Socio-political Events of Conflict and Unrest: A Survey of Available Datasets

Helene Bøsei Olsen[*] and **Étienne Simon**[*] and **Erik Velldal** and **Lilja Øvrelid**
University of Oslo, Language Technology Group

Abstract

There is a large and growing body of literature on datasets created to facilitate the study of socio-political events of conflict and unrest. However, the datasets, and the approaches taken to create them, vary a lot depending on the type of research they are intended to support. For example, while scholars from natural language processing (NLP) tend to focus on annotating specific spans of text indicating various components of an event, scholars from the disciplines of political science and conflict studies tend to focus on creating databases that code an abstract but structured representation of the event, less tied to a specific source text. The survey presented in this paper aims to map out the current landscape of available event datasets within the domain of social and political conflict and unrest – both from the NLP and political science communities – offering a unified view of the work done across different disciplines.

1 Introduction and background

Like in most social sciences, political scientists started to rely more and more on quantitative data to empirically test their hypotheses during the course of the 20th century. Hutter (1972) observes a rapid increase in the use of quantitative data, from 11.6% of political science articles in 1946–1948 to 58.5% in 1968–1970. To satisfy this demand for numerical data, researchers started manually collecting large databases of politically significant events from news journals (McClelland, 1978; Azar, 1980). These databases contain structured abstract descriptions of real-world events, enabling researchers to perform large-scale analysis. From an NLP perspective, these sorts of databases can be viewed as the desired output of the event extraction task. Event extraction models are trained on natural language texts, such as news or Wikipedia articles, annotated with event information at the token-level. Yet, while information extraction was originally motivated by practical endeavours (Sundheim and Chinchor, 1993; Grishman and Sundheim, 1996), modern event extraction is more closely associated with linguistic formalisations of sentential semantics and natural language understanding (Doddington et al., 2004).

When we look at both modern socio-political event databases and annotated NLP datasets, we observe several discrepancies that make annotated datasets less suited for the evaluation of socio-political event extraction systems. A first discrepancy pertains to the *link precision between text and events*. While the events encoded by database approaches commonly reflect information scattered in entire documents (typically one or multiple new articles), NLP events tend to be defined by word or phrase-level annotations tied to specific spans of text in a given document. A second and closely related discrepancy is what we refer to as the *abstraction gap*. For political science, the text of news articles is but a clue to what happened. Socio-political databases purpose to contain information about what actually happened in the real-world, which can only be elucidated through a combination of sources and expert knowledge. Moreover, the recorded events are typically defined within the context of the phenomena, theories, or research goals that are explored. In NLP, events are often defined based on linguistic motivations, meaning they are defined and specified within the text based on linguistic structures, patterns, or features present. The events defined in the text annotations of NLP datasets are usually more atomic and granular compared to the more aggregated and high-level events typically found in database resources. A third discrepancy has to do with *source text availability*, which is in turn closely tied to the underlying *purpose* of the data resource. While the main point of socio-political event databases is simply the set of

[*]These authors contributed equally to this work.

Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024), pages 40–53
March 22, 2024 ©2024 Association for Computational Linguistics

events themselves, i.e. the actual information that is recorded, the text annotations found within NLP, in contrast, are meant to enable training and/or testing of event extraction systems, i.e. systems that can map text into structured representations like those of the annotations. In NLP, therefore, it is generally seen as vital to make the annotated texts freely available, whereas it is significantly less common that the text sources used to build socio-political databases are shared. This has the unfortunate consequence of making many event databases not as directly applicable for NLP research as they might have been. A fourth discrepancy is related to the account of *temporal dynamics*. Socio-political event databases describe an evolving world, while annotated event extraction datasets are typically comprised of independent and identically distributed samples.

Several surveys in NLP describe annotated event datasets together with methodologies and techniques approaching the task of event extraction (Li et al., 2022; Yu et al., 2020; Xiang and Wang, 2019). Similarly, multiple articles describe socio-political databases, often with a focused comparison within the same domain, such as protest events (Hutter, 2014; Ward et al., 2013) or violent events (Hammond and Weidmann, 2014; Gleditsch et al., 2014). However, comprehensive studies linking the two fields together are notably lacking.

Considering the extensive data sources available, our survey does not aim to be exhaustive. Our primary focus is on central databases used in the social sciences and prominent annotated datasets in NLP concerning conflict and unrest. In structuring this survey, we classify datasets according to their purported goal. We start in Section 2 with datasets created for the main purpose of studying the recorded events themselves. We refer to these as *socio-political event databases*. The section will start by introducing manually annotated databases before we introduce databases created using automated methods. This naturally leads to Section 3 on *annotated event datasets* from the field of NLP covering socio-political events. The key characteristic of the datasets in this section is that they contain text-span annotations with the purpose of training and evaluating machine learning models for the event extraction task. We then describe and analyse the gap between the two types of event data and discuss works that can be seen as early attempts to bridge this gap in Section 4. Finally, we give special attention to our Ethics Section, as biases in the selection and description of datasets are critical when political analyses are derived from them.

As a note on terminology, while writing this survey, we opted to use the vocabulary of NLP, but also to make the parallel between the practices found between the two fields clearer. Instead of speaking of *annotation*, political scientists prefer the term of *coding*, which usually refers to manual annotation performed by human experts, but can also include *machine coding*, which refers to the automatic annotation of text by algorithms. Socio-political events usually involve one or more *actors*, those are entities, often states, armed groups, or other politically relevant organisations. Finally, the process of extracting political events from text is described in a *codebook*, which can be seen as similar in purpose to an annotation guideline.

2 Socio-political event databases

Early on, McClelland (1961) noted the necessity of building databases of politically relevant events to better our understanding of international politics. In contrast to annotated datasets geared towards training and evaluating systems for information extraction, these types of databases are built solely for the knowledge they encode, without much importance given to an underlying source text. The source texts are typically only included in the form of a reference for checking the validity of the event or indicating its provenance. However, most events recorded in such databases could, in principle, be automatically extracted from published texts.[1] Following this observation, there was an attempt to automatically extract these databases from news feeds in the late 1980s. These efforts resulted in the Kansas Event Data System (KEDS; Schrodt et al., 1994), extracting events from Reuters. This initiated the advance of machine-coded databases, which parallels the development of event extraction systems on the NLP side.

