

20060812 20060812_NYN_P1_P30378 | Guerrilleros de las FARC EP secuestraron en el sitio Piedras Blancas , vereda Santa Teresita , al funcionario de la Corporacion Regional de Uraba , Corpouraba , Antioquia , Mutata

¹See https://www.nocheyniebla.org/?page_id=566

20130720 20130720_NYN_P1_P40367 | Paramilitares autodenominados Los Rastrojos lanzaron una granada contra un establecimiento publico , ubicado en la Avenida 3 entre calles 6 y 7 , barrio Latino , dejando como resultado tres personas muertas y diecinueve heridas , ademas de danos materiales . Este hecho ocurrio a las 4:30 a . m . En la ciudad han circulado amenazas contra establecimientos publicos , que se han materializado en atentados y varios muertos y heridos , Norte de Santander , Cucuta.

In each narrative, the first element presents the date of the event. The second element is an unique identifying for each incident. The third element is an account of the events as *Noche y Niebla* described based on their multiple information sources. The final elements refer to the locations of the incident, starting with the department(s), and followed by the municipality(ies).

5. Computerized Textual Annotation

To generate the database of violent presence of armed actors in Colombia, the project relied on Eventus ID, a software for supervised event coding from text written in Spanish (Osorio and Reyes 2016). In contrast to more recent innovations in computerized event coding such as Universal Petrarch (Open Event Data Alliance 2014), Eventus ID belongs to the family of coders that rely on sparse parsing to process the text. Prominent programs such as KEDS, TABARI, and Petrarch belong to this category of coders (Schrodt, Davis and Weddle 1994; Schrodt 2009; Schrodt, Beiler and Idris 2014). As such, Eventus ID uses a set of dictionaries to identify event data, which is defined as a categorical description of someone (source actor), doing something (action), to someone else (target actor), in a give date (time), and in a specific place (location). In this way, the software processes textual information in order to identify who did what to whom, when and where. Once it identifies these elements in the corpus, the program extracts the textual information and transforms it into numeric format for data analysis.

In this particular project, the research is only focused in identifying the geographic location of armed actors as reported in the narratives. To do so, the coding protocol

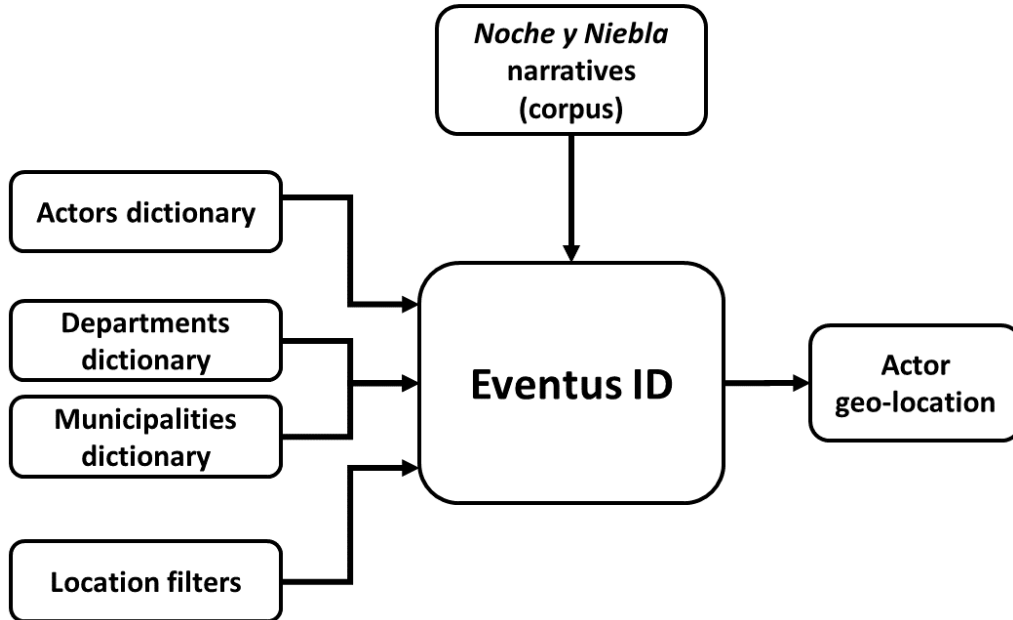
only uses a dictionaries of actors and locations (departments, municipalities, and location filters) to conduct this task. Advanced users will be able to find further technical details about Eventus ID and execution instructions in Github: https://github.com/javierosorio/Eventus_ID_2.0.

