

Integrating Conflict Event Data

Karsten Donnay^{1,2}, Eric T. Dunford³,
Erin C. McGrath⁴, David Backer⁵,
and David E. Cunningham^{3,6}

Journal of Conflict Resolution

1-28

© The Author(s) 2018

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0022002718777050

journals.sagepub.com/home/jcr

Abstract

The growing multitude of sophisticated event-level data collection enables novel analyses of conflict. Even when multiple event data sets are available, researchers tend to rely on only one. We instead advocate integrating information from multiple event data sets. The advantages include facilitating analysis of relationships between different types of conflict, providing more comprehensive empirical measurement, and evaluating the relative coverage and quality of data sets. Existing integration efforts have been performed manually, with significant limitations. Therefore, we introduce Matching Event Data by Location, Time and Type (MELTT)—an automated, transparent, reproducible methodology for integrating event data sets. For the cases of Nigeria 2011, South Sudan 2015, and Libya 2014, we show that using MELTT to integrate data from four leading conflict event data sets (Uppsala Conflict Data Project—Georeferenced Event Data, Armed Conflict Location and Event Data, Social Conflict Analysis Database, and Global Terrorism Database) provides a more complete picture of conflict. We also apply multiple systems estimation to show that each of these data sets has substantial missingness in coverage.

Keywords

event data, data integration, conflict, civil wars, protests

¹Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany

²Department of International Relations and Political Science, Graduate Institute of International and Development Studies, Geneva, Switzerland

³Department of Government and Politics, University of Maryland, College Park, MD, USA

⁴National Consortium for the Study of Terrorism and Responses to Terrorism, University of Maryland, College Park, MD, USA

⁵Center for International Development & Conflict Management, University of Maryland, College Park, MD, USA

⁶Peace Research Institute Oslo, Oslo, Norway

Corresponding Author:

Karsten Donnay, Department of Politics and Public Administration, University of Konstanz, Konstanz 78457, Germany.

Email: karsten.donnay@uni-konstanz.de

During the “Arab Spring,” antiregime action diffused across the Middle East and North Africa. Protests in rural Tunisia in December 2010 quickly spread to the rest of the country. Within weeks, protests were occurring in major urban areas of many countries around the region. From this initial spark, the specific path of contention varied widely. In Egypt and Tunisia, dissent remained primarily nonviolent and resulted in regime change. In Libya, Syria, and Yemen, governments responded to protests with massive repression, then dissidents turned to violence, leading to civil war. In other cases, protest campaigns died out without regime change or a transition to violent conflict (Lynch 2013, 2016).

When and where do protests occur? Under what conditions is violence more likely? Why do some protest campaigns dissipate, others expand, and select ones transition to full-scale civil war? Questions like these continue to be of great interest to conflict researchers. The ability to answer such questions has been transformed by the emergence of data sets of conflict phenomena providing event-level information that is spatially and temporally disaggregated, often measured specific to a place and a date. Scholars have used these event data sets to examine significant topics related to contentious politics, including variation in the location and timing of demonstrations, riots, terrorist attacks, battles between rebels and security forces, and atrocities against civilians.¹

In this article, we introduce Matching Event Data by Location, Time, and Type (MELTT), a novel methodology for integrating event data sets. Despite the availability of a growing number of these data sets, scholars tend to draw on just one of the data sets at a time when conducting analyses. While this choice may be appropriate for certain analyses, we contend that integrating information from multiple data sets has the potential to enhance conflict research. Different data sets often contain information about different types of contentious activity—such as protests, terrorist attacks, battles, state repression, and so on. More than any single existing data set, integration of multiple data sets facilitates studying whether and how these phenomena relate to one another and addressing questions such as why contention escalates in some cases and not in others.

Even when scholars are interested in a particular phenomenon, analyses could be improved by integrating information from multiple data sets. No single data set is necessarily comprehensive in coverage. Each data set may omit certain events of interest. If multiple data sets overlap in terms of the types of events they explicitly intend to capture, this overlap presents opportunities to augment coverage and to evaluate the extent to which each data set exhibits missingness.

Scholars have shown interest in integrating conflict data sets. For example, Findley and Young (2012) match terrorist events to polygons of armed conflict, while Fortna (2013) and Polo and Gleditsch (2016) match terrorist attacks to armed groups. These studies represent valuable progress in conflict research. With respect to integration, however, each study exhibits shortcomings: (1) event data

are matched to other forms of data, (2) the process is confined and tailored to the needs of a given research topic, and (3) matching is performed manually, which can be time-consuming, nontransparent, and difficult to replicate.

In contrast, the MELTT approach originates from the principles that data integration should be (1) automated, to facilitate efficient integration of large data sets in a manner that can be too resource intensive and even intractable to perform by hand; (2) transparent, written in the form of a structured protocol, to clearly communicate how integration is undertaken and to guarantee reproducibility; and (3) adaptable, to accommodate choices of data sets and parameters of comparison. MELTT embodies all three principles. The formal protocol we have developed entails systematically comparing all entries from multiple event data sets, identifying entries about events that co-occur spatiotemporally, examining other available information about each set of these possibly matching entries, and on this basis ultimately evaluating which entries code unique events or instead provide complementary or matching information about the same event. The full functionality of the MELTT protocol is made publicly available for use by the research community through the open-source R package `meltt`, accessible on the Comprehensive R Archive Network (CRAN).

We demonstrate the utility of event data integration in practice through three illustrative applications. First, we conduct Monte Carlo simulations on synthetic data sets to show that MELTT reliably identifies unique and matching entries in a wide range of typical situations encountered in the analysis of event data sets. Second, we rely on the protocol to integrate four well-established conflict event data sets: the Armed Conflict Location and Event Data (ACLED; Raleigh et al. 2010), the Uppsala Conflict Data Project–Georeferenced Event Data (UCDP-GED; Sundberg and Melander 2013), the Global Terrorism Database (GTD; START 2013), and the Social Conflict Analysis Database (SCAD; Salehyan et al. 2012). We manually validate the integrated data for a diverse selection of country-years: Nigeria 2011, South Sudan 2015, and Libya 2014. This validation demonstrates that MELTT identifies matching entries with high accuracy. Third, we use the integrated data for Nigeria 2011 to conduct multiple systems estimation (MSE) analysis (Manrique-Vallier, Price, and Gohdes 2013). The results of this analysis suggest that the events covered by the four data sets represent a biased segment of the larger universe of violent conflict in Nigeria. Taken together, these illustrations provide tangible evidence of the benefits of integration and of the reliability of our approach.

Why Integrate Conflict Event Data Sets?

Integration can benefit the quantitative analysis of conflict by allowing researchers to examine the relationship between types of events contained in different data sets and by providing more comprehensive measurement of phenomena. An increasing number of data sets contain event-level information on conflict, some capturing

countries or regions, others affording global coverage.² In this article, we focus on ACLED, UCDP-GED, GTD, and SCAD because they are the most comprehensive and widely used event data sets in quantitative studies of conflict. Citation statistics from the Web of Science indicate that the respective articles of record for the four data sets were cited a total of 377 times since their initial publication, including 81 times in 2017 alone. In addition, these four data sets offer a mixture of distinct and overlapping event types, with significant overlap of coverage for Africa since the 1990s. The interlinking lines in Figure 1 depict the coverage of different event types (right) and geographic regions (left) for each of the data sets (middle). The graphic confirms that these data sets exhibit significant conceptual and geographic overlap, but do not fully coincide.

A clear benefit of integrating these data sets is the ability to study the relationship between different conflict phenomena. In particular, researchers have recently been interested in understanding the relationship between civil war and terrorism, which has motivated efforts to link the GTD to UCDP-GED (Findley and Young 2012; Fortna 2013; Polo and Gleditsch 2016). There is still much to learn about how these types of violence relate to one another, as well as about the connections among many other types of conflict. Are violent actions (such as insurgency and terrorism) and nonviolent actions (such as protests, strikes, and demonstrations) substitutes for one another or complementary? How does the occurrence of violence or nonviolence in one area in one period influence actions in the same or neighboring areas in subsequent periods? When does state repression of one type of dissident activity lead to a switch to different activity? All these questions are raised by the complex pattern of protest, terrorism, governmental repression, and civil war observed across the Middle East and North Africa region during and after the Arab Spring. More broadly, these questions are central to understanding the dynamics of conflict occurring around the world. Studying those dynamics properly often requires pulling data from multiple diverse data sets.