In this section, we describe important databases of socio-political conflict and unrest. While the focus of this survey is on data rather than modelling, we do briefly touch on methodology when we discuss the automatically extracted databases, where modelling and data are inherently intertwined. The main manually annotated databases included in this

[1] For recent conflicts, some databases such as UCDP GED use other sources of information in addition to text sources, such as images or videos posted on social media, but this is still an uncommon practice.

section are listed in Table 1.

2.1 Manually annotated databases

The first two widely used databases for socio-political events are manually annotated by humans and include the World Events Interaction Survey (WEIS; McClelland, 1978) and the Conflict and Peace Data Bank (COPDAB; Azar, 1980). While both focus on inter-state political events, they diverge in their selection of news sources to extract the events, consequently resulting in distinct geographical focus (Howell, 1983).

Even though the WEIS and COPDAB projects cover a broad range of politically relevant events, one of the main limitations is that these events only cover a limited set of actors. Attempting to code every potentially relevant political event is time-consuming, resource-intensive, and costly, and might be beyond human capacity.

Consequently, more recently manually annotated databases tend to have a very restricted focus, particularly oriented towards addressing a specific research question. For example, Turchin (2012) attempts to find a temporally repeating pattern in the occurrence of violence in the United States. To do so, they compile a list of what they consider political violent acts over the last two centuries. Such highly specialised databases may have little to offer with respect to other types of research questions. On the other hand, some databases are used in the analysis of a wide variety of research questions. One of the most widely used comes from the Correlates of War Project (COW; Sarkees and Wayman, 2010), which lists all wars with more than a 1 000 battle-related deaths since 1816 and is a popular database for research on inter-state conflicts.

A particularity of these databases is that the coded information is not necessarily reliant on a specific underlying news article. As described in Section 1, the extracted events in databases are typically not designed to facilitate mapping from text to a structured event representation but rather focus on being faithful recordings of actual events in the world. This places them at a higher level of abstraction compared to the annotations commonly encountered in NLP. Moreover, it is common that multiple sources such as news articles,[2] and reports from non-governmental organisations are used by expert annotators in deducing information about the recorded event in the database.

The Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED; Sundberg and Melander, 2013) is one such database. It focuses on a single event type: fatalities from armed conflict involving at least one organised actor. The UCDP GED events go back decades and are continuously updated with the same coding process: every month, region-specialised human experts read news articles about violent events and transcribe them into the database following the UCDP GED codebook (Högbladh, 2023). The data is widely used in peace and conflict studies and for research projects such as conflict escalation prediction (Hegre et al., 2022).

A similar program is the Armed Conflict Location & Event Data project (ACLED; Raleigh et al., 2010). Although it covers violent deaths to a lesser extent compared to UCDP GED, ACLED includes a larger number of event types such as protests, territory changes, and troop movements. The database provides researchers with an alternative trade-off between domain coverage and data quality compared to UCDP GED. Similarly, the Social Conflict Analysis Database (SCAD; Salehyan et al., 2012) has an analogous purpose to ACLED. It contains 10 event types and is designed to supplement the UCDP GED specifically in the African, Latin American, and Caribbean regions. While having a more narrow event domain compared to ACLED, SCAD has the advantage of being easy to merge with the high-quality UCDP GED armed conflict events.

The NAVCO database (Nonviolent and Violent Campaigns and Outcomes; Chenoweth et al., 2019) is designed to answer the following research question: *do nonviolent campaigns have better or worse odds of success compared to violent ones?* (Chenoweth and Stephan, 2011). The criteria for nonviolent campaigns within this database are more restrictive compared to SCAD because they require comparability with violent ones. Consequently, only nonviolent campaigns with a *maximalist* goal are included, i.e. protests and strikes that in other contexts could be violent.

Rather than focusing on a specific research question, some databases concentrate on a set of events with high political significance. An example of this approach is the Iraq Body Count database (IBC; Hicks et al., 2011). This database records civilian casualties resulting from violence following the

[2] Many socio-political event databases still rely on specific news articles, typically sourced from news aggregators like Factiva and LexisNexis, which provide access to thousands of news sources.

Database	Domain	Sources	# Events ×1000	ML Filter	Reference
COW	wars	news	1	no	Sarkees and Wayman (2010)
USPVD	violence	other databases...	2	no	Turchin (2012)
UCDP GED	fatal organised violence	news, social media...	316	no	Sundberg and Melander (2013)
ACLED	conflict & protest	news, social media...	1 967	no	Raleigh et al. (2010)
SCAD	protest	news	23	no	Salehyan et al. (2012)
NAVCO	non-violent & violent	news	112	no	Chenoweth et al. (2019)
IBC	civilian deaths	news, NGO...	52	no	Hicks et al. (2011)
MMAD	protest	news	31	yes	Weidmann and Rød (2019)
GTD	terrorism	news...	200	yes	START (2022)
SPEED	protest	news	62	yes+	Nardulli et al. (2015)

Table 1: Manually annotated socio-political event databases described in Section 2.1. Note that some of these databases are still being actively updated, the number of events is given at the time of writing. The "ML Filter" columns indicate whether news articles are selected using a simple keyword system or a machine learning system. SPEED is going one step further by pre-extracting named entities and is thus labelled "yes+".

2003 invasion of Iraq. Until 2007, it only recorded fatalities reported in at least two different news sources, and from 2017 onward, it only reported aggregated death counts. One specificity of this database is that it targets personal information, such as names or demographic details about the victims whenever available. The Bosnian book of dead (BBD; Ball et al., 2007) is a similar endeavour for the 1992–1995 war in Bosnia and Herzegovina.