Figure 1 describes the general process of geo-locating actors using Eventus ID. The algorithm implements the following steps:

1. The software takes as the main input the corpus of *Noche y Niebla* narratives of human rights violations.
2. Eventus ID uses the content of the actors dictionary as search criteria to look for a match in each line of the corpus.
3. If there is an actor match, Eventus ID stores the matching text and identifies the corresponding numeric code as indicated in the actors dictionary.
4. For each corpus line containing an actor match, the program uses the dictionaries of departments and municipalities (which correspond to states and counties, respectively) to look for a location name (also know as toponym) in the same line in which it identified an actor.
5. If there is a match with the locations dictionaries, the program then verifies the location filters to identify whether or not this location name is a false positive or an actual toponym.
6. If there is no match in the location filters, Eventus ID saves the coded actor and locations in the output database.

The actors dictionary contains an encompassing list of 7,560 names of actors (last update on 8/28/2018). The menu of actors comprises four main types of armed groups: Government, Insurgents, Paramilitaries, and Criminals. The list of nouns contains a

Figure 1: Actor geo-location using Eventus ID



collection of names of state and non-state armed actors and their respective acronyms; it also includes a list of their subgroups (e.g. battalions, blocks, or fronts) and main leaders. For redundancy purposes, the actors dictionary considers variations of different armed actor's names. This is important as it allows to cover different variations in unstructured text. Table 1 presents the list of 30 armed groups considered in this research as grouped by the main type of actor.

Each actor name in the dictionary has a numeric code associated with it as corresponding to the codes in Table 1. This numeric code facilitates the statistical analysis. To identify actors in the text, Eventus ID uses list of nouns of the actors dictionary as search criteria to detect actors in the *Noche y Niebla* corpus. For example, if the actors dictionary includes the name "FARC", the system will take this name and look for it in the text. Once the program finds a match for "FARC" in the corpus, the system records the date of occurrence, extracts the textual match, and assigns the corresponding numeric code for this actor. These annotations are stored in a coding output database.

Table 1: List of actor types and organizations

Actor Type	Organization	Code
Government 01	Generic mentions of government	100
	Army	101
	Airforce	102
	Navy	103
	Police	104
	Attorney's General Office	105
	Intelligence Agency	106
Insurgents 02	Generic mentions of insurgents	200
	Revolutionary Armed Forces of Colombia (FARC)	201
	National Liberation Army (ELN)	202
	Popular Liberation Army (EPL)	203
	People's Revolutionary Army (ERP)	204
	9th of April Movement (M-19)	205
	Ernesto Rojas Commandos (CER)	206
	Simon Bolivar Guerrilla Coordinating Board (CGSB)	207
Paramilitaries 03	Generic mentions of paramilitaries	300
	United Self-defense forces of Colombia (AUC)	301
	Popular Revolutionary Anti-Terrorist Army of Colombia (ERPAC)	302
	United Gaitan Self-defense forces (AGU)	303
	Peasant Self-defense forces of Cordoba and Uraba (ACCU)	304
	Social Cleansing	305
	Independent paramilitary groups	306
Criminals 04	Generic mentions of criminal bands (BACRIM)	400
	Los Urabenos	401
	Los Rastrojos	402
	Aguilas Negras	403
	Los Paisas	404
	Envigado Office	405
	Los 400	406
	Independent criminal groups	407

The categorization of actor types as government, insurgents, paramilitaries, and criminal organizations follows the general convention of the main actors involved in the Colombian Conflict. However, we acknowledge that the behavior of some armed actors is not rigid and sometimes can reflect the characteristics of a different actor type. For example, paramilitary groups sometimes act as predatory criminal organizations. Or the distinction between military and paramilitary groups can sometimes be blurry. In addition, we acknowledge that the complexity and fluidity of the Colombian conflict often involves cases in which actors (implicitly or explicitly) move across categories. For

example, the armed group “Los Urabeños” has its origin in paramilitary groups, but later evolved into a criminal organization. In this version of the data base (1.1), we do not account for these fluid cases that change across categories. In consequence, the categorization of each specific organization remains fixed within their general category.

