
UCDP Georeferenced Event Dataset Codebook

Version 26.1

Uppsala Conflict Data Program

Department of Peace and Conflict Research
Uppsala University

This version compiled and updated by Stina Högladh (2026)

When using this data, please always cite:

Sundberg, Ralph, and Erik Melander, 2013, "Introducing the UCDP Georeferenced Event Dataset", Journal of Peace Research, vol.50, no.4, 523-532

When appropriate, also cite this codebook: Högladh Stina, 2026, "UCDP GED Codebook version 26.1", Department of Peace and Conflict Research, Uppsala University

Always include the Version number in analyses using the dataset.

*The official abbreviation of this dataset is **UCDP GED**.
The official full name of this dataset is **UCDP Georeferenced Event Dataset**.
The current version of the dataset is **26.1***

Data extracted from UCDP systems on 2026-03-30

Contents

1. Introduction.....	3
2. Definition of event in organized violence.....	4
3. Variables in the UCDP Georeferenced Event Dataset	5
3.1 API representation of variables	6
4. Data Collection Methods	13
4.1. First pass: Sources read each month.....	13
4.2. Second pass: Specialized sources	14
4.3. Further considerations for data validity and reliability	15
4.4. Quality assurance.....	15
5. Additional information on variables in the dataset	16
5.1. Event identifiers	16
5.2. Actors and dyads.....	16
5.3. Description of sources	19
5.4. Geography.....	21
5.4.1. Geo-referencing sources	22
5.4.2. Geo-precision and its values	23
5.5. Clarity	24
5.6. Time	25
5.7. Fatality figures	26
5.8. Variables present in previous versions of GED not used in version 26.1 ..	26
6. Format availability	27
7. References	28
Appendix 1 UCDP definitions.....	29
Definition of state-based armed conflict.....	29
Definition of non-state conflict.....	30
Definition of one-sided violence.....	31
Appendix 2 Temporal precision and date estimation rules.....	32
Appendix 3 Geo-precision coding rules.....	35
Appendix 4 Comparison with other event ontologies	38
Appendix 5 Main sources consulted during the 26.1 update	39
Appendix 6 Vague Number Translator	51

1. Introduction

This document describes the UCDP Georeferenced Event Dataset (UCDP GED), a project within the Uppsala Conflict Data Program (UCDP), at the Department of Peace and Conflict Research, Uppsala University. This version of the dataset is fully compatible with other UCDP datasets released in version 26.1.

GED 26.1 is a global dataset that covers the world between 1989-01-01 and 2025-12-31.

The purpose of this dataset is to provide the academic community with the most comprehensive structured event data available on organized violence in the post-1989 world, in response to the growing demand for geographically and temporally disaggregated data.

Whereas the ambition is to provide a dataset with both theoretical and practical relevance for researchers in a broad range of scholarly traditions, mainly pragmatic and practical decisions guide the construction of the dataset. This allows for effective coding procedures as well as disaggregated and flexible data without predetermined biases for certain research purposes. The geo-referenced event data may thus be used for purposes ranging from illustrating conflict behaviour geographically using geographic information systems software to studying causal pathways through a variety of statistical methods.

Whilst retaining the ambition to provide a dataset open to a broad variety of research purposes, the focus of the dataset on conflict dynamics and the effects of armed violence in the form of deaths still sets the parameters for users. This means that the UCDP GED is primarily directed toward quantitative and comparative researchers interested in the fatal outcomes of violent conflict behaviour at the subnational level.

Thus, the dataset is constructed in such a way as to maximize comparability and consistency across time and space and to provide a globally consistent representation of organized violence.

In effect, the goal of UCDP GED is not to present the most complete and accurate image of a certain conflict at a certain point in time, but rather to provide a tool for the comparative understanding of subnational conflict patterns and trends in global data.

All events are linkable to violence reported in the UCDP/PRIO Armed Conflict Dataset, the UCDP Non-State Conflict Dataset, or the UCDP One-Sided Violence Dataset. Events are included both for years in which the violence reaches the annual UCDP inclusion threshold of 25 fatalities and for years in which it does not reach this threshold.

The highest spatial resolution of the dataset is the individual village or town. The dataset is fully geocoded. The highest temporal resolution of the dataset is the day.

Earlier versions and definitions authored by:

Mihai Croicu,
Ralph Sundberg, Ph. D.

The UCDP geo-referencing and event data project is grateful for valuable external input from Halvard Buhaug at the Peace Research Institute, Oslo (PRIO), Nils B. Weidman at Konstanz University, Luc Girardin at ETH Zurich as well as Tomislav Dulic at the Hugo Valentin Centre, Uppsala University, for infrastructural support, ideas and comments in the early stages of this project.

2. Definition of event in organized violence

The basic unit of analysis for the UCDP GED dataset is the “event”, i.e. an individual incident (phenomenon) of lethal violence occurring at a given time and place.

UCDP define an event as: *An incident where armed force was used by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date”.*

The separate elements of the definition are operationalized as follows:

1. *Armed force*: use of arms in order to promote the parties’ general position in the conflict, resulting in deaths.

- arms: any material means e.g. manufactured weapons but also sticks, stones, fire, water etc.

2. *Organized actor*: a government of an independent state, a formally organized group or an informally organized group according to UCDP criteria:

- a. *Government of an independent state*: The party controlling the capital of a state.
- b. *Formally organized group*: Any non-governmental group of people having announced a name for their group and using armed force against a government (state-based), another similarly formalized group (non-state conflict) or unorganized civilians (one-sided violence). The focus is on armed conflict involving consciously conducted and planned political campaigns rather than spontaneous violence.
- c. *Informally organized groups*: Any group without an announced name, but which uses armed force against another similarly organized group (non-state conflict), where the violent activity indicates a clear pattern of violent incidents that are connected and in which both groups use armed force against the other

3. *Direct death*: a death relating to either combat between warring parties or violence against civilians.

UCDP GED provides three estimates for deaths for each event, thus creating an uncertainty interval:

- a low estimate, containing the most conservative estimate of deaths that is identified in the source material;

- a best estimate, containing the most reliable estimate of deaths identified in the source material;
- a high estimate, containing the highest reliable estimate of deaths identified in the source material. Note that UCDP attempts to distinguish and not include unreasonable claims in the high estimate of fatalities, and tends to be highly conservative when counting fatalities¹.

In order for an event to exist, at least one dead needs to be registered in the high, best or low estimate.

4. *Specific location*: a name and one pair of latitude and longitude coordinates that relate to the geographical information specified in the source material. A “specific location” refers to the best identifiable spatial reference available in the source material, which may range from an exact point to an administrative area or larger spatial feature.

5. *Specific date*: a specified time period during which armed interactions cause at least 1 fatality. The normal temporal unit to which an event can be related is a 24-hour day starting at midnight.

- In some cases it is impossible, based on the source material, to reduce the specific date to a single day as reporting only refers to wider time spans (multiple days) or information on the exact day is not clear. For these events, a wider time span is provided through the use of the *date_start*, *date_end* and *date_prec* variables.

For more definitions, please consult appendix I of this codebook.

3. Variables in the UCDP Georeferenced Event Dataset

This section provides brief descriptions of all variables in the dataset. More detailed definitions and coding rules for each variable are given in Section 5.

The event dataset has a dyad and actor focus, tracing the events of all UCDP conflict dyads² for both active years (years that have crossed the 25 battle related deaths threshold) and non-active years (the remainder).

Thus, if a dyad crossed the 25-deaths threshold in a single year, but did generate some events in either previous or subsequent years, all events belonging to the dyad are included, including those in years where the threshold was not crossed³.

The dataset includes all three types of UCDP organized violence: state-based conflict, non-state conflict and one-sided violence. All three categories of the UCDP annual data are

¹ For a more elaborate discussion on aspects concerning point 1-4, please refer to UCDP Codebooks for State-Based Armed Conflicts, Non-state Conflicts and One-Sided Violence.

² A dyad consists of two conflicting primary parties or party killing unarmed civilians. In state-based armed conflicts, at least one of the primary parties must be the government of a state.

A state-based conflict can include more than one dyad, if multiple groups oppose the government over the same incompatibility; non-state conflicts and one-sided violence instances are always equivalent to a dyad.

Further, in non-state armed conflicts, a dyad can only consist of formally versus formally organized groups or informally versus informally organized groups. A formally organized group can not be fighting an informally organized group to keep non-state conflicts and one-sided violence as independent categories.

³ E.g. State-based dyad 691 (Government of Uganda – UNRF II) crosses the 25 battle-related deaths threshold only in 1997. However, this dyad had some events, but did not cross the 25 battle-related deaths in 1996 and 1998. In versions 1.0 and 1.1 only those events belonging to the dyad in 1997 were included. In this version all the events belonging to the dyad (including those in 1996 and 1998) were included.

mutually exclusive and coded events will therefore be exclusive and non-overlapping. The data series start in 1989 and events before this calendar year are not included.

All the inclusion criteria are identical to UCDP GED version 1.5 – 26.1. The UCDP GED 26.1 is fully compatible with UCDP GED 25.1, 24.1, 23.1, 22.1, 21.1, 20.1, 19.1, 18.1, 17.1 and 17.2. It is (mostly) backwards compatible with UCDP GED versions 1.0-5.0. Check the compatibility notes below for further details. Note that starting with version 17.1, significant changes were made in the actor, dyad and actor/side id meaning those identifiers were no longer backwards compatible with older versions.

This version of GED corresponds with versions 26.1 of all other data in the UCDP data catalogue. Note that the release of UCDP GED is not synchronized perfectly with the above datasets, thus data discrepancies may appear due to data revisions.

3.1 API representation of variables

All variables described in this codebook are also available via the UCDP GED API. The API returns data in JSON format.

- Integer, numeric and date variables are represented as JSON numbers or strings in ISO format (YYYY-MM-DD) for dates. In the API, date_start and date_end are returned as ISO-8601 strings in the format YYYY-MM-DD. Users should convert them to date objects in their software of choice.
- Text variables are represented as JSON strings. When multiple values (e.g. multiple sources) are stored in a single field, they are separated by semicolons (;). Commas within titles or names are kept inside quotes.
- Missing values in the API are returned as null (for numeric and identifier fields) or as empty strings (for some text fields), depending on the historical implementation of the variable.

Users should always refer to the variable definitions in this codebook for conceptual meaning, and treat the API documentation as complementary technical documentation.