All of these news-sourced databases use a set of search terms to pre-filter articles from news aggregators (Yörük et al., 2022). For instance, the search string used by the UCDP GED contains terms such as "kill", "die" or "massacre". Additionally, these databases indirectly rely on automatic tagging by filtering out news articles based on topic tags automatically assigned by the news aggregators (e.g. to remove sport-related articles that may use similar terms metaphorically).

Furthermore, some databases take an extra step by employing their own machine learning models to filter news aggregators. Nevertheless, they continue to involve human experts in extracting the specifics of the events. An illustration of this is the Mass Mobilisation in Autocracies Database (MMAD; Weidmann and Rød, 2019). This database approaches the filtering as a binary classification task where articles are categorised based on their inclusion of an MMAD event. For the filtering process, they train an ensemble of Support Vector Machines (SVMs) and naive Bayes classifiers on a set of 250 000 manually annotated articles (Croicu and Weidmann, 2015). They report that their system reduces the workload for human coders by half while discarding 10% of relevant articles.

In the same vein, the Global Terrorism Database (GTD; START, 2022) compiles terrorist incidents. Initially, the news articles are filtered by an unspecified machine learning algorithm before the events are extracted by a human expert. The implementation of this filtration method began in 2012, with the sole mention of a deduplication algorithm using cosine similarity on n-grams at that time. This uncertainty about the underlying model is prevalent with numerous databases within political sciences; there is often a lack of comprehensive publication detailing the filtering mechanisms used.

An example of the next step towards automation is the Social, Political and Economic Event Database project (SPEED; Nardulli et al., 2015). In addition to the filtering of relevant news articles, they use statistical models to extract potentially relevant entities such as locations and actors. These entities are then reviewed and combined by a human expert to form events.

2.2 Automatically extracted databases

Automatically extracted databases allow for potentially broader coverage by reducing the costs of human expert annotation. However, this advantage is counterbalanced by reduced accuracy. Consequently, when political scientists select a database to address their research questions, they are faced with a trade-off between quantity and quality. In practice, hand-annotated databases are favoured if they cover the specific research question, while machine-coded ones are preferred otherwise.

Similar to how the schema of manually annotated databases is described by a codebook (annotation guidelines), automatically extracted databases follow an *event ontology* or *event coding scheme*. These ontologies define the set of event types with the meaning of the various arguments within the event. Usually, the set of possible arguments remains constant for all event types and includes at least a source and a target actor.

In contrast to manually annotated databases for which there is a one-to-one relationship between codebook and databases, automatic event ontologies are often used and reused to define several databases. Initially though, ontologies and databases were jointly developed relying on preexisting codebooks.

The extensively used WEIS ontology, derived from the manually annotated WEIS database detailed in Section 2.1, serves as a foundational ontology for several efforts aiming to automate event databases. These efforts often build upon the WEIS ontology, either augmenting or expanding it to align with specific research questions or targeted domains. The Kansas Event Data System (KEDS; Schrodt et al., 1994) adapted WEIS for developing a database on inter-state interactions, but WEIS was also extended in a KEDS-model-compatible way within the PANDA project (Bond et al., 1994) with a focus on nonviolent direct action.

The KEDS model uses symbolic rules for matching words to classify events and identify named entities. It focuses on the first sentence of news articles, using the structure to complete event details. This method involves a simple form of parsing, by examining how entities and action words are related without analysing the entire sentence structure. These KEDS ideas were later incorporated into a new model named Textual Analysis by Augmented Replacement Instructions (TABARI; Schrodt, 2001). This evolution was followed by formalisations of coding schemes specific to automatic event extraction.

Currently one of the most popular event ontologies for machine-coded databases concerned with inter-state events is the CAMEO event ontology (Conflict and Mediation Event Observations; Gerner et al., 2002). It is specifically designed for rule-based extraction models, such as TABARI, describing more than 20 event types with over 200 subtypes. Additionally, the CAMEO codebook details a hierarchical coding scheme for events and entities, distinguishing CAMEO as a genuine ontology rather than merely an event catalogue.

Another widely used ontology is IDEA (Integrated Data for Events Analysis; King and Lowe, 2003), an earlier alternative to CAMEO. It is a direct successor of the previously mentioned PANDA project, concentrating on intra-state conflict and citizen direct actions. The popularity of these ontologies comes mostly from the fact that they provided a list of patterns to be used with TABARI-like models, both for actors and verbs associated with the events. In practice, these patterns resemble simplified regular expressions, indeed some "verbs" given by CAMEO are not conventional grammatical verbs, similar to how nouns can be event triggers in NLP.

A given machine-coded event database can be defined as a combination of a model, an ontology, and the utilised news sources. For example, a popular machine-extracted database is ICEWS (Integrated Crisis Early Warning System; O'brien, 2010), created at the initiative of DARPA for conflict forecasting. ICEWS is a database extracted from several international and regional sources (AP, UPI, BBC Monitor, India Today, etc) using the TABARI model with classification into the CAMEO ontology. Similarly, GDELT (Global Database of Events, Language, and Tone; Leetaru and Schrodt, 2013) is an academic initiative, a database containing CAMEO-events extracted by TABARI from the LexisNexis news aggregator. GDELT is one order of magnitude larger than ICEWS, with a tendency to be less conservative in its inclusion of events (Ward et al., 2013).

In 2014, the TABARI model was phased out in favour of new models named PETRARCH (Python Engine for Text Resolution And Related Coding

Hierarchy; Norris et al., 2017). These models are still rule-based, however, the rules are designed on parse trees extracted by Standford CoreNLP (Manning et al., 2014) instead of using basic string templates. The PETRARCH-2 model is used by the TERRIER database (Temporally Extended, Regular, Reproducible International Event Records; Grant et al., 2019) to extract CAMEO events from newspapers from 1979 to 2016. The PHOENIX database (Salam et al., 2020) is also using a PETRARCH model (UD-PETRARCH) to extract CAMEO events from more than 250 news sources, including Spanish language sources.

