UCDP Geo-referenced Event Dataset (GED) Codebook

Version 1.1

In force as of 15 March 2012

Ralph Sundberg, Ph.D. Candidate and Project Manager

Mathilda Lindgren, Research Assistant

Ausra Padskocimaite, Research Assistant

Uppsala Conflict Data Program (UCDP)

www.ucdp.uu.se

When citing this dataset,please always cite:

The codebook
Sundberg, Ralph, Mathilda Lindgren and Ausra Padskocimaite, 2010, "UCDP GED Codebook version 1.0-2011", Department of Peace and Conflict Research, Uppsala University

The latest data presentation
Melander, Erik, and Ralph Sundberg, 2011, "Climate Change, Environmental Stress, and Violent Conflict – Test Introducing the UCDP Georeferenced Event Dataset", Paper presented at the International Studies Association, March 16-19, Montreal, Canada

This version of the codebook corresponds to the dataset named UCDP GED version 1.1-2011. Data extracted from UCDP systems on 15 March 2012

Table of Contents

1. INTRODUCTION 3

2. PURPOSE 3

3. DEFINITION OF THE EVENT 4

4. EVENT TYPES 6

4.1. SINGLE-DAY EVENTS (1) 7
4.2. SUMMARY EVENTS (2) 7
4.3. CONTINUOUS EVENTS (3) 8

5. EVENTS' TEMPORAL DIMENSION 8

6. SPLITTING OF THE EVENTS (DEATHSPLIT VARIABLE) 10

7. DATA INCLUSION 10

8. GEO-REFERENCING OF THE UCDP EVENTS 12

8.1. GEO-REFERENCING SOURCES 13
8.2. GEO-PRECISION VARIABLE AND ITS VALUES 13

9. GED STRUCTURE 14

10. RELIABILITY AND VALIDITY OF THE GED DATA 17

APPENDIX 1: EVENT TYPE CODING RULES 20

APPENDIX 2: TEMPORAL PRECISION CODING AND DATE ESTIMATION RULES 22

APPENDIX 3: GEO-PRECISION CODING RULES 27

REFERENCES 33

1. Introduction

This codebook describes the Geo-referenced Event Dataset (GED) (version 1.1) of the Uppsala Conflict Data Program's (UCDP) comprehensive data on organised violence. The document discusses and specifies the basic rules and methods underpinning the construction of this dataset, as well as provides an overview of its definitions and content.

Following the increase in demand from scholars of armed conflict to have access to reliable disaggregated data, the UCDP decided in February 2009 to convert all of its conflict- and dyad-level calendar-year data into a geo-referenced event data format. The project encompasses the conversion of annual data from the 1989-2010 time period in all three of the UCDP's categories of organised violence: state-based armed conflict (1), non-state conflict (2) and one-sided violence (3). By October 2011, the UCDP had compiled and coded information in event form for all three conflict types, covering the entire time period 1989-2010 for the African continent. These data are the basis of the UCDP GED version 1.1-2011.

The UCDP geo-referencing and event data project is grateful for valuable external input from Håvard Hegre and Halvard Buhaug at the Centre for the Study of Civil War, International Peace Research Institute, Oslo (PRIO), Nils B. Weidman at the Woodrow Wilson School of Public and International Affairs for beta testing, as well as Tomislav Dulic at the Hugo Valentin Centre, Uppsala University, for infrastructural support, ideas and comments.

2. Purpose

The purpose of this project is to provide the academic community with the most comprehensive event data available on organised violence in the post-1989 world, so as to answer the call for geographically and temporally disaggregated data. The first version of the geo-referenced event data will comprise actors and conflicts comparable with those found in the aggregated, annual UCDP data (see Data Inclusion). The UCDP aspires to expand future versions of the dataset to also include actors, incompatibilities and political violence that do not reach the 25-fatalities threshold of UCDP annual 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 wanting to illustrate conflict behaviour geographically, using geographic information systems software, to studying causal pathways by applying a variety of methods for statistical analysis.

Whilst retaining the ambition to provide a dataset open for 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 in effect primarily directed toward, and will most probably be useful to, researchers interested in the fatal outcomes of violent conflict behaviour at the level below the state.

The first version of the UCDP event data has several comparative advantages and strengths in the context of armed conflict event data. First, it offers systematic data on armed organised violence for the entirety of the African continent throughout 1989-2010, as opposed to other event datasets that focus on certain regions or countries and/or time periods. Second, it provides the users with reliable and detailed information on armed violence in three different types of conflict settings: state-based, non-state and one-sided violence. These categories are furthermore based on extensive research and clear definitions. Third, the new dataset contains extensive information on deaths resulting from armed violence, either in the form of single events or in the form of summary figures. All of this allows for high compatibility with the UCDP's country-year data as well as other global data. A more detailed comparison between the UCDP GED and other alternatives can be found in Appendix 4.

3. Definition of the Event

The basic unit of analysis is an "event". In general terms, this implies a phenomenon of lethal violence occurring at a given time and place.[1] More specifically, an event is defined as:

[1] As all events in the dataset are based on the UCDP definition of armed conflict, the actors involved in armed activities have at some point in the conflict stated a specified incompatibility for which they fight, or are involved in non-state conflict or one-sided violence. For a more detailed discussion of the UCDP definitions and delineations, please refer to the website, www.ucdp.uu.se (Definitions) or "UCDP/PRIO Armed Conflict Dataset Codebook Version 4-2011'".

"The incidence of the use of armed force by an organised actor against another organized actor, or against civilians, resulting in at least 1 direct death in either the best, low or high estimate categories at a specific location and for a specific temporal duration."