Variable name	Content	Type
id	A unique numeric ID identifying each event.	integer
relid	Users should rely exclusively on id for persistent event identification across versions. The relid can change when the event is updated Structure: country-year-type-dyad-sequence	
year	The year of the event	integer
active_year	1: if the event belongs to an active conflict/dyad/actor-year 0: otherwise	integer
type_of_violence	Type of UCDP conflict: 1: state-based conflict 2: non-state conflict 3: one-sided violence	integer
code_status	Status flag for events in the API. In the monthly candidate event feed this variable distinguishes between events that are still under review and events that have been fully coded. In the yearly	string

	released versions (e.g. GED 25.1, 26.1) the status is always “Clear”.	
conflict_dset_id	Only used in older versions of the dataset, exists but should not be used in the API 26.1 version, removed in the other formats for UCDP GED 26.1.	integer
conflict_new_id	A unique conflict identification code for each individual conflict in the dataset.	integer
	<p>UCDP Conflict ID for state based, non-state conflicts and one-sided violence as per the UCDP/PRIO Armed Conflict Dataset and UCDP Non-State Dataset and UCDP One-Sided dataset version 26.1.</p> <p>Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.</p> <p>This identifier is unique across the dataset (i.e. a non-state conflict cannot have the same identifier as a state-based conflict or a one-sided instance), irrespective of type of violence, and may be used for filtering and aggregation</p> <p>Warning: Not compatible with pre-17.1 versions of any UCDP datasets.</p>	
conflict_name	Name of the UCDP conflict to which the event belongs. For non-state conflicts and one-sided violence this is the same as the dyad name.	string(9999)
dyad_dset_id	Only used in older versions of the dataset, exists but should not be used in the API 26.1 version, removed in the other formats for UCDP GED 26.1.	integer
dyad_new_id	A unique conflict identification code for each individual dyad in the dataset.	integer
	<p>UCDP Dyad ID for state based conflicts, non-state conflicts and one-sided incidences as per the UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Datasets versions 26.1.</p> <p>Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.</p> <p>This identifier is unique across the dataset (i.e. a non-state conflict cannot have the same identifier as a state-based conflict or a one-sided instance), irrespective of type of violence, and may be used for filtering and aggregation</p>	

Warning: Not compatible with pre-17.1 versions of any UCDP datasets.		
dyad_name	Name of the conflict dyad creating the event.	string(9999)
<p>A dyad is the pair of two actors engaged in violence (in the case of one-sided violence, the perpetrator of violence and civilians).</p> <p>The two sides are separated by an ASCII dash (e.g. Government of Russia - Caucasus Emirate, Taleban - civilians).</p>		
side_a_dset_id	Only used in older versions of the dataset, exists but should not be used in the API 26.1 version, removed in the other formats for UCDP GED 26.1.	integer
side_a_new_id	A unique ID of side A.	integer
<p>Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.</p> <p>Warning: Not compatible with pre-17.1 versions of any UCDP datasets.</p> <p>Note that this ID is no longer the Gleditsch and Ward number for State actors/sides. If you need that identifier, use gwnoa described below.</p>		
side_a	The name of Side A in the dyad. In state-based conflicts always a government. In one-sided violence always the perpetrating party.	string(9999)
side_b_dset_id	Only used in older versions of the dataset, exists but should not be used in the API 26.1 version, removed in the other formats for UCDP GED 26.1.	integer
side_b_new_id	A unique ID of side B.	integer
<p>Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.</p> <p>Warning: Not compatible with pre-17.1 versions of any UCDP datasets.</p> <p>Note that this ID is no longer the Gleditsch and Ward number for State actors/sides. If you need that identifier, use gwnob described below.</p>		
side_b	The name of Side B in the dyad. In state-based always the rebel movement or rivalling government. In one-sided violence always "civilians".	string(9999)
number_of_sources	Number of total sources containing information for an event that were consulted.	integer

	<p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, -1. Note that -1 does not mean information on the source is missing; reference to the source material is always available in the source_article field.</p>	
source_article	References to the names, dates and titles of the source material from which information on the event is gathered.	text
<p>A reference to at least one source material is available for ALL EVENTS.</p> <p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For such older data, abbreviations are sometimes used for source agencies.</p>		
source_office	The name of the organizations publishing the source materials.	text
	<p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always available in the source_article field, for every event.</p>	
source_date	The dates the source materials were published on.	text
	<p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always available in the source_article field, for every event.</p> <p>1753-01-01 is set as a default date when the date is missing.</p>	
source_headline	The titles of the source materials.	text
	<p>Note that this variable is available for all data collected after 2014, for most data collected in 2013 and for recently revised events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always</p>	

	available in the source_article field, for every event.	
source_original	The name or type of person or organization from which the information about the event originates in the original report. e.g. "police", "Lt. Col. Johnson", "eyewitnesses", "rebel spokesman".	string(9999)
where_prec	The precision with which the coordinates and location assigned to the event reflects the location of the actual event. 1: exact location of the event known and coded. 2: event occurred within at maximum 25 km radius around a known point. The coded point is the known point. 3: only the second order administrative division where an event happened is known. That administrative division is coded with a point representing it (typically the centroid). 4: only the first order administrative division where an event happened is known. That administrative division is coded with a point representing it (typically the centroid). 5: the only spatial reference for the event is neither a known point nor a known formal administrative division, but rather a linear feature (e.g. a long river, a border, a longer road or the line connecting two locations further afield than 25 km) or a fuzzy polygon without defined borders (informal regions, large radiuses etc.). A representation point is chosen for the feature and employed. 6: only the country where the event took place in is known. 7: event in international waters or airspace.	integer
where_coordinates	Name of the location to which the event is assigned. Fully standardized and normalized.	string(9999)
where_description	Comment on the location coded, sometimes left empty can include area of the capital or name of a village that has not been found	string
adm_1	Name of the first order (largest) administrative division where the event took place	string(9999)
adm_2	Name of the second order administrative division where the event took place	string(9999)
latitude	Latitude (in decimal degrees)	numeric(9,6)
longitude	Longitude (in decimal degrees)	numeric(9,6)
geom_wkt	An Open Geospatial Consortium textual representation of the location of each individual point. Formatted as OGC WKT (<i>well known text</i>) without SRID.	string(9999)

priogrid_gid	<p>The PRIO-GRID cell id (gid) in which the event took place. Compatibility with PRIO-GRID (Tollefsen, 2012) is guaranteed for both PRIO-GRID 1 and 2.</p> <p>Warning: We associate every point to the PRIO-GRID that contains it, even if the point is in another country than the one officially assigned to the respective PRIO-GRID cell through their majority area rule. It is your responsibility to make sure the covariates for the PRIO-GRID cell are correct for each event. Further, for the same reason, DO NOT, under any circumstances, first clip out (subset) PRIO-GRID by country before merging with UCDP GED as data loss will certainly occur. Refer to your copy of the PRIO-GRID for further details on PRIO-GRID's majority assignment rule (p.3 in PRIO-GRID's original codebook).</p>	integer
country	Name of the country in which the event takes place.	string(999)
country_id	Gleditsch and Ward number of the country in which the event takes place.	integer
region	Region where the event took place. One of following: {Africa, Americas, Asia, Europe, Middle East}	string(999)
event_clarity	<p>1 (high) for events where the reporting allows the coder to identify the event in full. That is, events where the individual happening is described by the original source in a sufficiently detailed way as to identify individual incidents, i.e. separate activities of fighting in a single location:</p> <p>Example of such reporting: <i>"2 people where killed in Banda Aceh town on the 9th of December in fighting between the government and GAM when a car exploded in a main market."</i></p> <p>2 (lower) for events where an aggregation of information was already made by the source material that is impossible to undo in the coding process. Such events are described by the original source only as aggregates (totals) of multiple separate activities of fighting spanning over a longer period than a single, clearly defined day.</p> <p>Examples of such reporting: <i>"The Ukrainian government informs that 29 people have died in the past six days in a number of clashes with the separatists along the line of conflict"</i>.</p>	integer
date_prec	<p>How precise the information is about the date of an event.</p> <p>1: exact date of event is known;</p>	integer

	2: the date of the event is known only within a 2-6 day range.	
	3: only the week of the event is known	
	4: the date of the event is known only within an 8-30 day range or only the month when the event has taken place is known	
	5: the date of the event is known only within a range longer than one month but not more than one calendar year.	
date_start	The earliest possible date when the event has taken place.	Date YYYY-MM-DD
date_end	The last possible date when the event has taken place.	Date YYYY-MM-DD
deaths_a	The best estimate of deaths sustained by side a.	integer
	Always 0 for one-sided violence events.	
deaths_b	The best estimate of deaths sustained by side b.	integer
	Always 0 for one-sided violence events.	
deaths_civilians	The best estimate of dead civilians in the event.	integer
	For non-state or state-based events, this is the number of collateral damage resulting in fighting between side a and side b. For one-sided violence, it is the number of civilians killed by side a.	
deaths_unknown	The best estimate of deaths of persons of unknown status. Fatalities whose civilian/combatant affiliation could not be determined.	integer
best	The best (most likely) estimate of total fatalities resulting from an event. It is always the sum of deaths_a , deaths_b , deaths_civilians and deaths_unknown .	integer
high	The highest reliable estimate of total fatalities	integer
low	The lowest reliable estimate of total fatalities	integer
gwnoa	The Gleditsch and Ward number for Side A if the side is a state.	string(9999)
	Empty if Side A is not a state. In the API, unavailable values are returned as null.	
gwnob	The Gleditsch and Ward number for Side B if the side is a state.	string(9999)
	Empty if Side B is not a state. In the API, unavailable values are returned as null.	
geom / geometry	An Open Geospatial Consortium / ESRI binary representation of each individual point. Contains the SRID (4326) where supported.	geometry (Point,4326)

Due to the binary nature of this variable, this variable is contained only in the formats that support it.

4. Data Collection Methods

The UCDP GED is manually curated and compiled, with automatic assistance in data retrieval, filtering, data storage and manipulation, as well as data validation.

The original reporting underlying UCDP GED are collected from three sets of sources:

1. global newswire reporting
2. global monitoring and translation of local news performed by the BBC
3. secondary sources such as local media, NGO and IGO reports, field reports, books, X (twitter) feeds, social media such as Telegram etc.

The process is done in a "two-pass" system, first by consulting the above sources each month to provide candidate events for the entire globe (see appendix 5 for more details on sources). Then by consulting specialized sources based on information obtained from the first pass to find more information on incidents that can clear events or find more information on borderline cases.

Further, the UCDP GED is based on the same underlying data that all the other UCDP datasets are based on, i.e. UCDP GED is not built from e.g. the UCDP Dyadic Dataset, but rather both the UCDP Dyadic Dataset and UCDP GED are built from the same data.

4.1. First pass: Sources read each month

UCDP follows the media landscape closely and make an effort to change the sources read over the years to find the ones picturing the global organized violence the best. Whenever the program finds new information on old violence from sources such as truth commissions, UCDP makes an effort to include events from these sources. For a more detailed account of sources used in the 26.1 update, see appendix 5.

For the full dataset, Global newswire reporting including BBC Monitoring data has been sourced from the Dow Jones Factiva aggregator, using the following general search-string:

kill* or die* or injur* or dead* or death* or wounded or massacre*

This search is done globally, with "head and lead" and "intelligent indexing" being used for further filtering in those cases where this is feasible. In terms of sources, within the Factiva News aggregator, the most frequent global news sources used by UCDP are Reuters News, Agence France Presse (in English), Associated Press, Xinhua (in English) as well as BBC Monitoring. Note that for some of the years and geographic areas, reporting for some of the sources is extremely limited due to Factiva's non-inclusion of the whole corpus.

Similarly, media reporting is not consistent across time or space for any of the above-mentioned organizations. Changing managerial focuses, different organizational structures (such as field office locations), as well as different resource distributions and allocations (such

as, for example, the restructuring of BBC Monitoring in the early 2010s) make media reporting quality and quantity vastly different over various periods and over different areas. Furthermore, the ways in which conflicts are reported set the parameters for the preciseness of the data. In some countries and some phases of conflicts, the event data is based on either detailed daily reports or more summary-like reports covering larger areas.

Given this vast inconsistency in the source material generation process, we do not aim to be "source-consistent" - i.e. we do not aim to use the exact same set of sources to generate the entire dataset, since this would not provide anything else than a very poor convenience sample. The non-randomness in such a sample would be driven by news-agencies' procedures.

Instead, we drive our coding procedures with the set goal of obtaining a population dataset at the aggregate level (a complete yearly list) of all UCDP conflicts. This allows for consistency to be passed down to the event level and thus increases the quality and reliability of the sample presented in GED.

Thus, for many conflicts and periods we add secondary sources such as:

- local monitoring of various local media (e.g. Press Trust of India for India, or EFE news agency for Latin America or Radio Okapi for DR Congo),
- local monitoring and research organizations (such as SATP for India and Pakistan),
- global NGO reports (such as those coming from Human Rights Watch or Amnesty International),
- UN, EU, AU and other IGO reports,
- governmental publications (where considered reliable, such as those from Truth and Reconciliation Commissions),
- research articles or books etc.