Recently, Halterman et al. (2023a) introduced the PLOVER ontology (Political Language Ontology for Verifiable Event Records) together with the POLECAT dataset (Political Event Classification, Attributes, and Types) as a replacement for the CAMEO ontology and ICEWS dataset. The dataset is extracted using the NGEC model (Halterman et al., 2023b), which is composed of SVM, distilBERT and RoBERTa.

3 Text annotation for event extraction

On the NLP side, annotated datasets are created for the purpose of training models, shaping their design and annotation to align with the event extraction task's approach. Event extraction has been a central task in NLP, dating back to the Message Understanding Conferences (MUC) series in the 1990s. Initially, annotating event participants was formulated to fit a template-filling task, where information from a document is to be structured into a predefined set of fields such as finding the victims, time and location from a terrorist attack report. Following these early attempts, the highly influential Automatic Content Extraction (ACE) program released manual event annotations for text spans at the sentence level, performed jointly with annotation of rich information about entities, temporal expressions, and relations between entities. Below we describe these in more detail and also compare them with more recent annotation efforts. The NLP datasets covered in this survey are summarised in Table 2.

While looking at the 1990s MUC datasets, it is striking how closely they resonate with current socio-political event databases compared to modern NLP annotated datasets. The evolution of template filling into event extraction is not clearly defined, and similar models are used for the two tasks (Du et al., 2021). Indeed, both of them capture a semantic relationship between entities, as described by the template or event schema. Two other closely related tasks are relation extraction – which usually focuses on binary templates, often in the context of knowledge bases – and semantic role labelling – which usually focuses on the argument relations conveyed by specific predicates. Even though all of these tasks can be relevant to socio-political event databases, in this section, we only focus on annotated datasets for event extraction and template-filling, describing them in chronological order.

The Message Understanding Conferences (MUC; Grishman and Sundheim, 1996), held from 1987 to 1997 and funded by DARPA, are regarded as pioneering efforts in generating annotated datasets for information extraction. The conferences operated as shared tasks, where each MUC is associated with a designated dataset covering the corresponding information to be extracted and prepared by human annotators for training purposes along with a task definition. Although MUC maintains mainly a military theme, the various datasets focus on different types of events.

The first two conferences centred on military messages from the tactical Navy domain. In MUC-1 the participants were provided with merely 10 paragraphs as data without any formal evaluation. Building on MUC-1, MUC-2 introduced a dataset with 130 messages and 10 elements to be extracted, such as event type, agent, time, place, and the effect of the event (Sundheim and Chinchor, 1993).

Following the initial conferences, MUC-3 and MUC-4 introduced annotated datasets focused on terrorist events in Central and South America, reported by the Foreign Broadcast Information Service. These iterations of MUC marked a shift by increasing the complexity of the task, both by including several event and argument types, but also by moving from extraction of information from simple and short military messages to longer texts with more complex language. MUC-4 includes 4 event types `Arson`, `Attack`, `Bombing`, `Kidnapping`, with the 4 arguments roles Perpetrator, Instrument, Target, and Victim, which are shared across event types. Additionally, the datasets increased in size, with respectively 1 400 and 1 700 news articles for MUC-3 and MUC-4.

The last two instalments, MUC-6 and MUC-7, shift the focus towards domain-independent annotations, targeting named entity recognition, coref-

Dataset	Domain	Source	Annotation scope	# Doc	# Event Types
MUC-4	terrorist attack	news	document	1 700	4
ACE2005	general	news, conversation	sentence	599	33
Light ERE	general	news, discussion forum	sentence	902	33
Rich ERE	general	news, discussion forum	sentence	288	38
MAVEN	general	Wikipedia	sentence	4 480	168
WIKIEVENTS	general	Wikipedia, news articles	document	246	67
DocEE	historical & news	Wikipedia, news articles	document	27 485	59
MEE	general	Wikipedia	5 sentences	31 226†	16

Table 2: Overview of annotated text datasets in the field of NLP for event extraction. †: In the case of MEE, the "# Doc" column reports the number of 5 sentences spans in the dataset, not the number of documents.

erence resolution, and relation identification. This transition also includes an expansion to more languages. Interestingly, this shift was accompanied by a return to smaller training datasets, comprising only 100 documents. MUC-6 consists of events involving high-level officers joining or departing from companies, while MUC-7 targets satellite launch events, with event arguments such as Date, Country of Launch, and Payload Information.

These were followed by the automatic content extraction (ACE) program. The event annotation in the ACE tradition has become a de facto standard for the evaluation of event extraction systems in the field of NLP. The ACE dataset-2005 (Doddington et al., 2004) provides manual annotation for entities, relations, and events for joint evaluation of multiple IE tasks and in multiple languages (ACE05 in English, Chinese, and Arabic). The annotations distinguish specific text spans indicating the event trigger and associated arguments of an event at the sentence level. An event trigger is typically the word(s) in the text that most clearly describes an event, such as "bomb", which evokes an `Attack` event in the example sentence "U.S. forces continued to bomb Fallujah" where "U.S. forces" is the associated Attacker argument. ACE annotates 8 general event types, e.g. `Life`, `Conflict`, `Transaction` with 33 subtypes (e.g. `Conflict.Attack`) and 22 argument roles, e.g. Attacker, Agent and Recipient. Of particular relevance in the current setting are the `Conflict` event type (with subtypes `Attack` and `Demonstration`) as well as the `Life.Die` and `Life.Injure` event types.

More recently, the Entities, Relations and Events (ERE) annotation effort (Song et al., 2015) has contributed both data and annotation guidelines for event extraction purposes. From the Light ERE to Rich ERE datasets, the ERE effort has evolved from lightweight annotation automating the ACE guidelines to more complex treatment of entities and events aimed at paving the way for event coreference at the document-level. The Rich ERE annotation scheme extends on that of ACE, annotating 38 event subtypes under 9 main event types, including more fine-grained event subtypes in the `Movement`, `Contact`, and `Transaction` event types. In Light ERE, an event trigger can be associated with only one event. Still, in Rich ERE, an event trigger can be annotated for more than one event due to correlations of different event types. For instance, an `Attack` event and an `Injure` event can share the same event trigger; it is natural that when a person is attacked, the person is also injured. In Light ERE, only asserted events are annotated; in Rich ERE, apart from assorted events, events that did not actually occur are also annotated, hence annotating event modality.