Researchers interested in a particular conflict phenomenon—such as violence or protest—could (and usually do) conduct quantitative analyses on just a single event data set. Even in these instances, integration has potential advantages. Certain conflict event data sets, by nature, fail to capture the full extent of the phenomenon in which a researcher is interested. For example, UCDP-GED only covers the constituent events that comprise periods of active armed conflict, per UCDP's own definition. Consequently, UCDP-GED records violent events for Libya in 2011, when the country descended into civil war, but excludes events in both 2010 (prior to the civil war becoming active) and 2013 (when the civil war was deemed to be below the threshold of active armed conflict). This data set is well suited for measuring violent events within the context of active armed conflict. If researchers want to study violent events more broadly, however, integration with another data set (such as ACLED or GTD) that also captures events outside of active armed conflict could lead to a more comprehensive measurement of subnational patterns of violence.

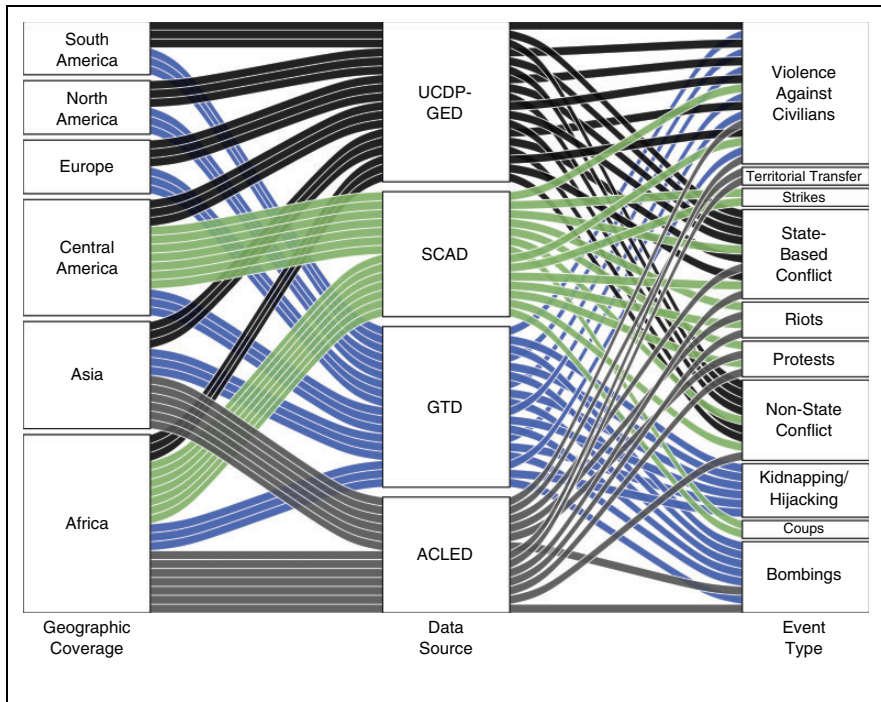


Figure 1. Geographic and event type overlap across four prominent event data sets.

When multiple available data sets seek to capture the same conflict phenomena, integration can help alleviate problems associated with missingness in individual data sets. All conflict event data sets are likely to mischaracterize and overlook at least some cases, in ways that are not necessarily consistent.³ A main source of most of these data sets is news media reports, which are known to be subject to discrepancies, omissions, vagueness, mistakes, and willful distortion (Donnay and Filimonov 2014; Weidmann 2015, 2016). Integrating information from multiple data sets can be a meaningful step in developing quantitative measures of the degree of missingness and biases in measurement of particular conflict phenomena.

Finally, even when the same event is reported in multiple data sets, they may capture and report different information about this event. Information in one data set may complement the information in another data set—filling in additional details on a given event. Alternatively, the information could deviate across data sets, in which case the compiled information serves to acknowledge uncertainty in knowledge about aspects of an event. Either way, integration tends to contribute to more careful measurement and offers a gauge of the consistency of the coding of individual events across sources.

An illustrative example illuminates how integrating complementary information from SCAD, UCDP-GED, and ACLED about events occurring on January 29, 2011 captures dynamics of violence in North-Central Nigeria that are not visible with just one data set. SCAD records sectarian riots that led police to use indiscriminate force, information corroborated by UCDP-GED, which adds an estimate of fifteen casualties. UCDP-GED further suggests that the relevant sectarian dimension is a Christian–Muslim divide. ACLED records that the use of force by police prompted student riots. Combining the perspectives of the three data sets paints a much fuller and more nuanced picture of how ethnoreligious riots, which were met by state force, led to youth riots, which were also met by state force, all of which then led to larger riots and destruction of infrastructure.

Thus, integration, if done properly, has the potential to advance understanding of conflict processes in multiple ways. Before proceeding, we emphasize that integration is not a panacea—in particular, an approach that solves all issues in the empirical measurement of conflict. If any of the data sets being integrated is prone to systematic gaps, biases, and inaccuracies, these shortcomings do not disappear upon integration, but rather will carry over to the output. The picture obtained from an event data set of good quality can be distorted by combining it with a data set of poor quality. Two data sets of poor quality, when integrated, can reinforce one another’s flaws, instead of yielding a picture with superior coverage, accuracy, and resolution.

Therefore, a choice to undertake integration and to give credence to the output and its applications in analysis hinges on what the researcher knows about the quality of individual data sets—especially the data-generating process, including their respective sources of information, coding protocols, and validation checks. Of course, these issues of quality are not avoided when relying on a single data set. Researchers should evaluate the quality of relevant data sets when deciding whether to use, or integrate, any data set. If researchers invest in evaluating data sets, make informed decisions, and articulate their rationale, integration of conflict event data sets is both valuable and viable.

Challenges of Integration

In practice, integration of conflict event data sets remains rare. Among the reasons are that integration presents methodological and practical challenges.

The first challenge is that the same event may be reported inconsistently across data sets because they rely on distinctive sources and coding protocols. An event could be coded as occurring on different days, in different locations,⁴ and with a different number of casualties, as well as being labeled as a different type of conflict activity. For example, UCDP-GED reports rebels pushing government forces from Misratah, Libya, and retaking the airport on May 10, 2011, as a state-based conflict. ACLED records what is apparently the same event as a battle

with a nonstate actor taking territory on May 11, 2011. Similarly, these data sets report from 17 to 31 casualties, with hundreds wounded, in fighting on the western outskirts of Misratah in Zlitan during June 2010. UCDP-GED records the event as state-based conflict on June 8, whereas ACLED records the event as a battle without change of territory on June 10. The details are close, but not identical. Integrating multiple data sets requires making decisions about whether differences mean that entries capture unique events or instead reflect uncertainty or variation in measurement.

The second challenge arises from the fact that conflict event data sets frequently have large numbers of entries. ACLED, UCDP-GED, GTD, and SCAD collectively comprise over 200,000 entries about events in Africa from 1997 to 2016 (the current universe of spatiotemporal overlap across all four data sets). Many of these events co-occur within narrow spatiotemporal windows. Reviewing all the entries by hand and coding which of them match or are unique would be a time-consuming process that is difficult to complete successfully. These problems are compounded by the regular updating and ongoing expansion of the coverage of data sets.

Finally, performing this work manually lacks transparency and adaptability. Researchers may wish to use data that someone else has integrated, but disagree about specific choices made during that process. Reproducing a manual integration effort—whether for purpose of validation or modification—is infeasible if the work was not sufficiently documented.

A New Methodology for Integrating Event Data

A robust methodology that addresses these challenges is a prerequisite for systematic data integration. The primary hurdle in designing such a methodology is specifying the means to determine which entries in multiple data sets capture the same event and which capture unique events.

Building on prior work in data comparison (Silva and Ball 2006; Herzog, Scheuren, and Winkler 2007; Christen 2012; Price, Gohdes, and Ball 2014), we pursue an algorithmic blocking approach for iterative pairwise comparison that is suited for event data sets. The application of this approach relies on the characteristic structure of geo-referenced, date-specific event data sets, in which entries consistently include both location geo-coordinates and time stamps (barring missing information). The underlying presumption of the approach is that entries not coded as occurring in or around the same location, at or around the same time, are unlikely to reflect the same event. Therefore, the set of entries that potentially capture the same event across data sets can plausibly be limited. Further in-depth comparisons are then performed only with the smaller subset of entries that capture spatially and temporally proximate events. What constitutes proximity is not definite and must be specified by the researcher with a given context in mind; we discuss this aspect in more detail below.