The choice of whether to include secondary sources is made by project leaders and UCDP coders together. The goal of this inclusion is to further specify and identify conflicts (at the aggregate country-year level) in places where detail from global news sources is seen as insufficient. Thus, by attempting to equalize the level of detail identified across the board we make the sample more normalized for substantive analysis.

Coders are experts in the coding process. Unlike any other data collection project in the field, all coders are full-time long-term employees of UCDP, typically following conflicts and countries for long periods, and attaining in many cases specialist status in certain geographical areas. Further, UCDP consults external area specialists as to best select sources appropriate for the second pass.

4.2. Second pass: Specialized sources

During the second pass, UCDP make extra searches to find sources that give a more comprehensive understanding of the violence in a country. UCDP make extra searches in the Factiva news aggregator using more sources and other languages, google searches as well as look for further information from local NGOs, local news sources and social media accounts. These sources are mainly consulted to find more information to clear events already entered into the coding interface, where UCDP needs more details for the violence to meet our definitional criteria. The searches are also made for cases that almost reach our criteria and where a few extra fatalities would make the conflict active on the aggregate level. For some unclear cases UCDP

take contact with other area specialists for suggestions on sources or help with clarifications of the context surrounding certain events.

4.3. Further considerations for data validity and reliability

In general, the codebook and its appendices aim to contribute to improve, as much as possible, the reliability of the data, by presenting clear and consequent definitions as well as transparent coding procedures and rules.

The constructed precision codes for time, geography and event clarity, however detailed and elaborated, may allow for differing interpretations and understandings. Though coding rules and precision codes have been extensively discussed with researchers and tested in a pilot phase of the project during the summer of 2009, the process of constructing the geo-referenced event dataset is based on several procedures that may not always correspond to the reality of the events. For example, when constructing the dataset, the UCDP coders have, for pragmatic reasons, worked from the assumption that all events referring to the same start and end dates and 1 location represent an event clarity of 1. However, due to changing coding rules over a long period of time for the annual UCDP data, some of the dates as well as the included information are not as precise as others. This is especially true for the years 2002 and 2003 during which the UCDP experienced major structural rearrangements and improvements.

Further, differences in reporting may affect not only how much of the real world population of events is coded (see above), but also what detail level can be extracted. As such, for some countries, precise locations might be uncommon in reports on armed violence. There might even be a preference towards reporting violent activities on the first-order administrative level or less, which decreases the geographical precision.

In relation to this, the coders of the GED are experts on the coding procedure, yet seldom on the geographical dimensions of each conflict. This opens up for an error marginal where unclear location phrases such as “area” or “zone” can be misinterpreted. To address this challenge, the UCDP begins with studying the geographical and administrative structures for each new country to code.

4.4. Quality assurance

The dataset includes an extensive series of procedures to assure the quality and reliability of the data.

A large number of routines are set in place at this stage for quality control, each coder being given a fixed, comprehensive set of protocols to be followed that ensure the consistent treatment of, amongst other, dyad names, dyad IDs, precision scores, geocoding locations, streamlining of names, integrity of fatality estimates etc. The exact coding procedure is highly formalized, with all the steps of the process being given to coders, together with algorithms to insure the correctness of all the decisions that coders have to take.

Identified inconsistencies are resolved through regular, frequent meetings (at least once a week) where all the coders and project managers take part.

Third, over 50 automatic tests are applied to the data, followed by a series of manual checks by a project leader. Algorithms verifying that locations are properly geo-coded (through the

usage of density analyses and interpolation techniques), that ADM1s and ADM2s are properly identified and linked to events, that death estimates are properly used, that IDs are properly used and consistent with the aggregated datasets are done etc. Some of these tests are done at point of input, GED using a state-of-the art, custom built data-management, data-input and data-storage facility, others at release time.

Visualizations of the data are provided to coders and project managers using a Google Maps API derived solution.

The automated routines do not make any modifications to the datasets, requiring a human coder to make all the changes for an added level of security. The automated tests are re-run for as many times as required, until the data is deemed as acceptable for release by a project manager.

5. Additional information on variables in the dataset

The dataset contents can be divided into 7 categories: **event identifiers; actors and dyads; sources; geography; time; clarity; fatality figures.**

5.1. Event identifiers

This section provides unique identifiers for every event (row/entry) in the dataset. All variables in this section can be used as a unique key for the dataset.

id	A persistent unique numeric ID identifying each event. integer
	The same id number in versions 1.9 – 26.1 identifies the same real-life event (incident). This allows changes between versions to be traced at event level.

5.2. Actors and dyads

This section provides variables that allow for linkages between the UCDP GED and all other UCDP datasets.

This section also provides with variables to allow you to aggregate/filter/extract data on conflict, dyad or actor.

active_year	1: if the event belongs to an active conflict/dyad/actor-year 0: otherwise integer
type_of_violence	Type of UCDP conflict: integer 1: state-based conflict 2: non-state conflict 3: one-sided violence
conflict_new_id	A unique conflict identification code for each individual conflict in the dataset. integer UCDP Conflict ID for state based, non-state conflicts and one-sided violence as per the UCDP/PRIO Armed Conflict Dataset and UCDP Non-

<hr/> State Dataset and UCDP One-Sided dataset version 26.1.		
Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.		
This identifier is unique across the dataset (i.e. a non-state conflict cannot have the same identifier as a state-based conflict or a one-sided instance), irrespective of type of violence, and may be used for filtering and aggregation		
Warning: Not compatible with pre-17.1 versions of any UCDP datasets.		
conflict_name	Name of the UCDP conflict to which the event belongs. For non-state conflicts and one-sided violence this is the same as the dyad name.	string(9999)
dyad_new_id	A unique conflict identification code for each individual dyad in the dataset.	integer
UCDP Dyad ID for state-based conflicts, non-state conflicts and one-sided incidences as per the UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Datasets versions 26.1.		
Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.		
This identifier is unique across the dataset (i.e. a non-state conflict cannot have the same identifier as a state-based conflict or a one-sided instance), irrespective of type of violence, and may be used for filtering and aggregation		
Warning: Not compatible with pre-17.1 versions of any UCDP datasets.		
dyad_name	Name of the conflict dyad creating the event. A dyad is the pair of two actors engaged in violence (in the case of one-sided violence, the perpetrator of violence and civilians). The two sides are separated by an ASCII dash (e.g. Government of Russia - Caucasus Emirate, Taleban - civilians).	string(9999)
side_a_new_id	A unique ID of side A.	integer
Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.		
<hr/>		

	Warning: Not compatible with pre-17.1 versions of any UCDP datasets.	
	Note that this ID is no longer the Gleditsch and Ward number for State actors/sides. If you need that identifier, use gwnoa described below.	
gwnoa	The Gleditsch and Ward number for Side A if the side is a state.	string(9999)
	Empty if Side A is not a state.	
side_a	The name of Side A in the dyad. In state-based conflicts always a government. In one-sided violence always the perpetrating party.	string(9999)
side_b_new_id	A unique ID of side B.	integer
	Fully compatible with UCDP/PRIO Armed Conflict Dataset, UCDP Non-State Dataset and UCDP One-Sided Violence Dataset versions 17.1 and later.	
	Warning: Not compatible with pre-17.1 versions of any UCDP datasets.	
	Note that this ID is no longer the Gleditsch and Ward number for State actors/sides. If you need that identifier, use gwnob described below.	
gwnob	The Gleditsch and Ward number for Side B if the side is a state.	string(9999)
	Empty if Side B is not a state.	
side_b	The name of Side B in the dyad. In state-based always the rebel movement or rivalling government. In one-sided violence always "civilians".	string(9999)

An important note on IDs

Most UCDP IDs (side/actor, dyad and conflict) have been changed starting with version 17.1 to resolve some severe problems with non-unique, conflicting IDs in the previous scheme.

As a feature, the new system allows you to merge data across datasets without having to always take the "type of violence" variable into account.

Thus, the side/actor, dyad and conflict IDs in version 26.1 are no longer compatible with IDs used in versions of UCDP datasets prior to 17.1.

Some external datasets that use UCDP data for anchoring purposes (e.g. ACD2EPR) have been updated to make use of the new ID structure at time of writing; others have not (e.g. the Non-State Actor Dataset). Be careful against which version you are matching such datasets.

Translation tables are available between the current version of IDs and the old versions at <http://ucdp.uu.se/downloads/>. These should only be used if you need to use version 17.1-26.1 datasets with "older" datasets using the old UCDP ID structure (either produced by UCDP or produced externally) OR if you upgrade a dataset to the new version ID structure.

Note also that side IDs are no longer the Gleditsch and Ward Number for state actors. Use the GWNoA/GWNoB variable instead.

Note also that the dyad and conflict ID of government-perpetrated one-sided violence are no longer the ID of the perpetrating state. Use GWNoA instead.

Note also that the dyad and conflict ID of rebel-perpetrated one-sided violence are no longer the ID of the perpetrating actor. Use SideA instead.

Note also that the dyad and conflict ID of non-state conflict are no longer identical. Both are indicated in both UCDP GED and the UCDP Non-State dataset.

5.3. Description of sources

This section contains references to the sources underlying each event. See section 4.2 for a description of the data collection processes and source selection process.

The full texts of these sources are often copyrighted to news agencies/publishers. If you need to obtain access to the full text of reports, you will either need to re-download them from Factiva/Lexis Nexis or other relevant source provider.⁴

UCDP does not store the unique identifiers that Factiva, Reuters, AFP etc. assigns to an article, as during the decades-long data collection process we observed such identifiers change multiple times, making them useless for tracing source material directly.

Source publication dates may postdate the actual event by months or years. This is common for retrospective reporting, truth commissions, court proceedings, academic studies, or later investigative journalism. Sources published prior to the event may also be referenced when they help identify or attribute responsibility for the violence.

For events with multiple sources, `source_article`, `source_office`, `source_date` and `source_headline` store all values in a single string separated by semicolons (;). Commas within titles and names are preserved.

Example (API and CSV representation):

```
source_article: "Jane's Intelligence Review,2006-01-13,Internal politics complicate counterterrorism in Yemen";"Cedre,2004-05-02,Limburg"
source_office: Jane's Intelligence Review;Cedre
source_date: 2006-01-13;2004-05-02
source_headline: Internal politics complicate counterterrorism in Yemen;Limburg
```

Users who need one row per source should split these strings on ; in their processing scripts, being careful not to split on commas inside quoted segments.

number_of_sources	Number of total sources containing information for an event that were consulted.	integer
--------------------------	--	---------

⁴ For very small samples or original reports or information on individual events, you are welcome to contact us.

	<p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, -1. Note that -1 does NOT mean information on the source is missing; reference to the source material is ALWAYS available in the source_article field.</p>	
source_article	<p>References to the names, dates and titles of the source material from which information on the event is gathered.</p> <p>A reference to at least one source material is available for all events.</p> <p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For such older data, abbreviations such are used. The most frequent are:</p> <p>R: Reuters News, BBC: BBC Monitoring AP: Associated Press Newswires AFP: Agence France Presse, X: Xinhua DOW: Dow Jones Wires</p>	text
source_office	<p>The name of the organizations publishing the source materials.</p> <p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always available in the source_article field, for every event.</p>	text
source_date	<p>The dates the source materials were published.</p> <p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always available in the source_article field, for every event.</p>	text
source_headline	<p>The titles of the source materials.</p> <p>This variable is systematically available for all data collected from 2014 onwards, for most data collected in 2013, and for recently revised older</p>	text

	events. For older data, the field is empty. Note that an empty field does not mean information on the source is missing; reference to the source material is always available in the source_article field, for every event.	
source_original	The name or type of person or organization from which the information about the event originates in the original report.	string(9999)
	May be empty where coder did not identify original source.	
	e.g. "police", "Lt. Col. Johnson", "eyewitnesses", "rebel spokesman".	

