Adding New Sources to ACLED Coverage

ACLED believes that only a tailor-made sourcing process for individual regions/countries will make data more reliable. Given the variation in types of disorder, available sources, and potential biases amongst countries, ACLED develops sourcing strategies adapted to the specific challenges at hand in each unique context. The goal of these strategies is to construct source combinations that approximate the reality of violence in each individual country/region *(for more on ACLED's sourcing strategy, see [this primer](#)).*

In line with this belief, as the types or patterns of disorder in a country or region shift, as the availability of sources changes, or as new biases may emerge as a function of the political landscape, it is important to consistently review each sourcing profile to ensure that the combinations continue to most accurately approximate local realities. Otherwise, shifts in the political landscape may impact trends in disorder and/or the media environment, which may in turn shift the reporting reliability of sources.

As such, ACLED strives to regularly review its sourcing across regions/countries to ensure that reported trends continue to be reliable.

Weighing real-time vs. historical integrity

However, adding new sources to coverage comes with important costs that must be weighed: namely, the impact on real-time versus historical trends in the data.

If a new source is identified that yields a number of unique events in real-time, adding the source immediately to the data would ensure that justice is done to real-time trends: capturing all reported events that are happening in real-time. However, the addition of such a source in an *ad hoc* fashion risks the integrity of historical trends as it will introduce an 'artificial spike' in the data. This refers to the phenomenon where if that same source was first back-coded before being introduced into the data, the 'spike' that its inclusion introduces in the data would be gone (or minimized) -- suggesting that the spike does not reflect a 'true spike' in disorder on the ground.

Avoiding such artificial spikes would require that the source first be back-coded to the beginning of ACLED coverage before being published. This process, however, takes significant time and resources. This is especially so in regions where ACLED coverage extends back in time significantly. Further, this is only made more difficult by the fact that there is a moving target: with the passage of time, the period of time that must be back-coded grows. The result: it can take significant time before such new sources are back-coded and then published. In cases where resources are limited

or unavailable, such a project can end up stagnated or tabled for significant periods of time, if not indefinitely, until resources become available.

While such a system may better ensure the integrity of historical trends in the data, during the time it takes to complete such a back-coding project, real-time trends in the data knowingly suffer. This is because there are events -- captured only by the new source, hence warranting its inclusion -- that are not being published in real-time.

While avoiding the introduction of artificial spikes in the data is important, there are cases in which such spikes are unavoidable. For example, the historical archives for a new source may not extend as far back in time as ACLED coverage for the country. In such cases, the addition of the source may result in an artificial spike at the date when historical archives begin. Or, a source may be born out of necessity given changes in a country's political landscape. Such sources will very often begin their coverage when the source is born. If this date does not extend as far back in time as ACLED coverage for the country, its inclusion may result in an artificial spike in the data.

While the introduction of such sources may introduce bias into the data -- since an artificial spike does not necessarily reflect a spike in disorder on the ground -- being resistant to changes in the media environment in a country is itself another form of bias. Often if a new source is born out of necessity, it is because other sources reporting in the region may not have been doing justice to the reporting of real-time trends. Avoiding inclusion of such new sources would mean being resistant to adaptation to ensure the best coverage of trends on the ground.

ACLED's decision: Accurately capturing both real-time and historical trends, relative to each other

In an effort to do justice to both real-time and historical trends, ACLED strives to capture both without privileging one period over another.

In an effort to accurately capture historical trends relative to present patterns, ACLED will not add a high-yielding source to the data on an *ad hoc* basis. If a new, high-yielding source is identified, ACLED will first back-code the source historically before publishing the events it yields.

However, in order to accurately capture real-time trends relative to historical patterns, ACLED will publish data coming from a new source in multiple tranches: first, once it is back-coded to 2018, and then again once it is back-coded to the start of ACLED coverage, or at minimum to the start of the source's coverage.

This system allows users who rely on real-time data trends to not have to wait long periods of time (or indefinitely) for a new source to be added to ACLED coverage. This is imperative especially in those cases where a new source contributes significantly to ACLED's real-time coverage -- and hence impacts users' understanding of trends in real-time.

Meanwhile, this system also allows users who rely on historical data trends to understand spikes in the data (especially around 2018) in order to better contextualize their findings.

2018 serves as a natural point of reference for ACLED's publication tranches. Nearly all regions of ACLED coverage extend back to at least 2018 (*for more on ACLED's temporal coverage across countries, see the continuously updated table accessible* [here](#)). Those wishing to do cross-regional analysis using ACLED data will hence often not use data before this date. Further, many sources, especially 'new media' sources, have been increasingly prolific since that time -- in line with internet penetration around the world. 2018 hence serves as a natural date to account for this increase in accessibility and availability as well.

The addition of a new source to the data may result in updating previously published events with further information and/or adding additional events to the ACLED dataset. However, it should be noted that increasing the number of sources in the data is not necessarily correlated with more events being reported in the dataset nor with increased reliability of the data. (*For more on this relationship, see the section titled "Do more sources mean that data are more reliable?" in* [this FAQ](#) *covering ACLED's sourcing methodology.*) In fact, more localized sources will often report more 'unique events' than other sources at higher scales. As such, adding a new local source may in turn render sources at a higher scale less useful -- e.g.. a single local source may be effective in capturing events previously captured across three sources at a higher scale. This may mean that those sources at higher scales may hence be dropped from ACLED's source lists should they no longer produce events which are not captured elsewhere. What this means tangibly is that there is not a single 'control variable' that can be used in temporal analysis to account for source variation over time. This is not only because the number of sources is not correlated with the number of events, but also the addition of sources to the ACLED dataset is not random, meaning that certain sources may have been added to ACLED's coverage *specifically because* they were better able to capture an emerging trend. It is not that the addition of a source results in event counts around a specific trend to rise; rather, it is often the case that a specific trend resulting in increased violence is *why* a certain source may have been added to ACLED's coverage. Singular, specific sources can better capture trends and reflect local reality. Users of ACLED's historical data should hence familiarize themselves with the contexts which they cover to determine how they choose to account for such changes in sourcing over time; unfortunately there is no easy answer in the form of a single 'control variable'. Nevertheless, users are always welcome to reach out to ACLED with questions and further guidance.

The regular review of sourcing is imperative in ensuring the most accurate and reliable information on the ground in real-time. ACLED, as a real-time crisis mapping project, strives to do justice to these trends. It is equally imperative, however, that users looking at historical trends in the data understand the sourcing strategy behind the trends in order to best contextualize any shifts in trends over time.