Annual Data Management and Cleaning

ACLED takes a variety of steps in ensuring that the data we publish are both accurate as well as easy for academics, policymakers, and practitioners to work with and implement into their own research. In addition to weekly review of the real-time conflict data, an overarching cleaning and data management effort is undertaken on an annual basis, leading to the annually-released dataset publication. Here, data cleaning and management are explored.

A list of events flagged over the course of a time period is reviewed for accuracy. Often, in the immediate aftermath of a conflict event, not all details may be known. For example, a group may not claim responsibility of an event until some time after the conflict event. Given ACLED's real-time coding of conflict, when information about the group is not known, ACLED will code the group as an unidentified armed group, and revise accordingly with new information. If new information about the group surfaces at a later date – e.g., a group comes forward claiming responsibility for an attack – then the event is updated to reflect the new information. Another example may be inconsistent reports in the aftermath of a conflict event, especially in regard to the number of fatalities. Over time, more in depth reports may surface, such as those by human rights organizations. These details are updated in the already-existing events in order to ensure the most accurate conflict coding is presented. Brief notes are also included with each conflict event in order to offer additional details about each event; these notes are also checked and updated to reflect the most accurate information.

In the same vein, any events that had not previously been coded and included in the dataset are included; again, as time passes, more information may surface about additional conflict events (e.g., an additional conflict location or date of conflict). These conflict events are added in order to be able to most accurately understand conflict patterns.

ACLED coders maintain a list of conflict actors, noting the name and classification of actors based on their stated goals and objectives. The names and spellings of conflict actors are reviewed as any differences in the name or spelling of a group can result in difficulties during data analysis, especially when relying on software where spelling differences in the name of an actor may result in assuming distinctly different actors.

Over time, these goals and objectives – and hence classification – can change, especially as groups grow or splinter. For example, what may at first be a part of a state military force may over time give way to a rebel movement, such as the mutiny of the military forces of South Sudan and the emergent SPLA/M-IO rebel movement. Or a political militia may take on a new goal of striving to overthrow a state, which would change their classification from a political militia to a rebel movement. Rebel groups may also splinter into new factions as different rebel leaders begin pushing varying agendas (e.g., the FDLR rebel group operating in Rwanda split into a number of different factions over time). ACLED classifies conflict actors as a (1) state force, (2) a rebel group, (3) a political militia, (4) a communal/ethnic militia, (5) a rioter, (6) a protester, (7) a civilian or an (8) external force. These individual numbered codes are aggregated to from a dyadic interaction code. Maintaining the most accurate and up-to-date code helps to ensure the most accurate representation of conflict dynamics. For more information on actor coding, please see the ACLED Codebook.

The location of conflict events is also checked in order to ensure the most accurate mapping of conflict events when exploring spatial relationships. The georeferencing of these conflict events is also further checked annually through the use of GIS software.

Lastly, after the compilation and cleaning of all events in the dataset, new event IDs are assigned to each conflict event in order to aid in data management for users. Each conflict event is assigned both an overarching event ID within the dataset as a whole, as well as a country-specific event ID. The data are publicly available both as a dyadic file (where each conflict event is a single unit of analysis) and a monadic file (where the involvement of each actor in a conflict event is a single unit of analysis). While the former is more useful in understanding conflict dynamics and patterns, the latter is useful in analyses of conflict actors and/or actor types. Conflict data are also available disaggregated by specific countries as well as by conflict type. Lastly, a shapefile of data is also publicly available for use in GIS analysis.