5.4. Geography

Data in the UCDP GED is geo-referenced, meaning that each event is connected to a specific location defined by a pair of latitude and longitude coordinates.

Each event is connected to a single location. If reporting talks about multiple locations but gives only one aggregated fatality figure is given, then the following procedure is applied:

- one separate event is created for each location;
- deaths are split between locations as evenly as possible in order to maintain the fatality figures as integers. The split is performed automatically by the data management system⁵. The splitting procedure applies to low, best and high estimates independently.

The coordinates are fixed to the World Geodetic System of 1984 (WGS 84), EPSG SRID 4326. These coordinates are specified in decimal degrees with a precision of 6 decimal figures (e.g. 75.920221). Coordinates (latitude and longitude) used in the GED are based on the most precise location mentioned in the source.

The lowest level of spatial disaggregation for an urban location is the town, for the rural areas, the village.

Street, neighborhoods, parts of towns are not coded, even when such information is available in the reporting. Thus, a town is always represented by a single pair of latitude and longitude coordinates.

Suburbs, as long as they can be seen as separate urban areas, distinct from the main town, are coded as individual towns. Similarly, airports are always coded as separate entities.

Other features such as "mountains", "peaks" and "forests" are also used to specify geographical location, as long as their size is comparable (same order of magnitude) to those of towns or villages.

⁵ if insufficient deaths exist to create the required number of events (e.g. the reporting speaks of three locations and of only two dead), no split is performed. Instead, the smallest geographical unit encompassing all mentioned locations is used with an appropriate precision score.

The next lowest levels of spatial disaggregation are the administrative division of the country.

UCDP uses two levels administrative divisions for every country, the first-order administrative division (referred to as the ADM1) and the second order administrative division (referred to as the ADM2).

In the case of multiple, contested administrative systems (such as in Sri Lanka or Nagorno-Karabakh), UCDP uses the administrative system of the government controlling the capital of the country where the event takes place in.

The highest level of spatial aggregation for location is the country, defined using the Gleditsch and Ward list.

Further, all the geocoding is **time-aware**, i.e. locations are coded to the place-names and administrative divisions that were in place at the time the event took place. For example, an event that took place in 1989 in what is today St. Petersburg, Russia, is geocoded as happening in Leningrad, Soviet Union. Thus, changes in administrative structures of countries, as well as changes in borders are visible in UCDP GED.

The name of the location whose coordinates were assigned to the event is also provided in the `where_coordinates` field. It is the closest location to the event that could be identified and has a pair of known latitude and longitude coordinates.

`where_coordinates` is always streamlined - a latitude/longitude pair will only ever link to one `where_coordinates`. Further in `where_coordinates`, all capitals are referred to as “cities”, all urban localities other than capitals as “towns” (New York City Town is a correct name in `where_coordinates`), all rural localities as villages or localities etc.

5.4.1. Geo-referencing sources

UCDP does not employ an over-arching source for geocoding, as experience has proven that there is no quality global source for location data, especially for conflict zones and least-developed countries.

As such, UCDP coders employ sources such as global gazetteers (such as the United States National Geospatial Intelligence Agency’s GEOnet Names Server, Geonames, Google Maps, Google Earth, Mapcarta, Maplandia, GeoHack, openstreetmaps, Wikimapia or the Google Geocoding API), local maps provided by governmental authorities, data from the humanitarian Data Exchange, UN agencies (such as UN OCHA, UNDP) or local gazetteers, as well as, on occasion, historical maps such as the US Army Map Service Global Topographic Maps series.

Extreme care is taken to insure the full consistency, coherence and reliability of the data across the dataset. UCDP maintains both a repository of all the names previously geo-coded, as well as internal automated systems designed to insure that consistency (such as 1:1 matches between place-names and coordinates) is maintained throughout the dataset.

Information used to determine administrative divisions (labelled ADM1 and ADM2) stem from several different sources, commonly from a government’s own website or reference literature that covers administrative divisions globally. The global ISO 3166-2 standard is further used for identifying administrative divisions.

Note that while in most cases ADM1s are the largest administrative divisions in a country, in some cases (such as Russia or Romania) they are not, as the largest administrative division is either solely a statistical reporting unit or simply a legal fiction.

Correspondence regarding geographical coordinates, administrative divisions and any general questions or comments regarding the geographic aspects of the coding should be emailed to the maintainer of the dataset. Also, please report any potential errors in the dataset.

5.4.2. Geo-precision and its values

In order to determine the precision with which specific latitude and longitude coordinates are connected to an event location, the dataset uses a geo-precision variable. Precise coding rules and examples of how the geo-precision values are assigned in the GED can be found in the Appendix.

The geo-precision variable can have seven values:

1. Event can be related to an exact location, meaning a place name with a specific pair of latitude and longitude coordinates;
2. Event can be “near”, in the “area” of or up to 25 km away from an exact location, meaning a place name with a specific pair of coordinates, this can also be areas such as administrative divisions below ADM 2, forests or mountains or roads up to a 25 km radius;
3. Event can be related to a second order administrative division (ADM2), such as a district, municipality or commune
4. Event can be related to a first order administrative division (ADM1), such as a province, state or governorate;
5. Event can only be specified to a feature that is neither a known point nor a known formal administrative division, but rather a linear feature (e.g. a long river, a border or a road) or a fuzzy polygon without defined borders (informal regions, large radiuses etc.). A representation point is chosen for the feature and employed. Similarly, if a location is only known to be between two points, and these two points are more than 25 km apart, such locations are coded with geoprecision 5.
6. Event can only be related to the whole country;
7. Event can only be related to an estimated pair of coordinates at sea or in the air (provided the airplane did not crash as a result of the event; in such cases the location of the crash is coded with the appropriate precision code).

where_prec	Described above	integer
where_coordinates	Described above	string(9999)
adm_1	Described above	string(9999)
adm_2	Described above	string(9999)
latitude	Latitude (in decimal degrees)	numeric(9,6)
longitude	Longitude (in decimal degrees)	numeric(9,6)
geom_wkt	An Open Geospatial Consortium textual representation of the location of each individual point. Formatted as <i>well known text</i> without SRID.	string(9999)
priogrid_gid	The PRIO-grid cell id (gid) in which the event took place. Compatibility with PRIO-grid 1 and 2 (Tollefsen, 2012) is assured.	integer

Warning: We associate every point to the PRIO-GRID that contains it, even if the point is in another country than the one officially assigned to the respective PRIO-GRID cell through their majority area rule. It is your responsibility to make sure the covariates for the PRIO-GRID cell are correct for each event. Further, for the same reason, DO NOT, under any circumstances, first clip out (subset) PRIO-GRID by country before merging with UCDP GED as data loss will certainly occur. Refer to your copy of the PRIO-grid for further details on PRIO-GRID's majority assignment rule (p.3).		
country	Name of the country in which the event takes place. The country variable refers to the location where the event physically took place, not the nationality of the actors involved. Note that this variable differs from the country variable in the annual UCDP data, which registers the country of the incompatibility/actor.	string(999)
country_id	The Gleditsch and Ward numeric identifier of the country where the event takes place	integer
region	Region where the event took place. One of following: {Africa, Americas, Asia, Europe, Middle East}	string(999)

5.5. Clarity

This codes whether the reporting was sufficiently clear for the coder to be able to fully identify the event itself or not.

1 : (denoting high clarity): events where the reporting allows the coder to identify the event in full. That is, events where the individual happening is described by the original source in a sufficiently detailed way as to identify individual incidents, i.e. separate activities of fighting in a single location:

Example of such reporting: *"2 people where killed in Banda Aceh town on the 9th of December in fighting between the government and GAM when a car exploded in a main market."*

2 : (denoting lower clarity): for events where an aggregation of information was already made by the source material that is impossible to undo in the coding process. The coder merely has access to sources saying that events have taken place (and has aggregated fatality figures), but cannot break apart the reporting into constituent events.

Such events are described by the original source only as aggregates (totals) of multiple separate activities of fighting spanning over a longer period than a single, clearly defined day. Given that the report aggregates multiple incidents into one story impossible to disaggregate back, it is unclear how many battles took place during the time period specified in the source. Thus they are "secondary events", because the form of reporting does not allow the coder to know exactly when the casualties occurred, and how the battles were fought, and the event thus summarises a series of clashes into one event.

Of course, UCDP has a preference for events with a clarity of 1; events with a clarity of 2 are just a complement to the former. In fact, often times, it is possible, usually by corroborating multiple reports, to identify some of the clarity-1 events contained in the description making up the event with clarity of 2. In such cases fatalities in such identified events are subtracted from those given in the clarity-2 event. This leads to clarity-2 events sometimes defying the parameters of the fatality estimates, as the 'high estimate' may at times be lower than the 'best' or 'low' estimate.

Examples of **clarity-2 events**:

"The Ukrainian government informs that 29 people have died in the past six days in a number of clashes with the separatists along the line of conflict".

"in the past 2 months 120 people were killed in operations throughout Assam".

"The responsible for the Aceh military operation indicates that 29 people have been killed in various incidents of fighting over the past five days".

event_clarity	described above	integer
---------------	-----------------	---------

5.6. Time

Each event is defined to have occurred at a certain date. The precision of the dataset is **one calendar day**, starting at 00:00 (midnight) and ending at 23:59 local time.

In many cases, the exact day an event has taken place is impossible to find out with any certainty. In those cases, a temporal precision variable is provided which denotes with what accuracy a specific time period in which the event occurred is known.

The temporal precision variable can have five values:

1. the exact day of the event is known;
2. the exact day of the event is not known, only time period between 2-6 days;
3. the exact day of the event is not known, only the week;
4. the exact day of the event is not known, only the month;
5. the exact day of the event is not known, only the year.

date_prec	How precise the information is about the date of an event. 1: exact date of event is known; 2: the date of the event is known only within a 2-6 day range. 3: only the week of the event is known 4: the date of the event is known only within an 8-30 day range or only the month when the event has taken place is known 5: the date of the event is known only within a range longer than one month but less than one calendar year.	integer
date_start	The earliest possible date when the event has taken place.	Date YYYY-MM-DD

date_end	The last possible date when the event has taken place.	Date YYYY-MM-DD
-----------------	--	--------------------

5.7. Fatality figures

This section provides fatality figures for each event.

A note on civilian deaths: Civilian deaths can exist in all three categories of violence.

In state-based and non-state violence, civilian deaths count “collateral” killings, i.e. when one or more civilians are killed as an effect of fighting between the two warring parties. At times, such fighting may even result in only the civilian bystanders receiving fatal injuries. Similarly, imprecise shelling or bombing in the context of an armed conflict is coded as state-based violence unless it is clear (from either reporting or context) that civilians have been explicitly targeted.

In one-sided violence, the targeted and killed civilians are always registered in the `deaths_civilians` column.

deaths_a	The best estimate of deaths sustained by side a.	integer
	Always 0 for one-sided violence events.	
deaths_b	The best estimate of deaths sustained by side b.	integer
	Always 0 for one-sided violence events.	
deaths_civilians	The best estimate of dead civilians in the event.	integer
	For non-state or state-based events, this is the number of collateral damage resulting in fighting between side a and side b. For one-sided violence, it is the number of civilians killed by side a.	
deaths_unknown	The best estimate of deaths of persons of unknown status.	integer
best	The best (most likely) estimate of total fatalities resulting from an event. It is always the sum of deaths_a , deaths_b , deaths_civilians and deaths_unknown .	integer
high	The highest reliable estimate of total fatalities	integer
low	The lowest reliable estimate of total fatalities	integer

5.8. Variables present in previous versions of GED not used in version 26.1

where_location (GED 1.0 – 1.9) and **where_description** (GED 2.0 – 5.0): removed as two separate textual fields for the name of the same location was both confusing and made the dataset both more unwieldy to users.