The MAssive eVENt detection dataset (MAVEN; Wang et al., 2020) is introduced to provide a large-scale annotated event dataset in the general domain, covering 168 event types. MAVEN follows the ACE terminology, targeting events at the sentence-level, and consists of event-related articles from English Wikipedia. FrameNet frames (Baker et al., 1998) are used to derive event types, with the lexical units serving as the corresponding triggers. Automatic POS-tagging and heuristic methods are used to narrow down trigger candidates and the corresponding event type candidates to aid human annotators. In MAVEN, the event types follow a hierarchical schema resembling a tree structure, prioritising the most detailed event types. If no fine-grained event type aligns with the event, the annotators resort to more general event types. For example, the most coarse-grained event type `Action` includes the event subtype `Violence`, which again contains subsubtypes such as `Killing`, `Attack`, `Terrorism`, and `Military Operation`. In the context of social and political conflict and unrest, the event types `Terrorism`, `Kidnapping`, `Violence`, `Use firearm`, `Military operation`, and `Attack` are especially relevant event types.

Li et al. (2021) presents WIKIEVENTS, a document-level annotated dataset based on Wikipedia articles and their referenced news articles. The annotations resemble ACE, but expand the number of sub-events from 33 to 67 following the KAIROS ontology. Additionally, it incorporates a more fine-grained event-type hierarchy. For instance, whereas ACE identifies the event type and subtype such as `Conflict.Attack`, WIKIEVENTS introduces event types at three levels, such as `Conflict.Attack.DetonateExplode`. Furthermore, Li et al. (2021) expand their annotations to include events that extend beyond the sentence boundary, capturing event arguments occurring in sentences lacking an explicit event trigger. Apart from the `Conflict.Attack` events, the dataset includes event types such as `Life.Die` and `Conflict.Demonstrate`, each with subtypes that are relevant in the socio-political domain context. Human annotators label event types, event mentions (triggers and arguments), and event coreferences across sentences in the document.

The DocEE dataset (Tong et al., 2022) is the largest document-level annotated dataset containing 27 485 documents and covers a wide range of event types in the socio-political domain, including `Armed Conflicts`, `Riot` and `Protest`. It includes two types of events, historical events, defined as events with their own Wikipedia page, and timeline events, which are news events organised in chronological order on Wikipedia. The Wikipedia article is annotated for the historical events, while the corresponding news article is used for the timeline events. Each document is manually given an event type based on the title and then annotated with event arguments from the event type schema. For example, the event type `Protest` is annotated with arguments Date, Location, Protesters, Cause, Slogan, Method, Arrested, Government Reaction, Casualties and Losses, and Damaged Property.

The recently released MEE dataset (Pouran Ben Veyseh et al., 2022) provides event-annotated data for eight typologically diverse languages (English, Spanish, Portuguese, Polish, Turkish, Hindi, Korean and Japanese). The data is based on Wikipedia articles under the subcategory *Event* from a number of different domains (e.g. Economy, Politics, Crimes and Military). The annotation scheme is based on the ACE guidelines and its 8 event types, however, limit the set of annotated subtypes to 16. Unlike ACE, the articles are split into 5-sentence segments and argument relations may span across the full-text segment. For the most relevant category in the current context, the dataset only includes the `Conflict.Attack`, `Life.Die` and `Life.Injure` event types.

4 Bridging the gap

In this section, we start by highlighting the main obstacles to transferring event extraction NLP expertise to the automatic extraction of socio-political event databases. One obstacle currently being addressed in the field of NLP is the restriction of events to single sentences. As we show in Section 3, document-level event extraction datasets are now starting to reemerge. In the second part of this section, we describe datasets that establish bridges between political science databases and annotated datasets.

Token-level annotations To facilitate model training, NLP event extraction datasets include token-level annotations delineating which words correspond to specific event triggers or arguments. On the other hand, manually coded socio-political event databases do not usually include this information, with the exception of the NER-automated SPEED database. Therefore, training machine learning models from socio-political databases requires either token-level annotation efforts or weakly supervised learning techniques. Alternatively, and perhaps more interestingly, one could directly prioritise research on end-to-end learning of document-level event extraction.

Source availability Regardless of the learning strategy used, a prerequisite is having available source texts, preferably in a free and open manner. The news articles used to code socio-political event databases are usually unavailable, mostly due to copyright restrictions. This significantly limits the appeal of these datasets within the NLP community.

Abstraction gap Furthermore, all the annotated datasets described in Section 3 solely capture the mapping between text and structured information, while the socio-political databases described in Section 2.1 attempt to record whether the event actually occurred in the real world. In the first case, only linguistic knowledge is necessary, even when encoding event modality. In the second, socio-political event databases require expert knowledge to evaluate and corroborate what is conveyed in the text. This implies that future research on learning to automatically extract document-level events should also address how to incorporate domain knowledge.

Temporal dynamic The socio-political databases describe an ever-changing situation with new actors regularly appearing and engaging in new conflicts. On the other hand, annotated datasets tend to be more stationary, with little to no temporal variation in the distribution of events.

In a way, the efforts to automatically create socio-political event databases overlook these issues because they tend to rely on older, rule-based models that do not necessitate data supervision. They fall into the abstraction gap by overcounting events, extracting from all uncorroborated news. Moreover, as they are typically not disclosed to the NLP community, there is no requirement to publish their source data.[3] This comes at the cost of reliability.

Some previous work has made efforts to bridge this gap between socio-political event databases and annotated event datasets. The MUC datasets, detailed in Section 2, represent the initial strides in this effort. We will here describe some of the more recent approaches.