To identify the locations, Eventus ID relies on two dictionaries, one for departments (equivalent to states) and another one for municipalities (equivalent to counties). In this way, the software is capable of geo-referencing actors at the sub-national level. The departments dictionary contains the names of all departments in Colombia. In a similar manner, the municipalities dictionary includes all the municipalities in the country and a variety of name variations. These dictionaries include the official department or municipal codes used by Colombia’s Census Bureau (*Departamento Administrativo Nacional de Estadística*, DANE). Using these official location codes facilitates merging the output coding data with official statistics and other databases.

A common challenge in computerized event data relates to correctly identifying the place of occurrence of an event (Lee, Liu and Ward 2016; Chalabi 2014). To address the challenge of geographic disambiguation, Eventus ID uses a filter of locations to eliminate matches that might look as a location names, but do not actually refer to locations. For example, the locations filter includes the name of “Universidad del Magdalena” to prevent the program erroneously identify this name as the department of “Magdalena.”

After detecting an actor in a specific line, Eventus ID uses the department and municipality dictionaries to identify the location of an event as mentioned in such line of the corpus. After identifying a location, the software verifies if there is a match in the location filter. If that is the case, the software ignores this name as a possible location. If there is no match, the program saves the location name and corresponding numeric code in the output database.

For each municipality, the database includes the geographic longitude and latitude

coordinates as reported by the Database of Global Administrative Areas.² This allows projecting the coded data into a map. It is important to clarify that the projection *does not provide the exact location* of an event as based on a neighborhood, specific street address, or a geo-tag marked in the field. The projection in the map is just a generic approximation of the location based on the geographic coordinates of the municipality.

After coding actors and locations using Eventus ID, the methodology considers a post-coding process for cleaning the data in Stata. This protocol includes a deduplication routine to eliminate multiple matches of the same actor in the same municipality day. In this way, the analysis avoids artificial inflation and considers a conservative approach for including in the cleaned database one single actor per organization, for each municipality-day. The protocol also includes a process for location deduplication and cleaning to reduce concerns of false positives. The final data set includes the type of actor, the specific organization recorded, the date of occurrence, as well as the department and municipality identified.

6. Technical Aspects of the Map

This section provides technical information related to the Geographic Information Systems (GIS) used in this project. We used ArcGIS Pro v2.2 to develop the map and specify its main configuration. After developing the map, we used ArcGIS Online to deploy the interactive map application in the web. Both ArcGIS Pro and ArcGIS Online are proprietary products of the Environmental Systems Research Institute (ESRI).³ We used these programs through the institutional licences of the University of Arizona. The coordinates system used in this project corresponds to WGS84 Web Mercator (Auxiliary Sphere).

²See <https://gadm.org/data.html>.

³See <https://www.esri.com/>.

The visualization presented in the map corresponds to heat maps of violent presence of armed actors. A heat map is a GIS analysis tool that provides a visual representation of high-density of occurrence or clustering of a phenomenon. In contrast to hot-spot analysis tools that rely on statistical measures of spatial auto-correlation, heat maps provide a more flexible (yet, less rigorous) way of clustering the observations and assigning a color gradient to indicate higher concentration.

The raw information used for creating the heat map came from the computerized data generated with Eventus ID. Each data point corresponds to the identification of a specific armed actor in a given municipality-day as mentioned in the *Noche y Niebla* narratives. As indicated before, the incidents do not have the specific xy coordinates of their occurrence. Instead, their position in the map is assigned to the corresponding latitude and longitude of the center of the municipality. As a result, points often overlap in the map projection. This becomes evident when users zoom into the map.⁴

To build a heat map, ArcGIS uses a normal distribution (also known as Gaussian distribution) to assess the strength of the influence of each point over an specific area as determined by a radius.⁵ The program renders a higher intensity to areas that concentrate a higher density of data points. To represent these areas, the program assigns a more intense color gradient than the one assigned to low intensity areas.

In contrast to a fixed raster visualization, the ArcGIS Online application dynamically adjusts the radius to render the heat map based on the projection scale that the user prefers. In this way, the heat map will provide a more aggregate or disaggregate clustering of the data points as the user zooms out or into the map.

⁴To identify how many points overlap in a specific location, users can use the mouse to make a right click on any cluster. This will open a pop-up window indicating the number of points overlapping in this location.

⁵For technical details about the methodology used to generate the heat maps, see: <https://developers.arcgis.com/javascript/latest/api-reference/esri-renderers-HeatmapRendererer.html>.

7. Research Team

- Dr. Javier Osorio, Principal Investigator, Assistant Professor, School of Government and Public Policy, University of Arizona.