Deciding whether proximate entries across multiple data sets concern the same event or unique events entails examination of information on other attributes such as the type, scale, and outcomes of the event and which actors were involved. Implementing such comparison systematically is not always straightforward. Measurement can be incomplete: one data set records information on a particular event attribute, whereas another data set does not (or the information happens to be missing for a given entry). Measurement can also be imprecise: one data set records less specific information on a particular attribute for a given entry than is recorded for a comparable entry in another data set (e.g., an approximate area rather than an exact location). In addition, measurement can be uncertain: the information that one data set records for a particular attribute may deviate, perhaps incidentally or erroneously, from the information in another data set as a function of the respective processes of data generation. In each instance, disparities between entries across data sets need not indicate that those entries capture different events.

These issues are common in conflict event data sets, even with careful data collection protocols and coding schemes. Available measurements of geo-coordinates, time stamps, and other attributes of event data are not always 100 percent exact and accurate. Instead, measurements tend to have an intrinsic “fuzziness.” In other words, a given measurement may be viewed as an approximation—perhaps a close approximation—of the actual value. For purposes of integration, a reasonable solution is not to require exact agreement, such that every attribute of two entries is identical. Rather, delineation of unique from matching entries ought to be approached fundamentally as a matter of degree, allowing for the potential of intrinsic fuzziness in measurement. How fuzziness is reflected in comparisons remains a judgment call in the absence of a definitive empirical basis for adopting particular heuristics.

We designed a data integration protocol that requires the specification of explicit rules governing whether entries are considered to capture unique as versus matching events, operationalizing fuzziness in the process of comparing any event attribute. This protocol represents the core of the MELTT methodology. Below, we walk through the steps involved in the integration of two data sets. In Online Appendix B, we show that the methodology generalizes readily to the integration of more than two data sets and can also be used for data that code episodes, that is, clusters, sequences, and other sets of events spanning multiple temporal units that are coded within the same entry.

The protocol proceeds in three steps. The first step reduces the number of required in-depth comparisons by applying spatiotemporal blocking to identify proximate entries across data sets. The second step performs in-depth comparisons within each set of proximate entries. More than two entries in one data set could potentially match a given entry in another data set. Taking into account the relative extent of correspondence among such entries, the third step then identifies which pair of entries is the most likely match. Figure 2 depicts the protocol in a step-by-step diagram. Each of these steps is described further below.

Step 1: Spatiotemporal Blocking

The starting point is to limit the comparison of entries across data sets to those entries about events that are proximate, that is, co-occur sufficiently close in space and time. Our framework allows for the possibility that entries coding the same event may not always be recorded with identical locations and timestamps across data sets. Formally, we define spatial fuzziness, Δs , as the radius (in kilometers) around a given entry within which another entry must fall to qualify as proximate. Similarly, temporal fuzziness, Δt , is defined as the number of time units (typically in days) before or after a given entry within which another entry must fall to qualify as proximate. The researcher must specify both Δs and Δt prior to integration.⁵

The protocol iterates through each entry in the first data set, checking whether or not any entries in a second data set fall within the ranges of spatial and temporal fuzziness specified by the researcher (Figure 2, left panel). When evaluating spatial proximity, the procedure uses great circle distances that correctly account for geographic distances at all scales. Temporal iteration proceeds in increments of the order of the units of Δt , using an efficient algorithmic approach (see Online Appendix).⁶ Entries for events that are found to co-occur within these windows are flagged as potentially coding the same event.

Step 2: In-depth Comparison of Entries

Among proximate entries, the correspondence of event attributes other than location and timing can be ambiguous. Each event data set can be distinctive in terms of the variables included, the reported values for these variables, the coding rules, and/or the sources of raw information on which coding is based. Such disparities complicate the process of identifying which entries code the same event (or not), ruling out straightforward, apples-to-apples comparisons across data sets.

Our solution is to require the researcher to define a taxonomy that formalizes the mapping of correspondences between values for a given event attribute in one data set and values for a relevant event attribute in another data set. This mapping does not necessitate exact and complete correspondence, which may be impossible for a particular attribute or constituent value. A taxonomy can be specified with multiple levels, mapping broader categories of values for a given attribute in one data set as being substantively equivalent to an attribute or values of an attribute in another data set. This option increases the flexibility in defining which entries qualify as corresponding entries and gauging the degree to which they correspond. The most fine-grained level of such a multilevel taxonomy would reflect the original coding of event attributes across data sets. Successive levels of the taxonomy subsume those values under broader categories, defined by the researcher in accordance with the qualitative understanding of each event attribute. The broader the category, the less specific the distinctions in characterizing events, but also the greater the prospects of identifying corresponding entries across data sets.

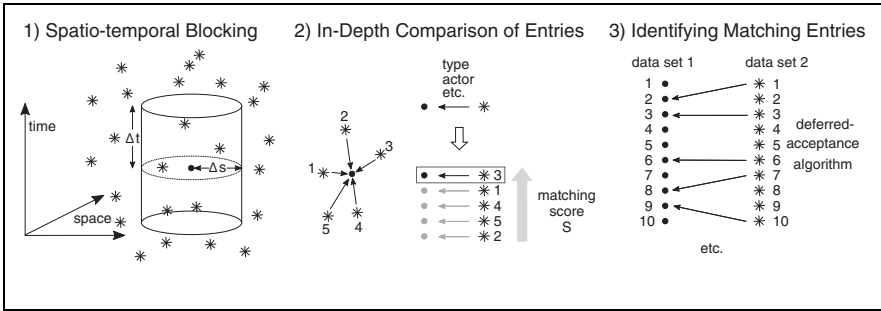


Figure 2. Illustration of the data integration protocol.

The taxonomy concept is best understood via an illustrative example using event-type attributes.⁷ Consider an event coded as a “Bombing/Explosion” in GTD. This event may be considered more broadly as a case of “strategic destruction” and at the broadest level as a case of “violent attack.” Similarly, an event coded as a “Battle-No change of territory” in ACLED may be considered more broadly as a “territorial dispute” and at the broadest level as a “violent attack.” In this example, the entries in the two data sets differ in both the original coding and for the broader level, but at the broadest level the entries correspond. Assuming the entries also coincide spatially and temporally, they might actually describe the same event—albeit with a significant degree of uncertainty about this assessment, since the correspondence for event type is only at the broadest level of the taxonomy.

This illustration highlights the flexibility gained by using a taxonomy to accommodate fuzziness in measurement of other attributes of events. If exact or close correspondence is required, the two entries would never be considered as coding the same event. If, however, the precision of measurement varies, the labeling of measurements differs, or the accuracy of measurement is uncertain, evaluating correspondence at a broader level may be reasonable.

In-depth comparison of event attributes (Figure 2, middle) is limited to separate subsets of proximate entries identified in the first step of the protocol. The procedure relies on a dynamic matching logic, iterating through each level of the taxonomy and finding the most fine-grained level on which two entries correspond, if at all. Some entries may correspond at the level of original values, while other entries may just correspond at the level of a broader category. Only proximate pairs of entries that correspond—at least on the broadest taxonomy level—for every event attribute included in the comparison are considered as potentially coding the same event.

The degree to which corresponding entries align with regard to m event attributes is formalized by the matching score S , calculated as follows:

$$S = \frac{1}{m} \sum_{i=1}^m w_i \frac{l_i}{L_i}. \tag{1}$$

In this equation, l_i is the taxonomy level at which the two entries correspond for attribute i ; $l_i \in \{0, 1, 2, \dots\}$, where $l_i = 0$ refers to the base category for each attribute. L_i is the “depth” of the taxonomy for a given attribute i ,⁸ and w_i is the weight assigned to that attribute. Weights must be chosen such that $\frac{1}{m} \sum_{i=1}^m w_i = 1$. The default assumption is that all attributes are weighted equally. A researcher may, however, choose to assign certain attributes greater weight than others. Among the rationales for a greater weight are a belief that an attribute is more fundamental to the measurement of a phenomenon, greater faith in how an attribute is measured by the data sets, or a clearer, more compelling mapping of values within the taxonomy.