***_dset_id (GED 5.0):** removed as ID systems for all UCDP datasets have been unified. A translation table between old and new ids is available at : <http://ucdp.uu.se/downloads/>

coordinate_location (GED 1.0 – 1.9) : renamed to **where_coordinates**

event_type (GED 1.0 – 1.9): given the high level of confusion experienced by our users with regards to the actual meaning of the variable, it was replaced by a new concept, **event_clarity**.

Geocomment (GED 1.0 – 1.9): eliminated as a human-legible free-text comment on location names, geocoding sources and alternative spellings proved to have more disadvantages than advantages for data usability in a large, quant-oriented dataset.

dyad_unique (GED 1.5): replaced by **dyad_new_id**. **Dyad_unique** was a temporary, stop-gap measure specific to GED 1.5 to prevent an acute problem originating from the merging of three separate systems. The construction of a new UCDP system together with introduction of an UCDP-wide system

uniq (GED 1.1 – GED 1.5): replaced by **id**. Compared to **uniq**, which was specific to each release of the data, **id** is persistent, i.e. consistent across releases of the datasets. An entry with the same **id** in version 1.9 describes the same real-life incident in versions 2.0 – 5.0.

6. Format availability

The UCDP GED is provided in a variety of formats for use by researchers within different fields and with different needs. All formats are available for download free of charge (no registration required) from the UCDP GED website (<http://ucdp.uu.se/downloads>).

The UCDP GED versions 5.0, 17.1, 17.2, 18.1, 19.1, 20.1, 21.1, 22.1, 23.1, 24.1 and 26.1 as well as monthly candidate events are also provided as an API service, allowing direct machine-to-machine communication (i.e. you can download the data, as well as filter and subset it on our servers directly from your script, do file, library etc. without downloading it to your own system). Documentation is available here: <http://ucdp.uu.se/apidocs/>. The API returns data in JSON format.

The GED full download is currently available in the following formats:

Comma Separated Values (CSV), Excel (XLSX), R Data Frame (Rds), STATA (dta),

A summary is provided here to help users understand each file format and its compatibility as well as to quickly get started:

- **CSV format:** A plain text file containing structured comma separated values. The file is suitable for usage with statistics packages, for processing with various programming languages, etc.

Note that the format implements the full CSV specifications as summarized in RFC4180⁶. The full CSV specifications are properly implemented by a large number of software packages, including *OpenOffice*, *Stata (using the insheet command or the menu)*, *SPSS*, *R (using the read.csv function)*, *PHP (using the fgetcsv and fputcsv functions)*, *Python (using the csv module)* etc.

Note that this file does not contain the **geom** variable as it is a plain text file.

⁶ The RFC, summarized, states that data is stored in UTF-8, column names are listed in the first row, lines (rows) are terminated in the Windows new-line system (CR LF), fields (record/cells) are separated with a comma (','), with each field containing a space, a double apostrophe (""), a comma (,) or a new line (CRLF) being enclosed within double apostrophes ("""). Double apostrophes are escaped by double apostrophes (""").

- Excel format: An Excel compatible file that can be used for visualizing the data in a simple Office-like system.
Note that this file does not contain the **geom** variable as it is a plain text file.
- Rds format: An R dataframe.
- STATA format: DTA file.

UCDP GED data can also be downloaded as a web-service through a RESTful API under the DaaS paradigm. Descriptions on how to use the UCDP GED API are available at <https://ucdp.uu.se/apidocs/>.

7. References

Boschee, E.; J. Lautenschlager, S. O'Brien; S. Shellman; J. Starz, James; M. Ward (2015), "*ICEWS Coded Event Data*", <http://dx.doi.org/10.7910/DVN/28075>, Harvard Dataverse, V7.

Deborah J. G., P. A. Schrod, O. Yilmaz (2009). *Conflict and Mediation Event Observations (CAMEO) Codebook*. <http://eventdata.psu.edu/data.dir/cameo.html>, 2009.

Raleigh, C., A. Linke, H. Hegre, J. Karlsen (2010), Introducing ACLED: An Armed Conflict Location and Event Dataset, *Journal of Peace Research* vol. 47(5):651-660.

Rød, E. G., N. B. Weidmann (2014), *Protesting Dictatorship: The Mass Mobilization in Autocracies Database*, APSA Annual Meeting in Washington, DC, 2014.

Salehyan, I., C. Hendrix, J. Hamner, C. Case, C. Linebarger, E. Stull, J. Williams (2012). Social conflict in Africa: A new database. *International Interactions*, 38(4):503-511.

Schrod, P.A., J. Beiler and M. Idris (2013), *Three's a Charm? Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance*. Paper presented at the International Studies Association Annual Convention, Toronto, March 2014.

Schrod, P.A. Twenty years of the Kansas event data system project. *The Political Methodologist*, 14(1):2-8.

National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2013). *Global Terrorism Database*. Retrieved from <http://www.start.umd.edu/gtd>

Also consult the following UCDP Codebooks:

- UCDP Dyadic Dataset Codebook
- UCDP/PRIO Conflict Dataset Codebook
- UCDP Non-State Conflict Dataset Codebook
- UCDP One-Sided Dataset Codebook
- UCDP Actor Dataset Codebook

Appendix 1 UCDP definitions

Definition of state-based armed conflict

UCDP defines state-based armed conflict as: “a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year.”

The separate elements of the definition are operationalized as follows:

- (1) *Use of armed force*: use of arms, resulting in deaths.
 - (1.1) *Arms*: any material means, e.g. manufactured weapons but also sticks, stones, fire, water etc.
- (2) *25 deaths*: A minimum of 25 battle-related deaths per year and per dyad (see Item 3.3 in this definition) in an incompatibility.⁷
- (3) *Party*: A government of a state or any opposition organization or alliance of organizations. UCDP distinguishes between primary and secondary parties. Primary parties are those that form an incompatibility by stating incompatible positions (see Item 5 in this definition). At least one of the primary parties is the government of a state. Secondary parties are states that enter a conflict with troops to actively support one of the primary parties. The secondary party must share the position of the primary party it is supporting in the incompatibility.
 - (3.1) *Government*: The party controlling the capital of a state.
 - (3.2) *Opposition organization*: Any non-governmental group of people having announced a name for their group and using armed force to influence the outcome of the stated incompatibility (see Item 5 in this definition). The UCDP only deals with formally organized opposition. The focus is on armed conflict involving consciously conducted and planned political campaigns rather than spontaneous violence.
 - (3.3) *Dyad*: A dyad consists of two conflicting primary parties. At least one of the primary parties must be the government of a state. In interstate conflicts, both primary parties are state governments.⁸ In intrastate and extrasystemic conflicts, the non-governmental primary party includes one or more opposition organization(s). A conflict can include more than one dyad. If e.g. a government is opposed by three rebel groups over the same incompatibility, the conflict is made up of three dyads. Note that secondary parties (i.e. intervening states supplying troops to one of the primary parties) do not lead to the formation of additional dyads.
- (4) *State*: A state is an internationally recognised sovereign government controlling a specific territory or an internationally unrecognised government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.

⁷ Note that an incompatibility involving two opposition groups, each involved in clashes with the government resulting in 20 deaths, would not be recorded as a conflict (neither dyad reached the minimum casualty threshold), whereas an incompatibility involving a single opposition group that caused 25 battle-deaths would be included in the dataset.

⁸ With three exceptions, the primary parties in interstate conflicts consist of only one actor on each side (and thus only constitute one dyad). See description of Location (Section 4.1) below for further information on these cases.

- (5) *Incompatibility concerning government or territory*: The incompatibility, as stated by the parties, must concern government and/or territory.
 - (5.1) *Incompatibility*: The stated general incompatible positions.
 - (5.2) *Incompatibility concerning government*: Incompatibility concerning type of political system, the replacement of the central government, or the change of its composition.
 - (5.3) *Incompatibility concerning territory*: Incompatibility concerning the status of a territory, e.g. the change of the state in control of a certain territory (interstate conflict), secession or autonomy (intrastate conflict).

As a country can experience several simultaneous conflicts, it is essential to differentiate between them, UCDP collects information on conflicts where the incompatibility, i.e. the general incompatible positions, concerns either government or territory or both. Note that the incompatibility expressed in terms of government or a specific territory is crude in the sense that possible underlying incompatibilities are not considered. In other words, the stated incompatibility is *what* the parties are (or claim to be) fighting over, but it says nothing about *why* the parties are fighting. However, classifying incompatibilities as stated manifestations of possible underlying incompatibilities or goals allows for the intersubjective compilation of the list of armed conflicts.

While a state can only experience one intrastate conflict over government in a given year, that same state can simultaneously be a primary party to one or more interstate conflicts over government and/or territory. In the case of intrastate territorial conflicts, multiple conflicts can be recorded over different territories in a state in a given year.

Definition of non-state conflict

A non-state conflict is defined by the Uppsala Conflict Data Program (UCDP) as “*the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year.*”

The separate elements of the definition are operationalized as follows:

- (1) *Use of armed force*: the use of arms, resulting in deaths.
 - (1.1) *Arms*: any material means, e.g. manufactured weapons but also sticks, stones, fire, water, etc.
- (2) *25 deaths*: a minimum of 25 battle-related deaths per year
 - (2.1) *battle-related deaths*: deaths directly related to the use of armed force between the warring groups
- (3) *Organized groups*: consists of either
 - (3.1) formally organized groups: any non-governmental group of people having announced a name for their group and using armed force against another similarly formally organized group, *or*
 - (3.2) informally organized groups: any group without an announced name, but who uses armed force against another similarly organized group, where the violent activity meets the following requirement:

(3.2.a.) there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other

(4) *State*: a state is

(4.1) an internationally recognized sovereign government controlling a specified territory, *or*

(4.2) an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.

(5) *Government*: the party controlling the capital of the state

Definition of one-sided violence

One-sided violence is the use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths. Extrajudicial killings in custody are excluded.

The separate elements of the definition are operationalized as follows:

(1) *Use of armed force*: use of arms in order to exert violent force, resulting in death

(1.1) Arms: any material means, e.g. manufactured weapons but also sticks, stones, fire, water, etc

(2) *25 deaths*: a minimum of 25 civilian deaths per year and per actor

(3) *Government*: the party controlling the capital of the state

(4) *Formally organized group*: any non-governmental group of people having announced a name for their group and using armed force

(5) *State*: a state is

(5.1) an internationally recognized sovereign government controlling a specified territory, *or*

(5.2) an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.

(6) *Civilians*: unarmed people who are not active members of the security forces of the state, or members of an organized armed militia or opposition group. Government officials, such as members of parliament, governors, and councilors, are also excluded and are instead seen as representatives of the government of a state.

Note that the UCDP definition of civilian does not refer to the definition of civilian under international law.

(7) *Extrajudicial killings in custody*: when the government of a state kills a person in its custody

(7.1) *custody*: when the person is located in a prison or another type of governmental facility

Appendix 2 Temporal precision and date estimation rules

This document specifies the qualifications for all temporal precision variable values according to the rules constructed by the UCDP for the GED. It also sets rules for interpretation of time-related expressions and estimation of events' start and end dates. The appendix presents concrete examples that guide temporal precision coding and date estimation procedures.

Estimation of Start and End Dates

1. Start and end dates of the events are set according to information in the original sources.
2. Ambiguous time-related expressions (e.g. past few days) are interpreted on the basis of the rules presented below. This ensures uniform estimation of the events' start and end dates throughout the entire GED.
3. If the source does not provide any information about the time period during which the event took place, dates are estimated for three days, counting backwards from the day of reporting:
 - a. "24 rebel soldiers were killed";
 - b. "Security forces stepped up operations against the largest insurgent group in Assam state, where a new government was set to take charge on Friday. A police spokesman said four members of the outlawed ULFA were killed in the battles";
 - c. "10 bodies found buried in a mass grave in territory controlled by the ULFA rebels".