Website: <https://www.javierosorio.net/>

Twitter: @javier_osorio

- MS. Mohamed Mohamed, graduate student, Master in Geographic Information System Technology, University of Arizona.
- Dr. Viveca Pavon, Postdoctoral Research Associate, School of Government and Public Policy, University of Arizona.
- Dr. Susan Brewer-Osorio, Center for Latin American Studies, University of Arizona.
- The project had superb support from an excellent team of graduate and undergraduate research assistants at the University of Arizona:
 - Alejandro Beltrán, PhD. student, School of Government and Public Policy.
 - Francly Luna, undergraduate student in Political Science and Law.
 - Juan Pablo Peñaloza, undergraduate student in Management Information Systems, and Spanish Translation and Interpretation.
 - Abel Macias, undergraduate student in Criminal Justice; ROTC, 5th BDE, Wildcat Battallion.
 - Rubyalejandra Najera, undergraduate student in Public Management and Policy, and Criminal Justice Studies.

8. Data Access

The database and replication files will be released after the first publication derived from this research project. Users will be able to find the full dictionaries, database, and replication files in the Github page of this project at: <https://github.com/javerosorio/>.

Once available, users can request access to the data at: <https://www.colombiaarmedactors.org/>

9. How to Cite

Osorio, Javier. 2018. *Violent Presence of Armed Actors in Colombia*, <https://www.colombiaarmedactors.org/>

BibTeX:

```
@misc{Osorio2018,  
author = {Osorio, Javier},  
title = {{Violent Presence of Armed Actors in Colombia}},  
url = {https://www.colombiaarmedactors.org/},  
year = {2018}  
}
```

10. Related Publications

- Osorio, Javier and Alejandro Reyes. 2016. "Supervised Event Coding from Text Written in Spanish: Introducing Eventus ID." *Social Science Computer Review*. Online <https://journals.sagepub.com/doi/abs/10.1177/0894439315625475>.

Version History

- **Version 1.0:** 11/12/2018
 - Beta version
- **Version 1.1:** 01/21/2019
 - First public version

References

- Arjona, Ana M. 2011. "Presencia vs. violencia: problemas de medición de la presencia de actores armados en Colombia."
URL: <http://focoeconomico.org/2011/12/20/presencia-vs-violencia-problemas-de-medicion-de-la-presencia-de-actores-armados-en-colombia/>
- Berti, Benedetta. 2016. "What's in a name? Re-conceptualizing non-state armed groups in the Middle East." *Nature* pp. 1–8.
URL: <https://www.nature.com/articles/palcomms201689.pdf>
- Centro de Investigación y Educación Popular. 2016. "Noche y Niebla: Banco de Datos."
URL: <http://www.nocheyniebla.org/>
- Chalabi, Mona. 2014. "Mapping Kidnappings in Nigeria."
URL: <https://fivethirtyeight.com/features/mapping-kidnappings-in-nigeria/>
- International Committee of the Red Cross. 2010. Protocols Additional to the Geneva Conventions of 12 August 1949. Technical report.
URL: https://www.icrc.org/en/doc/assets/files/other/icrc_002_0321.pdf
- Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War*. Cambridge, MA: Cambridge University Press.
- Lee, Sophie, Howard Liu and Michael Ward. 2016. "Lost in Space: Geolocation in Event Data."
URL: <https://arxiv.org/abs/1611.04837>
- Open Event Data Alliance. 2014. "Objectives - Open Event Data Alliance."
URL: <http://openeventdata.org/>
- Osorio, Javier and Alejandro Reyes. 2016. "Supervised Event Coding From Text Written in Spanish: Introducing Eventus ID." *Social Science Computer Review* . Online first.
URL: <https://journals.sagepub.com/doi/abs/10.1177/0894439315625475>
- Schrodt, Philip A. 2009. "TABARI. Textual Analysis by Augmented Replacement Instructions."
URL: <http://eventdata.parusanalytics.com/software.dir/tabari.html>
- Schrodt, Philip A., John Beieler and Muhammed Idris. 2014. Three's a Charm?: Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance. In *International Studies Association*. Toronto: .
URL: <http://parusanalytics.com/eventdata/papers.dir/Schrodt-Beieler-Idris-ISA14.pdf>
- Schrodt, Philip A., Shannon G. Davis and Judith L. Weddle. 1994. "Political Science: KEDS - A program for the machine coding of event data." *Social Science Computer Review* 12(4):561–587.