The metric S functions as a penalty score. A value of 0 indicates that entries correspond at the most fine-grained taxonomy level—the categories originally coded in the data sets being compared—for each event attribute. The broader the taxonomy level at which entries correspond, the larger the value of S . If proximate entries correspond at the broadest taxonomy levels for each other event attribute, S takes on the maximum value of 1. The score is used in the next step to rank pairs of proximate entries with corresponding attributes and establish which of those pairs is most likely an instance of matching entries across data sets.⁹

Step 3: Identifying Matching Entries

When comparing two data sets with overlapping coverage, a given combination of Δ_s , Δ_t , and event attribute taxonomies could yield instances of multiple entries in one data set that correspond to a proximate entry in another data set. Not all of those pairs of entries will necessarily correspond at the same level of every taxonomy. By implication, the pairs can differ in terms of their matching score S . A logical rule views pairs of entries that correspond most closely as the likeliest to be capturing the same event.

Our protocol strictly assumes that each entry in one data set should at most match one entry in another data set. This assumption proceeds from the expectation that each entry in each data set refers to a distinct event. What constitutes an entry could differ across data sets. For example, one data set may record each incident of conflictual activity as a separate entry (and implicitly, a distinct event), whereas another data set codes related incidents occurring in the same place and on the same day as one event. Our framework assumes that the researcher has already preprocessed the data to ensure that entries across data sets are indeed comparable in these respects.¹⁰

Isolation of matching entries requires identifying—among all corresponding entries—the pairs of entries that are maximally similar. This challenge is an instance of the “stable marriage” problem, which is well known in computer science and has also been extensively studied in economics in the context of assignment games (Shapley and Shubik 1972). The standard solution is a “deferred-acceptance algorithm” (Gale and Shapley 1962), which we employ as part of the protocol. The algorithm iterates through all pairs of corresponding entries. For those pairs that are not completely distinctive (i.e., share no entries with any other proximate pair), the algorithm defers pairs with inferior matching

scores and accepts pairs with superior matching scores. The process continues until changing any pairing of entries does not further improve the quality of matches for any of the pairings (Figure 2, right).¹¹

Given the results generated by this algorithm, integration is straightforward. First, all entries that were not identified as corresponding with any other entry (i.e., the unique events) are retained. Second, each pair of matching entries is merged into a single entry and added to the integrated data. As a default, the protocol preserves all information associated with every entry. In particular, merged entries carry the information of both original entries. The protocol also documents pairs of entries that were found to be proximate with corresponding information, but not ultimately deemed to be matching. Among the information preserved for each of these pairs is the matching score. This information supplies a means of exploring the overlap between data sets in greater depth, including through qualitative research that examines particular pairs of matching entries and their associated attributes.

Using MELTT

The MELTT methodology is flexible, allowing the researcher to vary the spatio-temporal window for determining proximity of events and to develop multiple taxonomies that vary in depth (i.e., the number of levels). This flexibility means that researchers have a number of decisions to make when using MELTT to integrate event data sets. We have developed the open-source `meltt` R software package that implements the full protocol and is freely available from CRAN. The package documentation and the Online Appendix accompanying this article provide further guidance to researchers in making these decisions.

Validation

An important consideration is the reliability of the methodology: does MELTT consistently generate accurate results? As a test, we demonstrate and formally validate the methodology in controlled settings, using synthetic data designed to correspond to different stylized scenarios that mirror situations commonly encountered in empirical analyses. Specifically, we systematically explore how the results are affected by two sets of factors. One set of factors concerns parameters that the researcher has to specify: different values of Δs and Δt and definitions of taxonomies. Another set of factors concerns properties of the data, for example, the number of data sets to be integrated, the nature of geospatial imprecision in measurement, and the degree of clustering of events. By virtue of having constructed the synthetic test data ourselves, we know with certainty the unique and matching entries, which afford a clear benchmark for evaluation of the protocol's performance.¹²

In each of our stylized scenarios, the performance of the methodology can be gauged by comparing two diagnostic indicators—the true positive rate (TPR, or the share of matching entries in the synthetic data that MELTT identifies correctly) and

Table 1. Systematic Tests of the Performance of the Data Integration Protocol.

Variation	TPR	FPR
MELTT protocol specification		
Order of data set integration	<i>No effect</i>	<i>No effect</i>
Increasing Δs	<i>No effect</i>	<i>No effect</i>
Increasing Δt	<i>No effect</i>	<i>No effect</i>
Increasing Δs and Δt	<i>No effect</i>	<i>No effect</i>
Data structure		
Vary number of data sets	<i>No effect</i>	<i>No effect</i>
Vary number of other event attributes	<i>No effect</i>	<i>No effect</i>
Increase depth of taxonomy	<i>No effect</i>	<i>No effect</i>
Increase density of observations	<i>No effect</i>	Increases weakly
Increase density of observations with a shallow taxonomy	Decreases weakly	Increases
Increase proportion of imprecise entries	Decreases	<i>No effect</i>

Note: FPR = false positive rate; MELTT = Matching Event Data by Location, Time and Type; TPR = true positive rate.

the false positive rate (FPR, or the share of unique entries in the synthetic data that MELTT identifies incorrectly as matching another entry). A TPR below 1 indicates that the protocol fails to recognize certain matching entries, while a FPR above 0 indicates that the protocol fails to identify certain unique entries.

Table 1 summarizes the results of our Monte Carlo analysis.¹³ We find that most of the scenarios affect neither the TPR nor the FPR to a substantial extent. The only exceptions involve certain structural characteristics of the data: increasing spatio-temporal clustering of events and increasing the proportion of imprecisely georeferenced data.¹⁴

Increasing the spatiotemporal density of events in the data alone does not substantially affect the ability of the protocol to recover matching entries. Higher density does, however, weakly decrease the TPR and increase the FPR when using a shallow taxonomy, that is, one with a limited number of levels and relatively broad categories. Intuitively, the rise in the FPR occurs because the higher the density of events, the more likely entries with similar attributes are spatiotemporally proximate, even though they code unique events. In such a circumstance, a deeper taxonomy—that is, one with more levels of differentiated categories—performs significantly better because it can more accurately discriminate which proximate entries truly match, especially among a large set of pairs.

Similarly, the TPR also decreases when the proportion of imprecise entries in the data increases. A primary reason is that matching entries are not even compared, since they are coded to centroid locations that are further apart than the specified window defining spatial proximity. In Online Appendix C, we demonstrate that this issue can be resolved by increasing the spatial fuzziness Δs . Doing so comes with a

trade-off: a larger spatial window means that more geographically distant entries are treated as proximate and thus become candidates to be considered as matches, increasing the chance of false positives. This trade-off is mitigated as the depth and quality of the taxonomies increase.¹⁵

The tests with synthetic data confirm that MELTT has robust performance in successfully integrating data sets. We also demonstrated that specific input choices made by a researcher can affect the relative performance of the protocol. When using data that are either highly clustered or coded with a low degree of spatial precision, the reliability of the output that MELTT generates can be enhanced by appropriate choices in the specification of taxonomies to discriminate between potentially matching entries. What constitutes proximity of events and how event attributes map across data sets must always be informed by strong understanding of the empirical context and the quality and characteristics of the data sets that are being integrated.

Conflict Dynamics in Africa: An Empirical Application

We posited earlier that a key advantage of integrated data could be to render a more complete picture of contentious activity, while still maintaining a high degree of reliability in measurement. To demonstrate these claims, we start with an examination of conflict dynamics in Nigeria during 2011, then conduct a further validation of the integration of data for South Sudan during 2015 and Libya during 2014. In all these instances, we integrate conflict event data from ACLED, UCDP-GED, GTD, and SCAD. By focusing on single years from this small selection of countries, the demonstration is limited enough in scale to enable careful qualitative review. At the same time, relevant diversity in data set properties is observed within and across the cases.