Temporal Precision 1 – Daily Precision of Time

1. If the exact date of an event is known the temporal precision code of 1 is applied. Such events have the same start and end dates that are precisely specified in the news sources either by dates, day names, hours or other specific temporal concepts:
 - a. "14th January", "today", "yesterday", "last Tuesday" - date for specified day;
 - b. "Monday night" - date for Monday;
 - c. "Last night" - date for preceding day of reporting;
 - d. "The other day" - date for the preceding day of reporting.

Temporal Precision 2 – Imprecise Time (2-6 days)

1. Temporal precision value of 2 should be used in those cases when start and end dates for events are of unspecified character, spanning more than one calendar day though no longer than six days, i.e. shorter than a week:
 - a. "Recently", "recent attacks" - dates for 3 days preceding and not including the day of reporting;

- b. "Past/last few days" - dates for 3 days preceding and not including the day of reporting;
- c. "Around 2 July" - dates for three days, 1-3 July, with the stated date +/- one calendar day;
- d. "Over the weekend" - dates for Saturday and Sunday, if source does not include Friday in the concept of weekend and unless specific dates/days for the weekend are provided in the source;
- e. "Since the beginning of the week", "this week" - dates from Monday to the day of reporting;
- f. "Night between Sunday and Monday" - dates for 2 days;
- g. "Past 24 hours" - dates for the day of reporting and the preceding day;
- h. "Past 48 hours" - dates for the day of reporting and 2 preceding days;
- i. "Past 72 hours" - dates for the day of reporting and 2 preceding days;
- j. "Past 2 days" - dates for 2 days preceding and not including the day of reporting;
- k. "Since Thursday" – dates from Thursday until the day of reporting;
- l. "Five-day offensive" - dates for 5 days of fighting including the day of reporting;
- m. "Continuous fighting between 13-16 February" - specified dates;
- n. "Night-long battle" - dates for 2 days covering the whole night;
- o. "Night of clashes" - dates for 2 days covering the whole night;
- p. "Last 6 days of January" - dates for 25-30 January, including final date of month;
- q. "Late last week" - dates for Friday to Sunday of the preceding week.

Temporal Precision 3 – Weekly Precision of Time

1. Temporal precision value of 3 should be used in those cases when start and end dates for events are specified to a certain week, but specific dates are not provided:
 - a. "Last week" - dates for Monday-Sunday of the preceding week. Exceptions can be made if there are reasons to believe that the event took place during the week of the reporting (e.g. sometimes "a raid last week" reported on Sunday might refer to the period Monday-Saturday of the same week, then dates for Monday-Saturday of that week should be used);
 - b. "Past week" - dates for 7 days including the day of the reporting, unless text indicates that past week refers to an ongoing week (starting Monday);

- c. "First week of August" - dates for August 1-7.
- d. "Week-old offensive" - dates for a week of fighting, 7 days, including the day of reporting;

Temporal Precision 4 – Monthly Precision of Time

1. Temporal precision value of 4 should be used in those cases when start and end dates for events are specified to a certain month, but specific dates are not provided:
 - a. "Beginning of/early March" – March 1 to March 10/day of reporting;
 - b. "Middle of March" – March 15 +/- 5 calendar days, i.e. March 10-20;
 - c. "End of/late March" – March 15 to the last day of March/day of reporting;
 - d. "A number of weeks", "recent weeks" - dates for 3 weeks counting backwards from the day of reporting;
 - e. "Several weeks" – dates for 3 weeks;
 - f. "Earlier this month" – starting the 1st day of the month and ending on the day preceding the day of reporting;
 - g. "Last month" - dates for the month preceding the one on which the event was reported;
 - h. "A fortnight ago" - dates for preceding 14 days including the day of reporting.

Temporal Precision 5 – Annual Precision of Time

1. Temporal precision value of 5 should be used in those cases when start and end dates for events are specified to a certain year, but specific dates are not provided:
 - a. "1995" - 1995-01-01 to 1995-12-31;
 - b. "Last year" - dates covering the year, YYYY-01-01 to YYYY-12-31;
 - c. "Past year" – All dates from the date of reporting back to YYYY-01-01
 - d. "Early 1999" – 1999-01-01 to 1999-04-30;
 - e. "Mid 1999" – 1999-05-01 to 1999-08-31;
 - f. "Late 1999" – 1999-09-01 to 1999-12-31;
 - g. "Past 3 months" - dates for 3 months counting backwards from the day of reporting (may not cross over into another calendar year);
 - h. "Past few months" – dates for 3 months counting backwards from the date of reporting (may not cross over into another calendar year).

Appendix 3 Geo-precision coding rules

This document gives an overview of the coding rules for geo-precision codes coupled with examples and comments.

General rules

1. All geographical locations are coded with moderation with preference given to more certain locations even if they represent a higher level of aggregation over those locations which are less certain but represent a lower level of aggregation.
2. Unclear geographical references with several possible levels of aggregation are coded as the highest possible one. For instance, if there is a town, a district (ADM2) and a province (ADM1) of the same name and the source does not specify to which type of location it refers, then the location will be coded as ADM1.
3. If event location (camp, bridge, road etc.) has the same name as a certain suburb, town or village (e.g. Uppsala IDP camp and Uppsala town), the coordinates for that town or village should be used only if it is known that the event location is within or close to (within 25 km) that town or village. If information about the locations' proximity to that town or village is not available, the location is aggregated to the lowest available administrative division. For instance, if it is not known that Uppsala IDP camp is within 25 km from Uppsala town, coordinates for Uppsala municipality (ADM2) should be used.
4. If the source refers to a certain location (e.g. river, forest, lake, park, mountains etc.) that is not similar in size with a locality, or that is not a point, a representation point is created with precision 5. If that location lies within an ADM2 or ADM1, the ADM2 or ADM1 is attached to the representation point. Do not aggregate e.g. rivers or national parks to administrative divisions if representation points can be made.
5. When coding historical observations the GED uses the names of the administrative divisions in force at the time of the reporting. If the boundaries of ADM1 have changed over time in a country, the dataset uses estimated coordinates for older provinces based on the relevant seat of the ADM1 at the time of the event.

A history of administrative changes is tracked internally by the UCDP system in a data structure referred to as a geotree. If you require access to such files, contact us.

Geo-precision 1

Geo-precision value of 1 is used if the location information corresponds exactly to the geographical coordinates available. Each pair of coordinates is also coupled with names for ADM1 and ADM2 when available.

1. "City", "town", "village", "location", "locality"- centroid point coordinates;
2. "District", "quarter", "neighbourhood", "locality" (of town) - coordinates for town centroid point are applied here, and not the specific section of it, though the name and details are kept in text in parenthesis in "Where";
3. Air battles if location is clear, i.e. "a plane was shot down over Kitgum".

Geo-precision 2

If the location information refers to a limited area around a specified location, coordinates for that location together with the geo-precision value of 2 are used.

1. “Near/in the vicinity of/adjacent to/just outside/around Kitgum town” – coordinates for Kitgum town;
2. “Pietermaritzburg area” – coordinates for Pietermaritzburg town;
3. “Outskirts/suburbs of Bujumbura city” – since outskirts and suburbs are understood as relatively independent and distant entities coordinates for Bujumbura city should be used;
4. “17 km from Uppsala town” – if the event takes place within a distance of 25 km from a specified location, coordinates for that specified location are used;
5. “North of Luanda city”, “southeast of Y mountain” - unspecified distances from a specified location are understood to be near the stated location;
6. “Bujumbura city towards Gishingano village” – if coordinates for Gishingano village can not be retrieved, then coordinates for Bujumbura city will be used;
7. “Niuland village near Dimapur town” - if coordinates for Niuland village are not available, but coordinates for Dimapur town exist, the latter are used;
8. “Dungu territory in DRC” – third level administrative divisions (ADM3), if small enough to have an approximate radius of 25 km or less, receive a precision code of 2.

Geo-precision 3

If the source refers to or can be specified to a larger location at the level of second order administrative divisions (ADM2), such as district or municipality, the GED uses centroid point coordinates for that ADM2. If these are not available, representation coordinates for a town within that ADM2 are used. The name of the ADM2 in force at the time of reporting is recorded in the variable ADM2.

1. “Arusha district, Arusha province” - coordinates for Arusha district (ADM2);
2. “Burambi commune, Burundi” – coordinates for Burambi commune (ADM2);
3. Air battles if unclear location - if the battle takes place “over” a certain ADM2, coordinates for that ADM2 will be used;

Geo-precision 4

If the location information refers to a first order administrative division, such as a province (ADM1), the GED uses the coordinates for the centroid point of ADM1.

1. “Cibitoke province, Burundi” – coordinates for Cibitoke province (ADM1);
2. Air battles if unclear location - if the battle takes place “over” a certain ADM1, coordinates for that ADM1 are used;
3. If the ADM2 in which the event took place is unclear (e.g. different sources refer to different ADM2s in which the same event took place), the location is aggregated to the ADM1 level;

Geo-precision 5

Geo-precision value of 5 is used in these cases:

1. If the location information refers to parts of a country which are larger than ADM1, but smaller than the entire country such as “Southern Lebanon”, “Northern Uganda”. In these cases, a representation point is created for that part of the country and used as a representation of that area together with geo-precision value of 5. Note that these points are stored and reused consistently by the UCDP (thus, all events assigned to “Northern DR Congo” will have the same coordinates recorded).
2. If a pair of coordinates is estimated as a representation point for a linear, non-administrative polygon or fuzzy geographic feature (river, informal area, large lake etc.). For example, if the location is on the border between two countries and the location of such point is not precisely known, a pair of estimated coordinates will be used together with geo-precision value of 5. For example, “on the border between Uganda and Sudan” will be coded as “Uganda/Sudan border” with the coordinates for a selected point on the border between Uganda and Sudan; Note that these points are stored and reused consistently by the UCDP (thus, all events assigned to “Uganda/Sudan border” will have the same coordinates recorded).
3. If the location information refers to islands which are not an ADM1 or 2 of their own. For example, “Zanzibar island” will be understood as eastern part of Tanzania and receive geo-precision value of 5. If a pair of coordinates for that island is not available in the gazetteers, it can be represented by an ADM1 in that island.
4. If the location is not specific and need to be estimated (for example, “road between Pader and Kitgum”, “along Aswa river” etc.), or the location is more than 25 km away from another location (for example, 75 km south of Kitgum town), then a representation point is created for that point. This is done even if the two points are located in the same ADM2. As such, if an event is described as taking place on “the road between Yei and Rasul in Yei district of Equatoria State”, then a point is estimated on that road, with precision 5, with both the ADM1 (Equatoria state) and the ADM2 (Yei district) coded.

Geo-precision 6

If the location information refers to an entire country, centroid point coordinates of that country are used. Also, if the location is not provided/is unclear/refers to several locations which cannot be split and covers the whole country and a particular activity area of the actor is not clear, centroid point coordinates of that country are used.

1. "Germany" - centroid point coordinates;

Geo-precision 7

If the event takes place over water or in international airspace, the geographical coordinates in the dataset either represent the centroid point of a certain water area or estimated coordinates according to similar techniques as presented above for geo-precision code 5.

For air events, precision code 7 is used only if the death is not the effect of or did not result in the airplane crashing (in such a case, 1-5 precision codes are used with the location of the crash).

1. "Southern ocean" – centroid point coordinates;
2. "Bay of Bengal" – centroid point coordinates;
3. "37 km off the coast from Stockholm city" – estimated coordinates for a point 37 km and 90 degrees off the coast of Stockholm.
4. "the minister was stabbed on an airplane en route to Delhi after departing Islamabad" – coordinates for Islamabad airport, precision code 7.