The Iraq body count corpus (IBC-C; Žukov-Gregorič et al., 2016) is introduced to automate the annotation process for the Iraq body count project (Hicks et al., 2011) discussed in Section 2.1. The corpus provides event annotations for whole documents, where each document contains references to one or multiple events. The annotations for the IBC-C are created through a form of distant supervision (Mintz et al., 2009), using different pattern matching and semantic functions to create named entity labels corresponding to ten argument roles, such as Fatality Numbers, Named Individuals, and Location. IBC-C provides token-level annotations, somewhat addresses the abstraction gap and can capture the temporal dynamic of the evolving war. Unfortunately, the complete dataset is no longer available due to copyright restrictions (and potential privacy concerns).

The Global Contentious Politics database (GLOCON; Duruşan et al., 2022; Hürriyetoğlu et al., 2021b; Yörük et al., 2022) is a partly automated protest event database. Part of the data used to train the event extraction model is referred to as GLOCON GOLD and is freely available upon request.[4] It includes manually annotated datasets for three sub-tasks: document classification, sentence classification, and event extraction. It encodes five specific event sub-types: `Demonstrations`, `Industrial actions`, `Group clashes`, `Armed militancy`, and `Electoral mobilisation`. Regarding the concerns we identified, the dataset includes token-level annotations, is associated with source texts, and preserves the temporal dynamic of the political system. However, although future or hypothetical events are not annotated, these types of events can be recognised from linguistic cues alone, leading to continued susceptibility to the abstraction gap. Subsequently, the GLOCON GOLD dataset was extended to define the CASE 2021 and 2022 shared task 1 on protest news detection (Hürriyetoğlu et al., 2021a, 2022). Compared to GLOCON, the shared task datasets include more source articles and define an additional subtask: event sentence coreference identification. Finally, shared task 2 in 2021 (Haneczok et al., 2021) and 2023 (Tanev et al., 2023) attempt to bridge the gap more directly as they use data annotated following the ACLED codebook for evaluation. The 2023 task 2 tackles the prediction of battle events from social media messages in the Russo–Ukrainian war. On the prediction of whether a PRIO-grid cell contained a battle event, the two systems submitted for the task reached F_1 scores of 0.04 and 0.152, demonstrating the considerable amount of work that lies ahead.

5 Limitations

Due to space constraints, we needed to limit the number of datasets discussed. We strive to high-

[3]One exception is that ontologies underlying automatically extracted databases provide some short examples in their codebook. For example, the PLOVER ontology comes with a small (323 samples) hand-annotated dataset from the CAMEO codebook.

[4]https://github.com/emerging-welfare/glocongold

light datasets that are both central and relevant to the domain of political conflicts and unrest and showcase the evolution of practices in their respective fields. However, most of the datasets we selected are based on English-language news, even when used to analyse the political situation in non-English-speaking countries.

Some notable mentions that could not be included for relevancy or duplication concerns are POLDEM (Kriesi et al., 2020), MAR (Gurr, 2000), ICBe (Douglass et al., 2022), UCDP VPP (Svensson et al., 2022), PITF's WAD (Schrodt and Ulfelder, 2016), RAMS (Ebner et al., 2020), etc. Additionally, there has been a rise in annotated datasets made from user-generated text not encompassed in this survey, such as the Twitter-based datasets on civil unrest CUT (Sech et al., 2020) and G-CUT (Chinta et al., 2021).

6 Ethics

Working on event data concerning sensitive topics such as armed conflict, protest data, or other socio-political events necessitates a high degree of ethical consideration and responsibility.

The fact that the main source for several of the socio-political event databases is news articles, one should raise awareness of the inherent bias when reporting on these topics in the news. In the context of creating databases for conflict events based on media reporting, Chojnacki et al. (2012) highlights the importance of awareness towards both description bias, meaning errors in how conflicts are reported, and selection bias, meaning which conflicts are reported, and more importantly, those that remain unreported. Regarding selection bias, Chojnacki et al. (2012) suggests that researchers can solely make assumptions about the representativeness of the reported news, while for description bias, efforts should be made to mitigate and reduce potential bias in the extracted events. While similar biases can be present in manually annotated databases (McClelland, 1983), both description and selection biases from media sources can be partly mitigated using human experts to assess the validity of the reported events and or seek out sources to confirm the information. However, these types of biases do not seem to be addressed for annotated datasets in NLP.

Another concern is that this paper describes a dataset (IBC-C) that has been retracted due to copyright restrictions and is no longer accessible because of the mishandling of sensitive personal data. Other datasets are still accessible but do not clarify the handling of personal data and/or licences for redistributing data. Access to data while upholding copyright and privacy considerations is crucial to ethical research practice. Including these datasets in this work does not represent endorsement but is necessary to discuss different approaches and challenges associated with socio-political event data.

An important consideration when dealing with annotated datasets and databases involves the annotators' exposure to distressing or harmful content. Constantly engaging with descriptions of conflict and violence can lead to desensitisation, emotional numbness, and potential emotional and psychological distress. Recently, more attention has been directed toward the impact of secondary or vicarious trauma and the psychological well-being of annotators, content moderators, and others handling harmful content (Das et al., 2020; Steiger et al., 2021; Kirk et al., 2022). However, strategies and specific actions to alleviate potential risks for annotators, such as providing psychological support, remain limited or inadequately addressed in the datasets described in this paper. We strongly advocate for a more focused approach on supporting annotators to mitigate the effects of exposure and encourage leveraging existing datasets in research on automatic event extraction instead of creating new event datasets in order to minimise exposure.

We now address the concern of misuse and misinterpretations of socio-political event data. For instance, the GLOCON dataset strives to use neutral terms to describe different actors, e.g. using *militant*, instead of *terrorist*. Other datasets vary in their approaches when dealing with language that might be insulting, marginalising, or criminalising. The extent of this handling often depends on factors such as the use of standardised actor lists and whether the datasets are manually annotated by experts. Notably, in annotated datasets used for NLP, with a one-to-one mapping between text-span and label, this issue remains unaddressed. The vocabulary used to describe individuals and groups, particularly those from minority communities, holds the dual power to shape our perceptions of said groups and might impact the reliability of extracted events and subsequent analyses derived from event databases.