Integrating Conflict Event Data for Nigeria 2011

In 2011, Nigeria experienced an assortment of different types of contentious activity. Some of the events are characteristic of a civil war context, involving rebel activity against the state and civilian communities. Other events (e.g., election-related violence) may fall outside the scope of civil war. In March, April, and May 2011, the process to elect Nigeria's president was inundated by violence. After southern Christian minority leader Goodluck Jonathan was declared the winner, over 800 people were killed and tens of thousands displaced in the country's Muslim-majority North during three days of violence. Sectarian violence in the wake of the election was soon overshadowed by a series of terrorist attacks led by Boko Haram, which bombed military and police targets within hours of the inauguration and throughout the summer. Boko Haram's strategy evolved throughout 2011 from local assaults to Nigeria's first-ever suicide bombing in August 2011. A string of bombings on Christmas Day, carried out by Boko Haram, resulted in the declaration of a state of emergency on December 31, 2011.

Collectively, ACLED, UCDP-GED, GTD, and SCAD report that a total of 904 events occurred in Nigeria during 2011. About three-fifth of these events are spatiotemporally clustered, that is, they co-occur within three kilometers and one day of an event reported in at least one other data set.

Data Integration. We created three taxonomies that generalize variables present in the ACLED, UCDP-GED, GTD, and SCAD data sets, capturing key attributes of conflict events. The first taxonomy defines how each *type* of event coded in any of the four data sets maps onto any other type. The second taxonomy provides a comprehensive mapping of all *actors* coded across the data sets into broader categories (government, political opposition group, or civilians). The third taxonomy reflects the degree of *geo-precision* in the coding of event locations, mapping specific precision codes of each data set onto common broader categories ranging from precise (exact town, area, city or village) to imprecise (unknown subnational location). The taxonomy tables and accompanying descriptions applicable to the case of Nigeria are provided in Online Appendix D.

To integrate the data sets, we used these taxonomies together with spatiotemporal fuzziness parameters of $\Delta s = 3$ km and $\Delta t = 1$ day. These choices arise from comparing integration results for a range of values. The chosen values are large enough to allow for coding uncertainty, but also small enough to mitigate against false identification of matching entries.¹⁶

Figure 3 summarizes the data after integration. We find that 23.9 percent of the entries across the four data sets have matching entries, that is, they are recorded in two or more of the four data sets.¹⁷ Merging the matching entries reduces the integrated data to entries covering 688 events versus the original total of 904 events reported across the four data sets.

The manner in which the protocol distinguishes between matching and unique entries can be illustrated by considering two concrete examples. In the first example, all four data sets report an event on August 19, 2011, in Maiduguri, a northern Nigerian city, with the entries coding the event location to within two kilometers of one another. The entries agree that an unknown number of suspected Boko Haram militants broke into the house of a police officer, and ultimately three policemen and one civilian were shot and killed. The only distinction among the entries is that ACLED labels the event as “battle-no change of territory,” UCDP-GED as “state-based violence,” GTD as “armed assault,” and SCAD as “anti-government violence.” With an appropriate taxonomy that formalizes the mapping of event types across data sets, the protocol correctly finds a correspondence among these entries and deems them as capturing the same event, despite differences in the original values for the event types. In the second example, GTD reports an assault on a prison guard at his home, while SCAD reports a strike by the cell phone retailers union in response to the killing of their leader, both occurring in Maiduguri on October 13, 2011. The entries record these events as taking place within two kilometers of one another on the same date. Leveraging the event-type taxonomy, the protocol

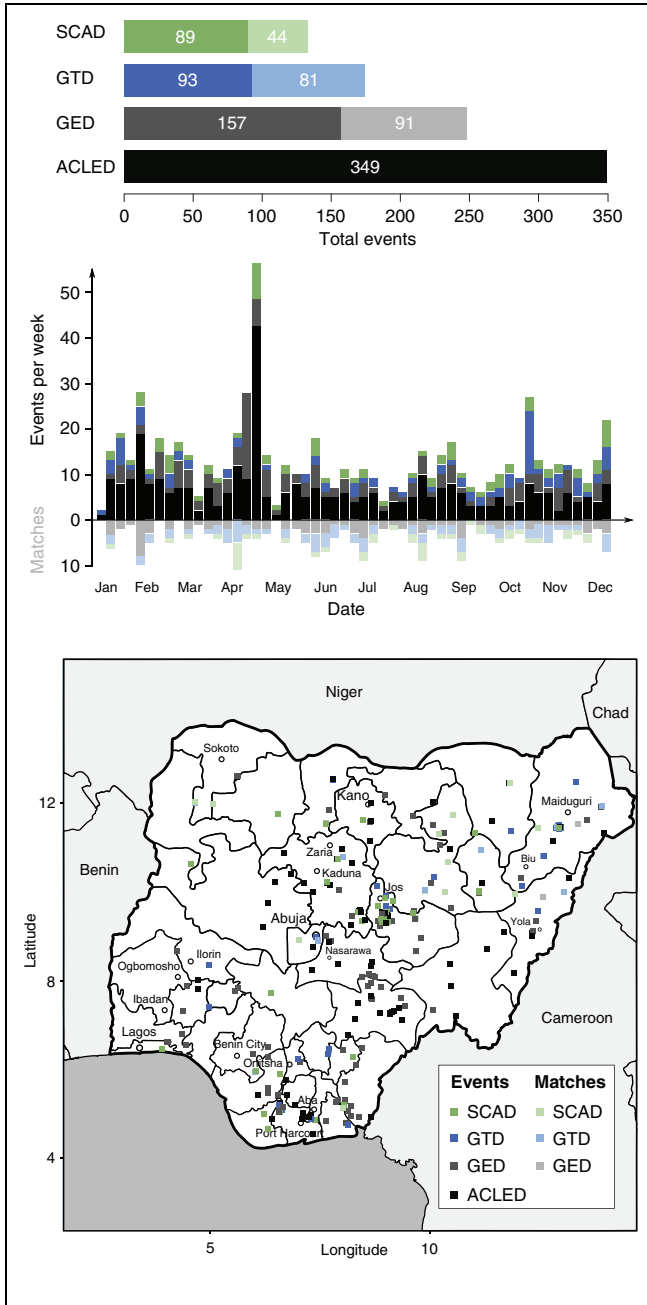


Figure 3. Unique and matching entries for events reported for Nigeria in 2011.

Table 2. Performance of the Data Integration Protocol for Nigeria 2011.

	Indicator	Entries	Percentage of Sample	Total Percentage
Correct	True positives	344	38.1	92.2
	True negatives	489	54.1	
Incorrect	False positives	16	1.8	7.8
	False negatives	55	6.0	

correctly identifies that these two entries do not correspond: the assault is a violent event, whereas the strike is a nonviolent event. Thus, the entries are evaluated as recording unique events.

To judge the reliability of the MELTT protocol, we manually conducted an in-depth qualitative review of the output for Nigeria 2011.¹⁸ We classified entries identified to be matching as either true positives (containing information matching at least one other entry) or false positives (containing information about seemingly different events). Conversely, we classified entries that the protocol identified as unique as either true negatives (confirmed to have no matching entry in any of the other three data sets) or false negatives (should have been matched to at least one entry in another data set, but was not). Table 2 summarizes the performance statistics. For over 92 percent of entries across the four data sets, the protocol accurately determined whether or not matching entries exist among the other data sets. For less than 8 percent of the entries, the procedure did not correctly adjudicate between unique and matching entries. The results of this review suggest that the MELTT protocol is highly reliable.

Overlapping Coverage. Next, we examine what integrated data reveal about the degree of unique and overlapping coverage across the four data sets. In the process, we consider how those aspects of coverage relate to event type, time period, and geographic region. The goal is to identify where overlapping coverage is most prevalent. The first column in Table 3 provides the counts of events in the integrated data that are unique to one data set. The remaining columns present the counts of events coded in two, three, or all four of the data sets. The results show that 79 percent of entries capture events unique to one data set. The remaining 21 percent of entries in the integrated data reflect events recorded in at least two of the four data sets. The greatest overlap among the four data sets relates to violent events, which is to be expected. ACLED and UCDP-GED aim to capture types of violent events that should be substantially overlapping. Meanwhile, the overlap of GTD events with these two data sets confirms that measures of terrorist and civil war activity are not mutually exclusive on a consistent basis.

Of the four data sets, only ACLED and SCAD cover protests and riots. Unexpectedly, the protocol finds that all 92 entries in the two data sets that reflect protests (58 in ACLED, 34 in SCAD) are unique, that is, no matches

Table 3. Summary of Overlap for Nigeria 2011.