Appendix 4 Comparison with other event ontologies

The **UCDP GED** is an "**incident dataset**", sharing a highly similar conceptualization of "events" with datasets such as the Global Terrorism Dataset (START, 2013), ACLED (Raleigh, 2009) or SCAD (Salehyan, 2012) in the sense that each entry represents an **incident**, a real-life series of actions circumscribed to a certain typology and resulting in a certain outcome or set of outcomes. In the case of UCDP GED this typology is: **fighting resulting in the death of at least one person**.

This differs markedly from the conceptualization of events in most "machine" datasets and coding systems such as PHOENIX/EL:DIABLO (OEDA, 2014), CAMEO/TABARI (Gerner et. al, 2014) , ICEWS/JABARI (Boschee et. al., 2015) or KEDS (Schrodt, 2006). In these datasets, an event represents an action between two actors (e.g. taunt, attack, fight, retreat, mediate etc.). Since typically multiple actions lead to an outcome (an incident) a UCDP GED event is equivalent to a collection or aggregation of multiple events in such "machine" datasets. Further, as compared to these datasets, UCDP GED is more geared towards extracting information relating to the incident (such as fatality figures) rather than binning actions into sets of categories.

Further, UCDP GED differs markedly to event data collection efforts where the event is defined as the individual source report, since UCDP GED collates information from all sources referring to the same information and interprets all of them in order to extract information and define an event. Thus, UCDP GED would be a filtered version of such a dataset, with filtering performed according to certain interpretation rules, uncertainty management and definitions.

Of course, this comparison only describes the conceptual differences between our conceptualization of "events" and other conceptualization of "events" for users either more familiar with other datasets or users wanting to use UCDP GED together with other event-type data. As such, no other event dataset replicates or is replicated by UCDP GED as UCDP GED collects information on a specific, unique type of social behaviour (as defined above).

Appendix 5 Main sources consulted during the 26.1 update

Africa

Algeria, Tunisia, Morocco, Western Sahara

Factiva sources: Agence France Presse, All Africa, Associated Press Newswires, BBC Monitoring Africa, BBC Monitoring Media, BBC Monitoring Middle East, Reuters (Full article)

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Algeria Ministry of National Defence, ANSTRAT Tunisia, Terrorism Monitor, UN Mission reports to Western Sahara

Angola, Republic of Congo, Zimbabwe, Uganda, Rwanda

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Newsfile, Reuters, Factiva Portuguese search (“matar”, “morrer”, “ferir”, “morto”, “morte”, “ferido”, “massacre”, “FLEC”)

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Angola24Horas.com, Cabinda Skyrock, The Independent (Uganda), Uganda Police Force, VOA Português, Zimbabwe Human Rights NGO Forum

Social Media sources: X: Caleb Weiss (@caleb_weiss7), FlecFac, JClaudeNzita

Benin, Cape Verde, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea Bissau, Liberia, Sierra Leone, Togo, Cote d'Ivoire, Senegal

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Newsfile, Reuters, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group, Terrorism Monitor

Extra sources area or country specific: Dakaractu, Senepius

Social Media sources X: Caleb Weiss, EliasuAlhaji, Kris à Ouaga, Nadia Nata, Wamaps

Botswana, Comoros, Lesotho, Madagascar, Malawi, Mauritius, Namibia, South Africa, Zambia

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, Cape Argus, Cape Times, News24 Online (South Africa), Times of Swaziland, The Swazi

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: GI-TOC South Africa Organized Crime Observatory, Swaziland News, The Bridge

Burkina Faso, Mali, Mauritania

Factiva sources: Agence d'Information du Burkina, Agence France Presse, All Africa, Associated Press Newswires, BBC Monitoring Africa, BBC Monitoring Middle East, Reuters, RFI (French), Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: CTC Sentinel, CPS spokesman, Echoes of the Sahel, Kal Akal, Malian Armed Forces website, Terrorism Monitor

Social Media sources: X: @18Echo0321, @abousaib, @AESinfos, @ag_anara, @AgMohamed60, @BrantPhilip, @DirpaFa, @FLAZAWAD, @IlyasseAg241786, @inkinane1, @Intelligence225, @lavoixdelazawad, @Mossa_ag, @ocisse691, @simnasr, @walid_leberbere, @Wamaps_news, @WerbCharlie, Facebook: mohamed.ramadane.56, Facebook: naim.toure, Telegram: @NaimInfo2

Burundi

Factiva sources: All Sources

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: ARIB.info, Burundi Human Rights Initiative, Igihe, IWACU, Ligue Iteka, Radio Publique Africaine, SOS Medias Burundi, SOS Torture Burundi, UNHRC Commission of Inquiry on Burundi

Social Media sources: X: red_tabara

Cameroon, Niger, Nigeria

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Newsfile, Journal du Cameroun, News Agency of Nigeria, Premium Times (Nigeria), Punch (Nigeria), Reuters, Vanguard (Nigeria)

Extra Sources: Africa Confidential, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Anglophone Crisis Timeline, CTC Sentinel, Echoes of Sahel, HumAngle, Nigeria Watch, Terrorism Monitor

Social Media sources: X: @AESinfos, @BrantPhilip, @hamid_gade, @MimiMefo, @SembeTV, @Wamaps_news, @WassimNasr, @WerbCharlie, @ZagazolaMakama

Central African Republic

Factiva sources: All Sources

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Corbeau News, Humanitarian Response documents and data, LRA Crisis Tracker, MINUSCA reports, OCHA weekly, Radio Ndeke Luka, RJDH, UN Panel of Experts reports, UN Secretary-General reports

Chad

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Newsfile, Reuters, Alwihda (French), Jeune Afrique (French), Journal du Tchad (French), RFI.fr (French), Tchadinfos (French)

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: CTC Sentinel, Terrorism Monitor

Social Media sources: X: @Wamaps_news

Democratic Republic of Congo

Factiva sources: All Sources

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Actualite.cd, Bunia Actualité, Congo Research Group, Kivu Morning Post, LRA Crisis Tracker, MediaCongo.net, MONUSCO reports, Radio Okapi, UN Group of Experts Reports

Social Media sources: X: @afcongo, @CNangaa, @DRCNewsToday, @LawrenceKanyuka

Eswatini

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, News24 Online (South Africa), Times of Swaziland, The Swazi, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Swaziland News, The Bridge

Ethiopia, Eritrea, Djibouti

Factiva sources: Addis Standard, Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Middle East, Reuters, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Amhara Association of America's weekly "War Updates from the Amhara Region, Ethiopia", EEPA Situation Reports Horn, Ethiopian Human Rights Commission (EHRCO), OCHA Ethiopia Humanitarian Access Situation Reports, OLLAA Exclusives, QBO-ABO-WBO, "My Views on News" YouTube channel

Social Media sources: X: @odaatarbiwbo, @oromiacmc, Facebook: fdredefense.official

Kenya, Somalia

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, Reuters, The Star Kenya, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Africa Center, Critical Threats “Gulf of Aden Security Review”, CTC Sentinel, Goobjoog News, Halbeeg News, ISS Today, Long War Journal, Panel of Experts on Somalia, Secretary-General Reports (Somalia), Somali Digest “Security & Conflict”, Somaliland Standard, Somaliland Sun, Strategic Intelligence, Terrorism Monitor

Libya

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Central Asia, BBC Monitoring Europe, BBC Monitoring Middle East, BBC Monitoring Newsfile, Libya Herald, Reuters, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Airwars, alsaaa24, CTC Sentinel, Eye on ISIS in Libya, Libya24, Libya Observer, Libya Security Monitor, Terrorism Monitor, UNSMIL Human Rights Report on Civilian Casualties, UN Secretary-General Reports

Mozambique, Tanzania

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, Club of Mozambique, Deutsche Welle, Mozambique News Agency, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Cabo Ligado, Carta de Mocambique, CTC Sentinel, Terrorism Monitor

South Sudan, Sudan

Factiva sources: Agence France Presse, All Africa, BBC Monitoring Africa, BBC Monitoring Middle East, Reuters, Sudan Tribune, Xinhua

Extra Sources: Africa Confidential, Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Africa Centre for Justice and Peace Studies (ACJPS), CTSAMVM reports, HUDO Centre, IOM Sudan Regional Crisis Situation Updates, LRA Crisis Tracker, OHCHR, ReliefWeb/OCHA, reports from UN missions in the countries, Small Arms Survey, Sudan Conflict Observatory, Sudan Shahid map, Radio Dabanga

Americas

Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Jamaica, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Sao Tome and Principe, Trinidad and Tobago

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Americas, EFE News Service, Reuters

Extra Sources: Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Antigua Observer, Diario Libre, Dominica News Online, Jamaica Gleaner, The St Kitts Nevis Observer, Times Caribbean Online

Argentina, Bolivia, Chile, Ecuador, Guyana, Paraguay, Peru, Suriname, Uruguay

Factiva sources: Agence France Presse, Associated Press Newswires, BBC Monitoring Americas, EFE, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Times Caribbean Online

Belize, Costa Rica, El Salvador, Honduras, Nicaragua, Panama

Factiva sources: Agence France Presse, Associated Press Newswires, BBC Monitoring Americas, EFE News Service, Reuters, El Herald (Spanish), La Estrella (Spanish), La Nación (Spanish), La Prensa (Honduras, Spanish), La Prensa (Nicaragua, Spanish), La Prensa (Panama, Spanish)

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Amandala, Diario Extra, El Faro, La Prensa Grafica

Brazil

Factiva sources: Agence France Presse, BBC Monitoring Americas, EFE News Service, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group, Jamestown, Latin News

Extra sources area or country specific: A 10 Mais, AC Jornal, Ac24 Horas, Agora RN, Bahia190, Blog Waldiney Passos, Carlos Britto, Contil Net Notícias, Correio24h, Correio do Povo, De Fato Notícias, Diário do Amapá, Diário do Nordeste, Diário Sul Maranhense, Ecos da Notícia, Folha BV, Folha do Acre, G1, Gazeta Online, GP1, Jornal A Tribuna, Jornal Imirante, Jornal Pequeno, Mossoró Hoje, Nossa Voz, Novo Notícias, O Alto Acre, O Câmera, O Imparcial, O Povo, PB Agora, Portal Correio, Portal Potiguar, Portal R10, Portal Tucumã, Por Dentro do RN, RN News, Roraima 1, Seles Nafes, Zero Hora

Canada

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Americas, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Colombia

Factiva sources: El Colombiano, El Espectador, El Tiempo, El Universal (Colombia), Infobae, Factiva Spanish search string (“muerto”, “asesin”, “enfrentamiento”, “tortura”, “masacre”, “combate”, “homicid”, “abatido”)

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Alerta Paisa, Alerta Santanderes, CERAC, Centro Nacional de Memoria Histórica (CNMH), CINEP / Programa por la Paz, Cronica del Quindio, Defensoria del Pueblo, Dejusticia, Diario del Huila, Diario del Sur, Ejercito de Colombia, Fundación Ideas para la Paz (FIP), Fundación Paz & Reconciliación (Pares), Hoy Diario del Magdalena, INDEPAZ, InSight Crime, JEP dataset, La Opinion, La Prensa Araucana, Observatorio de Derechos Humanos y DIH, Policía de Colombia, Razón Pública, Rutas del Conflicto

Guatemala

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Americas, EFE News Service, Reuters

Factiva in Spanish (asesinad# or pandill# or marero or maras or mara or muerto or muerta) All sources

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Haiti

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Americas, EFE, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: AyiboPOST, BINUH reports, Haiti Libre, ReliefWeb Situation Reports, RNDDH, Security Council reports, Vant Bèf Info