Finally, some datasets described in this work may contain fine-grained details about individuals, organisations, or groups, which can be used maliciously. Some datasets, such as GLOCON enforce responsible data use and seek to mitigate unethical usage by assessing the declared research intentions before granting access to the dataset (Yörük et al., 2022).

Acknowledgments

We would like to thank the anonymous reviewers as well as Mert Can Yilmaz for their valuable comments. We are also grateful to the Peace Science Infrastructure project collaborators from PRIO and UCDP for insightful discussions on the peace science perspective.

References

Edward E Azar. 1980. The conflict and peace data bank (COPDAB) project. *Journal of Conflict Resolution*, 24(1):143–152.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Patrick Ball, Ewa Tabeau, and Philip Verwimp. 2007. The bosnian book of dead: assessment of the database (full report). Technical report, Households in Conflict Network.

Doug Bond, Brad Bennett, and William Vogele. 1994. Data development and interaction events analysis using KEDS/PANDA: an interim report. *International Studies Association, Washington*.

Erica Chenoweth, Jonathan Pinckney, and Orion A. Lewis. 2019. NAVCO 3.0 Dataset.

Erica Chenoweth and Maria J Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

Abhinav Chinta, Jingyu Zhang, Alexandra DeLucia, Mark Dredze, and Anna L. Buczak. 2021. Study of manifestation of civil unrest on Twitter. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 396–409, Online. Association for Computational Linguistics.

Sven Chojnacki, Christian Ickler, Michael Spies, and John Wiesel. 2012. Event data on armed conflict and security: New perspectives, old challenges, and some solutions. *International Interactions*, 38(4):382–401.

Mihai Croicu and Nils B Weidmann. 2015. Improving the selection of news reports for event coding using ensemble classification. *Research & Politics*, 2(4):2053168015615596.

Anubrata Das, Brandon Dang, and Matthew Lease. 2020. Fast, accurate, and healthier: Interactive blurring helps moderators reduce exposure to harmful content. In *AAAI Conference on Human Computation & Crowdsourcing*.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *International Conference on Language Resources and Evaluation*, volume 2, pages 837–840. Lisbon.

Rex W. Douglass, Thomas Leo Scherer, J. Andrés Gannon, Erik Gartzke, Jon Lindsay, Shannon Carcelli, Jonathan Wilkenfeld, David M. Quinn, Catherine Aiken, Jose Miguel Cabezas Navarro, Neil Lund, Egle Murauskaite, and Diana Partridge. 2022. Introducing the icbe dataset: Very high recall and precision event extraction from narratives about international crises.

Xinya Du, Alexander Rush, and Claire Cardie. 2021. Template filling with generative transformers. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 909–914, Online. Association for Computational Linguistics.

Fırat Duruşan, Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Çağrı Yoltar, Burak Gürel, and Alvaro Comin. 2022. Global contentious politics database (GLOCON) annotation manuals.

Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. 2020. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 8057–8077, Online. Association for Computational Linguistics.

Deborah J Gerner, Philip A Schrodt, Omür Yilmaz, and Rajaa Abu-Jabr. 2002. Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*.

Kristian Skrede Gleditsch, Nils W Metternich, and Andrea Ruggeri. 2014. Data and progress in peace and conflict research. *Journal of Peace Research*, 51(2):301–314.

Christan Grant, Andrew Halterman, Jill Irvine, Yan Liang, and Khaled Jabr. 2019. OU event data project.

Ralph Grishman and Beth Sundheim. 1996. Message Understanding Conference- 6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*.

Ted Robert Gurr. 2000. *Peoples versus states: Minorities at risk in the new century*. US Institute of Peace Press.

Andrew Halterman, Benjamin Bagozzi, Andreas Beger, Phil Schrodt, and Grace Scraborough. 2023a. PLOVER and POLECAT: A new political event ontology and dataset.

Andrew Halterman, Philip A. Schrodt, Andreas Beger, Benjamin E. Bagozzi, and Grace I. Scarborough. 2023b. Creating custom event data without dictionaries: A bag-of-tricks. *arXiv preprint arXiv:2304.01331*.

Jesse Hammond and Nils B Weidmann. 2014. Using machine-coded event data for the micro-level study of political violence. *Research & Politics*, 1(2).

Jacek Haneczok, Guillaume Jacquet, Jakub Piskorski, and Nicolas Stefanovitch. 2021. Fine-grained event classification in news-like text snippets - shared task 2, CASE 2021. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 179–192, Online. Association for Computational Linguistics.

Håvard Hegre, Paola Vesco, and Michael Colaresi. 2022. Lessons from an escalation prediction competition. *International Interactions*, 48(4):521–554.

Madelyn Hsiao-Rei Hicks, Hamit Dardagan, Gabriela Guerrero Serdán, Peter M. Bagnall, John A. Sloboda, and Michael Spagat. 2011. Violent deaths of iraqi civilians, 2003–2008: Analysis by perpetrator, weapon, time, and location. *PLOS Medicine*, 8(2):1–15.

Llewellyn D Howell. 1983. A comparative study of the WEIS and COPDAB data sets. *International Studies Quarterly*, 27(2):149–159.

Ali Hürriyetoğlu, Osman Mutlu, Fırat Duruşan, Onur Uca, Alaeddin Gürel, Benjamin J. Radford, Yaoyao Dai, Hansi Hettiarachchi, Niklas Stoehr, Tadashi Nomoto, Milena Slavcheva, Francielle Vargas, Aaqib Javid, Fatih Beyhan, and Erdem Yörük. 2022. Extended multilingual protest news detection - shared task 1, CASE 2021 and 2022. In *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pages 223–228, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ali Hürriyetoğlu, Osman Mutlu, Erdem Yörük, Farhana Ferdousi Liza, Ritesh Kumar, and Shyam Ratan. 2021a. Multilingual protest news detection - shared task 1, CASE 2021. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 79–91, Online. Association for Computational Linguistics.

Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Fırat Duruşan, Çağrı Yoltar, Deniz Yüret, and Burak Gürel. 2021b. Cross-context news corpus for protest event-related knowledge base construction. *Data Intelligence*, 3(2):308–335.