Dimensions	Unique Coverage	Overlapping Coverage		
	One Data set	Two Data sets	Three Data sets	Four Data sets
Event type				
Violent attacks	319	55	29	7
Violence against civilians	118	24	22	3
Protests	92	0	0	0
Riots	14	5	0	0
Time period				
January–March	152	21	8	3
April–June	168	18	19	1
July–September	91	20	12	4
October–December	133	23	13	2
Region				
North Central	190	13	7	1
North East	164	62	41	9
North West	86	7	2	0
South East	10	0	0	0
South	66	2	1	0
South West	27	0	0	0
Total	543	84	51	10

across the data sets could be identified. We manually reviewed these results and confirmed their validity.¹⁹ This finding indicates that relying on just ACLED or SCAD to study contentious activity in Nigeria during 2011 would result in undercounting the prevalence of protests.

The total number of events varies by the time of year, yet the proportion of overlapping entries remains relatively similar across time periods. Overall, about 20 percent of events are overlapping between at least two of the data sets for each period, with a somewhat larger degree of overlap from July through September. These results suggest that timing does not have a systematic effect on whether an event is recorded in either one or multiple data sets.

An examination of the results by region indicates that both the overall coverage and the degree of overlap are greatest in the North East region of the country, which most directly experienced the impact of the rising Boko Haram insurgency. With the corresponding increase in violent attacks, the high extent of regional overlap simply tracks the overlap in the coding of violent events, as was discussed above.

Degree of Missingness. The fact that the integrated data feature both unique and matched entries form the basis for using statistical techniques to extrapolate the degree of missingness across the four data sets. Missingness is an important issue for the analysis of any data. In regard to conflict event data, a basic repercussion of

missingness can be understating the prevalence and miscalculating distributions of particular phenomena and/or their associated attributes. If one further relies on incomplete data when evaluating relationships, the results can be systematically biased.

Intuitively, the greater the extent of overlap among several independently compiled data sets, the more certain we can be that they cover a larger share of the events that actually occurred. This intuition is formalized in techniques of MSE (cf. Manrique-Vallier, Price, and Gohdes 2013), which can be applied if three or more data sets with overlapping coverage exist. In practice, these techniques have been used effectively to estimate the degree of missingness of data on atrocities in civil conflicts (cf. Ball, Kobrak, and Spierer 1999; Silva and Ball 2006; Price, Gohdes, and Ball 2014). The ability to apply MSE to extrapolate to a “true” total number of events rests on two assumptions. First, the type of event should be conceptually consistent across the data sets used for extrapolation. In our illustrative analysis, we opt to focus exclusively on violent events, which all four data sets encompass (albeit with some different labels). Second, extrapolation hinges on the assumption that the data sets do not use the exact same sources of raw information. In our analysis, all four data sets rely on media reporting, but they augment their coding with distinct additional nonmedia sources.²⁰

The literature on possible biases in conflict event data (cf. Donnay and Filimonov 2014; Weidmann 2015) argues that quality of coverage likely varies across types, conflict periods, and geographical regions. The examination of overlapping coverage described above suggests that these dimensions affect the integration of the data on Nigeria 2011 as well. Following best practices laid out by Manrique-Vallier, Price, and Gohdes (2013), we stratify the data along those dimensions (as in Table 3) and apply MSE with respect to each dimension.²¹ The results suggest that the four data sets capture about 35 percent to 40 percent of all violent events. In other words, based on the patterns of matching (and unique) entries across data sets, we can reasonably anticipate that for every recorded violent event, at least one other such event was not captured.²²

This missingness is not necessarily an inherent problem for quantitative inferences, so long as the recorded events are a random sample of the actual universe of violent events in Nigeria during 2011. Yet we observe some variation in extrapolated totals, depending on how we stratify the data. The estimated coverage is lowest when stratifying by event type (33 percent), increases when stratifying by time period (37 percent), and reaches the highest rate when stratifying by region (41 percent).²³ These results indicate disparities in coverage of different types of events, time periods, and regions, which could affect certain quantitative analyses using these data.

Assessing Performance on Alternative Cases

The analyses of Nigeria 2011 substantiate that MELTT is accurate and efficient at discriminating between unique and matching entries across conflict event data sets. A possible concern, however, is that the integration protocol is sensitive to data

quality, in conjunction with spatiotemporal windows specified by the researcher. If the quality of data for one country-year is relatively poor, integration may increase the rates of false positives or false negatives. The choice of windows could then exacerbate the number and unreliability of identified matches. To examine these potential issues associated with data quality, we assess MELTT's performance in two additional country-years.

We purposely sought cases where the data quality differs significantly from that of Nigeria 2011. To do so, we generate an overall measure of the degree of geo-precision in the coding of events for country-years with overlapping coverage in ACLED, UCDP-GED, GTD, and SCAD. This measure uses the geo-precision variable associated with the location coding of events in each data set, computing the share of observations for which the coding reflects the actual location of an event. The ability to precisely geo-code an event is often closely correlated with the extent to which other event attributes are known. Thus, our measure also proxies for general quality of reporting.

Some entries in the data are *precise*, providing coordinates at or nearby the actual location of the event. Other entries are *imprecise*, providing coordinates at the centroid of the first-order administrative division or the entire country. The proportion of precisely coded events for Nigeria 2011 (83 percent) is close to the average proportion observed across all country-years for which the data sets offer overlapping coverage. Therefore, we select two additional country-years that represent the ends of the precision spectrum. South Sudan 2015 (58 percent) is the country-year with the lowest rate of precise entries, while Libya 2014 (95 percent) is the country-year with the highest rate of precise entries.²⁴

Applying the same event type, actor,²⁵ and precision taxonomy schema used for the Nigeria 2011 data, we separately integrate data for South Sudan 2015 and Libya 2014, in each case using two spatiotemporal windows: three kilometers and one day (identical to the Nigeria 2011 integration) and fifty kilometers and one day. We add the additional specification using the fifty-kilometer window for two reasons. First, past research undertaking a similar integration task manually used this larger spatial window when considering potential matches (Weidmann 2015). Second, a likely effect of imprecision in the coding of event locations is the inflation of distances between potentially matching entries.

Validating the output of MELTT is important in terms of refining assumptions and identifying any fundamental issues. A crucial consideration, however, is that manual validation of each and every entry—as performed for Nigeria 2011—proves to be an exacting, time-consuming process that becomes increasingly less feasible as the number of entries increases. Libya 2014 and South Sudan 2015 collectively total more than 3,000 entries across the data sets. To facilitate evaluation of the performance of MELTT on larger-scale cases such as these, we designed a computer-assisted validation protocol.²⁶ This protocol starts by randomly sampling pairs of entries from the set of matching entries identified by MELTT. For each such pair, two additional proximate entries are drawn as controls. The controls are entries

Table 4. Integration Results for Libya 2014 and South Sudan 2015.

Country-year	Window	Total Events	Matches	Accuracy (%)
Libya 2014	3 km, 1 day	2,382	327 (13.7%)	78.3
	50 km, 1 day		716 (30.0%)	69.9
South Sudan 2015	3 km, 1 day	1,027	46 (4.4%)	90.8
	50 km, 1 day		71 (6.9%)	86.4

capturing events that occurred within the relevant spatiotemporal proximity of a pair, but were not identified by the protocol as matches to either entry in this pair.

In a simple interactive interface, the researcher is then shown descriptive information of one of the entries MELTT has matched (dubbed the “main entry”), to be compared to analogous descriptive information for a set of three other entries (“candidate entries”)—the matching entry identified by MELTT, as well as the two proximate but nonmatching entries that are randomly selected controls. For every such set, the researcher has to select the candidate entry that she or he deems most likely to match the main entry based on the descriptive information provided. Behind the scenes, the protocol automatically assesses whether the match or one of the controls was chosen, that is, confirming or rejecting the output of the automated approach. Performing this comparison for a representative set of entries provides an straightforward, rigorous, reasonable, and flexible way to generate performance statistics (i.e., true/false positive/negative rates) comparable to those of a full comparison, without the need to review every entry.²⁷

Table 4 reports the results from utilizing the computer-assisted validation protocol to assess the performance of MELTT in integrating the usual four data sets for Libya 2014 and South Sudan 2015.²⁸ The results show that the performance for South Sudan 2015 is on par with Nigeria 2011. The implication is that MELTT performs well across different levels of geographic precision—recall that a much larger proportion of entries for South Sudan 2015 have imprecise geographic locations, relative to Nigeria 2011. The performance for Libya 2014 is good as well, although the validation protocol finds lower accuracy rates than in the other two cases. A possible contributing reason is that most activity in Libya at this time occurred in or around a limited number of population centers. Consequently, events tend to be highly clustered spatially. As shown in the analysis using synthetic data presented earlier (see Table 1), the more the events cluster in close proximity, the greater the need for deeper taxonomies that can sufficiently discriminate between unique and matching entries. In order to achieve accuracy for Libya 2014 that is on par with what was observed for the other cases, the taxonomies for this specific case would need to be refined further.