Mexico

Factiva sources: El Universal, Infobae, La Jornada, Milenio, Reforma, Factiva Spanish search (“ejecutar”, “ejecución”, “ejecutad”, “balazo”, “enfrentamiento”, “descuart”, “calcina”, “desmembra”, “decapita”)

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Borderland Beat, El Heraldo de Juarez, El Sol de Acapulco, El Sol de Hermosillo, El Sol de Irapuato, El Sol de Morelia, El Sol de Puebla, El Sol de Sinaloa, El Sol de Tijuana, El Sol de Zacatecas, InSight Crime, Lantia Intelligence, Proceso, Stratfor

United States

Factiva sources: Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Police Shooting Database from Washington Post

Venezuela

Factiva sources: Agence France Presse, Associated Press Newswires, BBC Monitoring Americas, EFE, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: ACAPS, CEPAZ, COFAVIC, El Nacional, El Pitazo, Foro Penal, Fundaredes, InSight Crime, Notas de Hoy, PROVEA, Runrun, Ultimas Noticias, Venezuela Organized Crime Observatory, VPI TV

Asia

Afghanistan

Factiva sources: Agence France Presse, Associated Press, Bakhtar News Agency, BBC Monitoring Central Asia, BBC Monitoring Former Soviet Union, BBC Monitoring South Asia, Pajhwok Afghan News, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area or country specific: Afghan Islamic Press, CTC Sentinel, Etilaat Roz, Hasht-e Subh, Kabul Now, Khaama Press, Terrorism Monitor, UNAMA

Social Media sources: X: @afghananalyst2, @afgfreedomaff, @AfghanistanInt6, @AfghanistanNMF, @alinazary, @aip_news, @ariananews_, @khorasandary, @natsecjeff, @NR_Front, @nrfafg, @sibghat_ah, @tolonews

Australia, Cook Island, Fiji, Kiribati, Micronesia, Nauru, New Zealand, Niue, Palau, Samoa, Seychelles, Solomon Islands, Tonga, Tuvalu, Vanuatu

Factiva sources: Agence France Presse, Bangkok Post, BBC Monitoring Asia Pacific,

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Bangladesh

Factiva sources: Agence France Presse, BBC Monitoring Central Asia, BBC Monitoring South Asia, Indo-Asian News Service, Press Trust of India, Reuters, The New Nation, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: CTC Sentinel, SATP Bangladesh Timeline, Terrorism Monitor

Bhutan, Maldives, Mongolia, Nepal, Singapore

Factiva sources: Agence France Presse, Bangkok Post, BBC Monitoring South Asia, Reuters, The Nation (Thailand)

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: SATP Timeline for Nepal & Maldives

Brunei, East Timor, Japan, North Korea, South Korea, Taiwan

Factiva sources: Agence France Presse, Associated Press, Bangkok Post, BBC Monitoring Asia Pacific, The Nation (Thailand), Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Cambodia, Laos, Paracel Islands, Vietnam

Factiva sources: Agence France Presse, Bangkok Post, BBC Monitoring South Asia, BBC Monitoring Former Soviet Union, BBC Monitoring Central Asia, BBC Monitoring Asia Pacific, The Nation (Thailand), Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

China

Factiva sources: Agence France Presse, BBC Monitoring Asia Pacific, BBC Monitoring Central Asia, BBC Monitoring Former Soviet Union, BBC Monitoring South Asia, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

India

Factiva sources: Press Trust of India, Telegraph (India), Deccan Chronicle, Imphal Free Press, The Pioneer, and the Assam Tribune (Search String: (ultra* or rebel* or insurgent* or militant* or jawan* or naxal* or maoist* or terrorist* or communal or clash*) and (kill* or die* or dead* or death* or "shot to" or gunned or hack* or succumb* or blast* or detonat* or lynch*))

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: SATP

Indonesia

Factiva sources: Agence France Presse, BBC Monitoring Asia Pacific, The Jakarta Post, Xinhua News Agency

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Antara News, CTC Sentinel, IPAC, Tempo Indonesia, Terrorism Monitor

Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan Uzbekistan

Factiva sources: Agence France Presse, BBC Monitoring Central Asia, BBC Monitoring Former Soviet Union, Interfax Central Asia General Newswire, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Malaysia

Factiva sources: Agence France Presse, Associated Press, Bangkok Post, BBC Monitoring Asia Pacific, New Straits Times, Reuters, The Nation

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Myanmar (Burma)

Factiva: Associated Press, BBC Monitoring Asia Pacific, Frontier Myanmar, Mizzima News, Reuters, Shan Herald Agency for News, The Irrawaddy, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Assistance Association for Political Prisoners (AAPP), Fortify Rights, Free Burma Rangers, Mizzima, Myanmar Now, Radio Free Asia, Terrorism Monitor, The Irrawaddy

Pakistan

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring South Asia, Press Trust of India, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: SATP, Terrorism Monitor

Papua New Guinea

Factiva sources: Agence France Presse, ForeignAffairs.co.nz, IHS Global Security Insight Daily Analysis, Radio New Zealand News

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area specific: Loop PNG, Post Courier, the Guardian

Philippines

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Asia Pacific, Inquirer.net, Philippines News Agency, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: CTC Sentinel, IPAC, Terrorism Monitor

Sri Lanka

Factiva sources: Agence France Presse, Asian News International, Associated Press, BBC Monitoring South Asia, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: SATP timeline for Sri Lanka

Thailand

Factiva: Bangkok Post, BBC Monitoring Asia Pacific, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Benar News, Thaiger, The Nation Thailand

Europe

Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Kosovo, Latvia, North Macedonia, Moldova, Montenegro, Romania, Serbia, Slovenia

Factiva sources: Agence France Presse, BBC Monitoring Europe, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Andorra, Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Italy, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, Norway, Poland, Portugal, San Marino, Slovakia, Spain, Sweden, Switzerland, Turkish Cyprus, Vatican

Factiva sources: Associated Press, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra sources area specific: Actu Marseille, SVT, Kvartal

Armenia, Azerbaijan, Georgia

Factiva sources: Agence France Presse, BBC Caucasus, BBC Central Asia, BBC Former Soviet Union, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Caucasus Watch, Live U Map, OSCE, Terrorism Monitor

Ireland, United Kingdom

Factiva sources: Agence France Presse, Associated Press, Reuters

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Russia, Ukraine

Factiva sources: ITAR-TASS World Service (Russia, Russian Language) Full article Russian language search: Keywords: погиб\ *or Потери or умер* or убит\ or ранен* or уничтож or ликвидированный or убили or гибели or устранять or ликвидировали or обезвредили

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Caucasian Knot, Mediazona, Poteru, UA Losses, Ukrinform, Warspotting

Middle East

Bahrain, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates

Factiva sources: Agence France Presse, Al Jazeera, BBC Monitoring Middle East

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Egypt

Factiva sources: Agence France Presse, Associated Press, BBC Monitoring Africa, BBC Monitoring Middle East, Reuters, Xinhua

Extra Sources: Africa Research Bulletin, Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Al-Araby, Mada Masr, Terrorism Monitor, CTC Sentinel

Iran

Factiva sources: Agence France Presse, Associated Press Newswires, BBC Monitoring Middle East, BBC Monitoring Newsfile, Al Jazeera, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Abdorrahman Boroumand Center, AVA Today, CTC Sentinel, Rudaw, Terrorism Monitor

Iraq

Factiva sources: Agence France Presse, Associated Press Newswires, BBC Monitoring Middle East, Al Jazeera English, Reuters, Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Airwars, Community Peacemaker Teams Iraqi Kurdistan, Hawar, Iraq Body Count, ISHM, PKK-HPG, Rudaw, UNAMI, Terrorism Monitor

Israel

Factiva sources: Agence France Presse, BBC Monitoring Middle East, Reuters, Haaretz, Jerusalem Post, Ma'an News Agency

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: B'Tselem, FDD, Janes Intelligence Review, Ministry of Foreign Affairs, OCHA, Terrorism Monitor, Wafa

Lebanon

Factiva sources: Agence France Presse, BBC Monitoring Middle East, Reuters, National News Agency Lebanon, The National (United Arab Emirates), Xinhua

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Al Manar, CTC Sentinel, El Nashra News, IDF sources, Terrorism Monitor

Social Media sources : Telegram @janoub_lebnan23, @dahieh4all

Syria

Factiva: No factiva sources

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Air Wars, SOHR

Turkey

Factiva: Agence France Presse, Reuters, BBC Monitoring Europe, BBC Monitoring Middle East, BBC Monitoring Central Asia, BBC Monitoring Newsfile, National Iraqi News Agency, FARS News Agency

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: CTC Sentinel, Jamestown Foundation, PKK-HPG, TAK, Turkish Ministry of Interior, Turkish Police (General Directorate of Security), Terör Şehitleri

Yemen:

Factiva: Agence France Presse, BBC - All sources, Reuters, Al Jazeera, Xinhua, Yemen News Agency (SABA), Al Arabiya, Al Araby Al Jadeed, Al-Arab, Anadolu

Extra Sources: Amnesty International, Crisis Watch, Human Rights Watch, International Crisis Group

Extra Sources area specific: Airwars, Barran Press, CIMP, Chatham House, Critical Threats, CTC Sentinel, ICRC, Mwatana, Sana'a Center, Sheba Intelligence, Stratfor, Terrorism Monitor

Appendix 6 Vague Number Translator

This appendix outlines UCDP's approach to interpreting vague or imprecise fatality counts in media reports and other sources. The vague number translator provides a framework that guides UCDP analysts in deriving fatality estimates from ambiguous language in source material.

UCDP applies moderate estimates, favoring conservative interpretations when ambiguity is present. The translator also underscores UCDP's commitment to methodological rigor and cautious estimation under uncertain conditions.

Each fatality figure derived from vague terms is based on the translator detailed below and the judgment of UCDP analysts. These assessments & decisions are documented internally but are not visible in the publicly available dataset. Users should be aware that vague terms have been interpreted using this consistent framework. However, estimates may still vary depending on contextual factors, the phrasing used, or the reliability of the original source.

Interpretations for Common Vague Terms

<i>Phrase</i>	<i>Estimated Fatalities</i>	<i>Notes</i>
<i>a couple</i>	2	Indicates two.
<i>multiple</i>	2	Indicates more than one.
<i>some</i>	2	Similar to "multiple."
<i>a few</i>	2	Suggests more than one.
<i>many</i>	3	Usually implies a larger quantity than "a few" or "some."
<i>numerous</i>	2	Indicates more than one.

<i>several</i>	Best: 3 High: 3 Low: 2	Indicates more than one & plurality with some uncertainty.
<i>group</i>	2 or 3	Context matters when choosing either 2 or 3
<i>casualties</i>	2	Only coded 2 when the term reflects fatalities in the source material.
<i>heavy casualties</i>	3	Only coded 3 when the term reflects fatalities in the source material.
<i>over a dozen</i>	13	Slightly more than twelve.
<i>dozens</i>	24	Indicates at least two dozen.
<i>a score</i>	20	A term for 20.
<i>scores</i>	40	Plural of "score"; implies at least two scores.
<i>over a hundred</i>	101	Minimal interpretation beyond 100.
<i>hundreds</i>	200	Indicates at least two hundred.
<i>over a thousand</i>	1001	Just above one thousand.
<i>thousands</i>	2000	Indicates at least two thousand.
<i>killed or injured/captured</i>	0	None may have been killed.
<i>killed and injured/captured</i>	1	At least one killed.
<i>neutralized</i>	0	Treated as 0 unless context confirms use as a euphemism for killing.

Disclaimer

Please note that this method of translating vague numbers has been implemented since 2010 to improve consistency in coding. However, UCDP GED data from earlier years may have been coded using slightly different interpretations. When conducting longitudinal analysis or comparing historical data, users should keep this potential variation in mind.