James L. Hutter. 1972. Statistics and political science. *Journal of the American Statistical Association*, 67(340):735–742.

Swen Hutter. 2014. Protest event analysis and its offspring. In *Methodological Practices in Social Movement Research*. Oxford University Press.

Stina Högbladh. 2023. UCDP georeferenced event dataset codebook version 23.1. *Department of Peace and Conflict Research*.

Gary King and Will Lowe. 2003. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(3):617–642.

Hannah Kirk, Abeba Birhane, Bertie Vidgen, and Leon Derczynski. 2022. Handling and presenting harmful text in NLP research. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 497–510, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Hanspeter Kriesi, Bruno Wüest, Jasmine Lorenzini, Peter Makarov, Matthias Enggist, Klaus Rothenhäusler, Thomas Kurer, Silja Häusermann, Patrice Wangen, Argyrios Altiparmakis, et al. 2020. Poldem-protest dataset 30 european countries.

Kalev Leetaru and Philip A Schrodt. 2013. GDELT: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer.

Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, and Philip S. Yu. 2022. A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21.

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 894–908, Online. Association for Computational Linguistics.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Charles McClelland. 1978. World event/interaction survey, 1966-1978. *WEIS Codebook ICPSR*, 5211(640):49.

Charles A McClelland. 1961. The acute international crisis. *World Politics*, 14(1):182–204.

Charles A McClelland. 1983. Let the user beware. *International Studies Quarterly*, 27(2):169–177.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the* ACL *and the 4th International Joint Conference on Natural Language Processing of the* AFNLP, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Peter F. Nardulli, Scott L. Althaus, and Matthew Hayes. 2015. A progressive supervised-learning approach to generating rich civil strife data. *Sociological Methodology*, 45(1):148–183.

Clayton Norris, Philip A Schrodt, and John Beieler. 2017. PETRARCH2: Another event coding program. *J. Open Source Softw.*, 2(9):133.

Sean P O'brien. 2010. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International studies review*, 12(1):87–104.

Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Nguyen. 2022. MEE: A novel multilingual event extraction dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9603–9613, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Clionadh Raleigh, Andrew Linke, Havard Hegre, and Joakim Karlsen. 2010. Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5):651–660.

Sayeed Salam, Patrick Brandt, Vito D'Orazio, Jennifer Holmes, Javiar Osorio, and Latifur Khan. 2020. An online structured political event dataset based on CAMEO ontology.

Idean Salehyan, Cullen S Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. Social conflict in africa: A new database. *International Interactions*, 38(4):503–511.

Meredith Sarkees and Frank Wayman. 2010. *Resort to War, 1816-2007*. CQ Press.

Philip A Schrodt. 2001. Automated coding of international event data using sparse parsing techniques. In *annual meeting of the International Studies Association, Chicago*. Citeseer.

Philip A. Schrodt, 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.

Philip A Schrodt and Jay Ulfelder. 2016. Political instability task force atrocities event data collection codebook. *Parus Analytics*.

Justin Sech, Alexandra DeLucia, Anna L. Buczak, and Mark Dredze. 2020. Civil unrest on Twitter (CUT): A dataset of tweets to support research on civil unrest. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 215–221, Online. Association for Computational Linguistics.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ERE: annotation of entities, relations, and events. In *Proceedings of the the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98.

START. 2022. Global terrorism database 1970–2020.

Miriah Steiger, Timir J. Bharucha, Sukrit Venkatagiri, Martin Johannes Riedl, and Matthew Lease. 2021. The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Ralph Sundberg and Erik Melander. 2013. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4):523–532.

Beth M. Sundheim and Nancy A. Chinchor. 1993. Survey of the Message Understanding Conferences. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

Isak Svensson, Susanne Schaftenaar, and Marie Allansson. 2022. Violent political protest: Introducing a new uppsala conflict data program data set on organized violence, 1989-2019. *Journal of Conflict Resolution*.

Hristo Tanev, Nicolas Stefanovitch, Andrew Halterman, Onur Uca, Vanni Zavarella, Ali Hurriyetoglu, Bertrand De Longueville, and Leonida Della Rocca. 2023. Detecting and geocoding battle events from social media messages on the russo-Ukrainian war: Shared task 2, CASE 2023. In *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pages 160–166, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

MeiHan Tong, Bin Xu, Shuai Wang, Meihuan Han, Yixin Cao, Jiangqi Zhu, Siyu Chen, Lei Hou, and Juanzi Li. 2022. DocEE: A large-scale and fine-grained benchmark for document-level event extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3970–3982, Seattle, United States. Association for Computational Linguistics.

Peter Turchin. 2012. Dynamics of political instability in the united states, 1780–2010. *Journal of Peace Research*, 49(4):577–591.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. MAVEN: A Massive General Domain Event Detection Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1652–1671, Online. Association for Computational Linguistics.

Michael D Ward, Andreas Beger, Josh Cutler, Matthew Dickenson, Cassy Dorff, and Ben Radford. 2013. Comparing GDELT and ICEWS event data. *Analysis*, 21(1):267–297.

Nils B Weidmann and Espen Geelmuyden Rød. 2019. *The Internet and political protest in autocracies*, chapter 4. Oxford Studies in Digital Poli.

Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.

Erdem Yörük, Ali Hürriyetoğlu, Fırat Duruşan, and Çağrı Yoltar. 2022. Random sampling in corpus design: Cross-context generalizability in automated multicountry protest event collection. *American Behavioral Scientist*, 66(5):578–602.

Manzhu Yu, Myra Bambacus, Guido Cervone, Keith Clarke, Daniel Duffy, Qunying Huang, Jing Li, Wenwen Li, Zhenlong Li, Qian Liu, et al. 2020. Spatiotemporal event detection: A review. *International Journal of Digital Earth*, 13(12):1339–1365.

Andrej Žukov-Gregorič, Zhiyuan Luo, and Bartal Veyhe. 2016. IBC-C: A dataset for armed conflict analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 374–379.