With both South Sudan 2015 and Libya 2014, expanding the spatiotemporal window consistently yielded an increase in the number of matching entries across data sets, accompanied by a drop in the rate of accuracy. These results highlight the trade-off intrinsic to specifying wider spatial windows: this decision allows for

identifying matching entries that happen to be coded as spatially distant due to coding imprecision, but doing so comes at the risk of overidentifying entries as matches that actually capture unique events.

Conclusion

In this article, we presented a novel automated protocol that allows for systematic, transparent, reproducible, and flexible integration of conflict event data sets. Our quantitative and qualitative analyses demonstrate that the protocol we developed is accurate and efficient in identifying entries that match across multiple event data sets, while also distinguishing entries that capture unique events. The Monte Carlo simulations we performed, using synthetic data, show the protocol is robust to a range of scenarios that are relevant for empirical applications.

The detailed qualitative analysis of the integration of data on conflict events in Nigeria during 2011 reveals that the procedure has a high accuracy rate even for a context in which multiple data sets record numerous events that cluster in space and time. Subsequent analyses using MSE techniques suggest that the violent events recorded across leading conflict event data sets represent only a subset of the full universe of activity in Nigeria, with some biases in the extent of coverage. Finally, the analyses of Libya 2014 and South Sudan 2015 show that MELTT also performs well in contexts of varying data quality.

Our analyses indicate that MELTT serves as a valuable tool to enhance the quantitative study of conflict dynamics. Researchers interested in particular phenomena, such as protests or civilian targeting, are able to obtain a more accurate picture by integrating multiple data sets that provide complementary information. Researchers interested in the relationship between phenomena typically covered in different data sets—such as protest and terrorism—can use integrated data that cover a more diverse assortment of event types. Regardless of the specific topical interest, using MELTT for data integration provides assurances that the resulting data, as compared to naive poolings of multiple data sets, mitigate the risk of including duplicate entries about the same events.

Our empirical demonstrations focused on integrating four of the most commonly used conflict event data sets. Researchers can also opt to use our protocol to integrate any of the growing array of other data sets with spatiotemporal information on conflict events. In addition, the methodology has the potential to contribute to data generation and to the evaluation of the quality of different data-generating processes. This includes applications within institutional data collection efforts such as ACLED and UCDP-GED, which already rely on data from multiple primary data sources (news media, local NGO reports, etc.) that can have overlapping coverage. MELTT could also be used to compare machine- to human-coded data sets. If entries from the former do not consistently match to entries from the latter, this may be evidence that the machine-coded data set is generating a large number of false positives, presuming a degree of confidence in the human-coded data. We

may not know definitively which of these entries are false positives, absent a careful qualitative validation. Regardless, the extent to which human- and machine-coded data align could be highly informative for the design of data development initiatives.

Using an automated data integration procedure allows researchers to integrate data rapidly on scales that were previously infeasible. A further benefit is that the procedure is much more transparent and replicable than manual data integration. As a best practice, we recommend that researchers who use MELTT in published work should make their taxonomies and other input specifications public, alongside the usual replication materials. In this way, other researchers can review and evaluate choices related to the integration of data sets. Those who disagree with these choices can opt to modify them and rerun analysis to determine whether or not any findings are sensitive to how integration was implemented.

The protocol is well suited to deal with the regular updating and expanding coverage of conflict event data sets. Many of these event data sets have at least an annual release, which can extend the spatial and temporal coverage of the data, as well as reflect certain historical revisions. Consequently, a data set integration performed at a given point may already be outdated by the time related studies are published. Using MELTT, any researcher can easily and quickly rerun the integration to reflect the most current data available, as well as undertake replication analyses with greater ease when additional data become available.

We reiterate that integration is not always desirable, nor does it necessarily guarantee improvements in the empirical measurement of conflict. Integrated data sets may be no better at capturing a true record of events as compared to what a single, high-quality data set may capture. In fact, issues of data bias are likely to carry over to an integrated data set—especially if the same systematic biases are present in many or all of the data sets that are integrated. Thus, using integrated data does not free the researcher from the responsibility to carefully consider the quality of individual data sets, including their sources of information, coding protocols, and validation checks. Whether integrated data should be used is initially a theoretical question based on the research topic and an understanding of the strengths and weaknesses of relevant data sets, some of which may only become clear *after* integration has clarified the degree of overlap among them, or a lack thereof. Thus, integration could have a place in most analyses of conflict event data—if only to bolster the argument for using a specific data set to tackle the empirical research question at hand.

Authors' Note

Authors are listed in order of contribution. Previous versions of this article were presented at conferences organized by the Midwest Political Science Association, the Peace Science Society (International), and the Society for Political Methodology and seminars held at the Peace Research Institute Oslo, Uppsala University, the University of Maryland, and the University of Konstanz. The open source `meltt` R software implementing the approach presented in this

article is available at <https://cran.r-project.org/package=meltt>. The views and findings expressed in this article are those of the authors alone.

Acknowledgments

The authors thank Juan Montoya-Bayardo, Taehee Kim, Karim Makkady, and Rithvik Yarlagadda for their research assistance. The authors are also grateful to Gary LaFree and Andrew Linke for their support throughout the development of MELTT. The authors appreciate the helpful feedback from the discussants and other participants at workshops and presentations, as well as from Kristine Eck, Kristian Skrede Gleditsch, and two anonymous reviewers.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Work on this article was funded through grants to the University of Maryland from (1) the U.S. National Science Foundation (#SES-1343123) and (2) the Air Force Office of Scientific Research (#FA9550-14-1-0353) under the Minerva Initiative of the U.S. Department of Defense.

Supplemental Material

Supplementary material for this article is available online.

Notes

1. Studies of subnational spatiotemporal patterns of conflict events have confirmed several influential factors, including population (Raleigh and Hegre 2009), demographics (Weidmann, Rød, and Cederman 2010), elements of physical geography such as terrain and the distance from the capital (Buhaug, Gates, and Lujala 2009), rainfall (Hendrix and Salehyan 2012), elections (Salehyan and Linebarger 2015), peacekeeping (Hultman, Kathman, and Shannon 2013), and battlefield losses (Wood 2014).
2. The recently launched xSub project (Zhukov, Davenport, and Kostyuk 2017) draws on geolocated event data from 21 distinct conflict data sets. We expect that MELTT is similarly useful for integrating data from these data sets and others.
3. In a detailed comparison of Uppsala Conflict Data Project-Georeferenced Event Data (UCDP-GED) and Armed Conflict Location and Event Data (ACLED), Eck (2012) shows that the number of events reported in each data set varies widely across countries for periods on which both have coverage. Generally, ACLED reports more violent events than UCDP-GED. For some cases (e.g., Algeria), however, UCDP-GED contains more events.
4. The use of different geo-location techniques across data sets can also yield different coordinates for the same location.

5. For example, certain entries may be coded with a significant degree of spatial imprecision such as at the administrative centroid or country level. In that case, the researcher could calculate the distance between those imprecise locations and the array of exact point locations, then set the spatial window such that plausibly matching entries would be considered proximate.
6. See Online Appendix B.2 for details. For event data, the units of Δt would typically be days, but the protocol also generalizes to units of years, months, hours, or seconds.
7. Event-type attributes used in this illustration are “attacktype1” for GTD and “EVENT_TYPE” for ACLED.
8. L_i reflects the number of taxonomy levels beyond the base category for every attribute i . When a taxonomy only contains the base category, L defaults to 1.
9. The protocol also allows for static matching at a prespecified taxonomy level (as compared to flexible matching at any level). In this instance, the matching score S is irrelevant since entries either do or do not match at the specified taxonomy level.
10. We relax the assumption of one-to-one matches only when comparing entries about standalone events to entries about episodes (see Online Appendix B.4 for details).
11. When two or more corresponding entries have the same matching score, the algorithm randomly breaks ties among those pairs. This rule is necessary to ensure the convergence of the process to distinctive matching pairs.
12. For details on the synthetic test data and its coding, please refer to Online Appendix C.1; information on replicating our analysis using the synthetic data may be found in Online Appendix A.
13. “No effect” means that the true positive rate remains close or equal to 1 and the false positive rate close or equal to 0.
14. Please refer to Online Appendix C.2 for graphical illustrations of both scenarios.
15. We discuss this trade-off in detail in Online Appendix C.3.
16. See Online Appendix D.3 for sensitivity analyses. To address potential concerns over geo-spatial imprecision, we also integrated the Nigeria 2011 data using a much larger fifty-kilometer window. We found that overall accuracy decreased because a larger spatial window contributed to a greater number of geographically distant events being compared, resulting in more false positive matches.
17. Our graphical illustrations treat ACLED—the data set with the most reported events for Nigeria in 2011—as the basis for comparison. This choice only affects the graphical representation, with no impact on which entries are matched.
18. Details on the manual validation procedure can be found in Online Appendices D.4 and D.5.
19. We concluded that although some entries may describe the same event, not enough information is available to be certain. Details of the manual comparison may be found in Online Appendix D.6.
20. For example, ACLED and UCDP-GED rely on additional on-the-ground reporting by (local) human rights organizations or nongovernmental organizations whenever available.
21. For further details on the method and analysis, refer to Online Appendix D.8.
22. The work of Krüger and Lum (2015) finds similar degrees of missingness in conflict event data for Kosovo 1999, corroborated by validating the results obtained using multiple systems estimation techniques against reliable ground-truth data.

23. Estimates vary more widely when considering individual violent event types, time periods, or regions. See Online Appendix D.8 for a detailed breakdown.
24. See Online Appendix D.9 for further details on how these two cases were selected.
25. Lists of actors are adjusted to reflect those active in the respective country-years.
26. The computer-assisted validation protocol has been implemented as part of the `meltt R` package.
27. See Online Appendix B.1 for a full description of the computer-assisted validation protocol and its interface.
28. The validation was performed by two research assistants, with a subset further cross-validated by the authors. The intercoder reliability for each country-year window was about 80 percent. Accuracy scores reported in Table 4 are the average of the results from the validation by the two research assistants. See Online Appendix D.9 for further description of the validation. We also used this protocol to validate the output for Nigeria 2011 and found similar performance statistics to those reported in Table 2.

References

- Ball, Patrick, Paul Kobrak, and Herbert F. Spirer. 1999. *State Violence in Guatemala, 1960–1996: A Quantitative Reflection*. Washington, DC: AAAS.
- Buhaug, Halvard, Scott Gates, and Päivi Lujala. 2009. “Geography, Rebel Capability, and the Duration of Civil Conflict.” *Journal of Conflict Resolution* 53 (4): 544–69.
- Christen, Peter. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Heidelberg, Germany: Springer.
- Donnay, Karsten, and Vladimir Filimonov. 2014. “Views to a War: Systematic Differences in Media and Military Reporting of the War in Iraq.” *EPJ Data Science* 3 (1): 1–29.
- Eck, Kristine. 2012. “In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets.” *Cooperation and Conflict* 47 (1): 124–41.
- Findley, Michael G., and Joseph K. Young. 2012. “Terrorism and Civil War: A Spatial and Temporal Approach to a Conceptual Problem.” *Perspectives on Politics* 10 (02): 285–305.
- Fortna, Virginia Page. 2013. “Choosing Terror: Rebels’ Use of Terrorism in Internal Armed Conflicts 1970–2010.” Paper presented at the Annual Meeting of the American Political Science Association, Chicago, August 29–September 1.
- Gale, David, and Lloyd S. Shapley. 1962. “College Admissions and the Stability of Marriage.” *The American Mathematical Monthly* 69 (1): 9–15.
- Hendrix, Cullen S., and Idean Salehyan. 2012. “Climate Change, Rainfall, and Social Conflict in Africa.” *Journal of Peace Research* 49 (1): 35–50.
- Herzog, Thomas N., Fritz J. Scheuren, and William E. Winkler. 2007. *Data Quality and Record Linkage Techniques*. New York: Springer.
- Hultman, Lisa, Jacob Kathman, and Megan Shannon. 2013. “United Nations Peacekeeping and Civilian Protection in Civil War.” *American Journal of Political Science* 57 (4): 875–91.
- Krüger, Jule, and Kristian Lum. 2015. “An Exploration of Multiple Systems Estimation for Empirical Research with Conflict-Related Deaths.” Paper presented at the Visions in Methodology Conference, University of Kentucky, May 13–15.

- Lynch, Marc. 2013. *The Arab Uprising: The Unfinished Revolutions of the New Middle East*. New York: Public Affairs.
- Lynch, Marc. 2016. *The New Arab Wars: Uprisings and Anarchy in the Middle East*. New York: Public Affairs.
- Manrique-Vallier, Daniel, Megan E. Price, and Anita Gohdes. 2013. Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflict. In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, edited by Taylor B. Seybolt, Jay D. Aronson, and Baruch Fischhoff, 1-31. Oxford, UK: Oxford University Press.
- Polo, Sara M. T., and Kristian Skrede Gleditsch. 2016. "Twisting Arms and Sending Messages: Terrorist Tactics in Civil War." *Journal of Peace Research* 53 (6): 815-29.
- Price, Megan, Anita Gohdes, and Patrick Ball. 2014. "Updated Statistical Analysis of Documentation of Killings in the Syrian Arab Republic." Human Rights Data Analysis Group, commissioned by the United Nations Office of the High Commissioner for Human Rights (OHCHR). Accessed June 15, 2017. <https://hrdag.org/wp-content/uploads/2014/08/HRDAG-SY-UpdatedReportAug2014.pdf>.
- Raleigh, Clionadh, and Håvard Hegre. 2009. "Population Size, Concentration, and Civil War. A Geographically Disaggregated Analysis." *Political Geography* 28 (4): 224-38.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED—Armed Conflict Location and Event Data." *Journal of Peace Research* 47 (5): 651-60.
- Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Lineberger, Emily Stull, and Jennifer Williams. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38 (4): 503-11.
- Salehyan, Idean, and Christopher Lineberger. 2015. "Elections and Social Conflict in Africa, 1990–2009." *Studies in Comparative International Development* 50 (1): 23-49.
- Shapley, Lloyd S., and Martin Shubik. 1972. "The Assignment Game 1: The Core." *International Journal of Game Theory* 1 (1): 111-30.
- Silva, Romesh, and Patrick Ball. 2006. "The Profile of Human Rights Violations in Timor-Leste, 1974–1999." A Report by the Benetech Human Rights Data Analysis Group to the Commission on Reception, Truth and Reconciliation of Timor-Leste. Accessed June 12, 2017. <http://hrdag.org/content/timorleste/Benetech-Report-to-CAVR.pdf>.
- START (National Consortium for the Study of Terrorism and Responses to Terrorism). 2013. Global Terrorism Database. Accessed April 20, 2018. <http://www.start.umd.edu/gtd>.
- Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523-32.
- Weidmann, Nils B. 2015. "On the Accuracy of Media-based Conflict Event Data." *Journal of Conflict Resolution* 59 (6): 1129-149.
- Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60 (1): 206-18.
- Weidmann, Nils B., Jan Ketil Rød, and Lars-Erik Cederman. 2010. "Representing Ethnic Groups in Space: A New Dataset." *Journal of Peace Research* 47 (4): 491-99.

- Wood, Reed M. 2014. "From Loss to Looting? Battlefield Costs and Rebel Incentives for Violence." *International Organization* 68 (4): 979-99.
- Zhukov, Yuri M., Christian Davenport, and Nadiya Kostyuk. 2017. "xSub: Cross-National Data on Subnational Violence." Produced and distributed by Center for Political Studies, University of Michigan, Ann Arbor, MI. Accessed April 16, 2018. <http://www.cross-sub.org/>.