The separate elements of the definition are:

1. *Armed force*: use of arms in order to promote the parties' general position in the conflict, resulting in deaths.
    a. *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 meets the following requirements:
        i. there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other.
3. *1 direct death*: one death as
    a. directly related to combat between the warring parties (state-based and non-state conflict), *or*
    b. directly related to one-sided violence against civilians.
        i. *Civilians*: unarmed populations that are not part of the organized actors.

4. *Casualty estimate categories:* best, high and low casualty estimates defined by UCDP as
    a. *Best estimate:* the aggregated most reliable numbers for all incidents of organized violence during one event. If different reports provide different estimates, an examination is made as to what source is most reliable. If no such distinction can be made, UCDP as a rule includes the lower figure given.
    b. *Low estimate:* the aggregated low estimates for all incidents of organized violence during one event. If different reports provide different estimates and a higher estimate is considered more or equally reliable, the low estimate is also reported if deemed reasonable.
    c. *High estimate:* the aggregated high estimates for all incidents of organized violence during one event. If different reports provide different estimates and a lower estimate is considered more or equally reliable, the high estimate is also reported if deemed reasonable. If there are incidents when there is some uncertainty about which party has been involved, these may also be included in the high estimate.[2]
5. *Specific location*: a name and one pair of x and y coordinates that relate to the geographical information specified in the source material.
6. *Specific temporal duration*: a specified time period during which armed interactions cause at least 1 fatality. The smallest possible temporal unit to which an event can be related is a calendar day (24 hours) starting at midnight.

4. Event Types

The unit of analysis in the UCDP GED is an event. In the dataset each row constitutes an event of violence (see Definition of the Event). Each event is supplemented by additional information on the date, scale, perpetrator and other aspects of the event in question.

UCDP GED contains three different types of events based on the above definition: single-day events, summary events, and continuous events. Different types of events differ in these aspects: duration, temporal precision and continuity in armed violence.

[2] 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.

All types of events are restricted to only one specified spatial location, whilst the temporal aspects vary between different event types. Please refer to Appendix 1 for more details on how each event type is determined by the individual coders.

In the dataset the event type is indicated by the event type variable which can have three values:

- 1 - Observation is a *single-day* event;
- 2 - Observation is a *summary* event;
- 3 - Observation is a *continuous* event.

4.1. Single-day events (1)

Single-day events represent either a single incident of armed violence or separate incidents of fighting that result in at least one fatality within one calendar day (24 hours) at only one specified location.

Since the exact number of events is not of interest in this dataset (only the fatalities), multiple incidents of combat on the same day in the same specified location may, at times, be aggregated into a single event that covers one calendar day. This is done only if the fighting encompasses the same place, same date and the same actors.

4.2. Summary events (2)

Summary events, on the other hand, refer to deaths occurring from separate, multiple incidents of violence covering multiple days for which no exact disaggregated information is available. In other words, it is unclear how many battles took place during the time period specified in the source. These events are referred to as summary because the form of reporting does not allow knowing exactly when the casualties occurred, and how the battles were fought, and the event thus summarises a series of clashes into one event.

Commonly, summary events cover reporting that states that “in the past 2 months X people were killed”. Summary events can, however, be relevant for users

interested in understanding general casualty rates.[3] Since summary events often provide fatality estimates that cover long periods of time, several single-day and continuous events may occur during the same time period. Commonly, fatalities in such other events may be subtracted from a summary event (as we assume that the summary event and the other events to some extent cover the same fighting at times). This leads to summary events sometimes defying the parameters of the fatality estimates, as the 'high estimate' may at times be lower than the 'best' or 'low' estimate.

4.3. Continuous events (3)

Some of the armed activities occurring in one specified location and covering multiple days are interrelated and represent continuous fighting to a certain degree. These events are referred to as continuous events as the fighting is said to be continuous across a larger temporal unit than one day.

The burden of determining what qualifies as "continuous fighting" rests upon each coder and is always guided by moderation. This implies that key phrases in the sources used must clearly indicate continuity and interrelatedness in fighting in one location over multiple days in order for an incidence of violence to be coded as a continuous event.[4]

5. Events' Temporal Dimension

In addition to incidences of violence varying in their specific type, each event may also vary in its temporal dimension, or temporal precision. The temporal dimension of each event is thus made clear to the user through the application of a temporal precision variable which denotes with what accuracy a specific time period in which the event occurred is known.

[3] Note that fatalities caused by daily events during the same period and the same location as the summary event refers to, are always subtracted from the summary figures included in the data to avoid constructing a misleading overlap of casualty rates. More precisely, when a summary figure for a month refers to 40 dead and other sources report on 5 independent events in which a total of 36 fatalities were recorded, the summary figure entering the dataset will contain 40-36 fatalities, i.e. 4 fatalities. These coding procedures apply to all UCDP annual data as well as event data.

[4] The continuity of these battles is determined by each coder according to the provided information, the context of the event and other information on the conflict dynamics. This process is guided by a strong preference for moderation where summary event categorisation takes precedence in situations of unclear forms of longer fighting.

Each entry and all types of events receive a start and end date in the dataset that represent the specified time period within which the event took place. The start and end dates vary depending on how precisely the time period of the event is specified in the original data source.

Temporal precision variable can have six values:

- 0 - not applicable since the event is a summary event;
- 1 – the exact day of the event is know;
- 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.

The most precise are the *single-day events* in which information about the exact day of their occurrence is available. These events have the same start and end dates and receive a temporal precision value of 1. In those single-day events where the information about the exact day of their occurrence is not available, temporal precision values can vary from 2 to 5.

Summary events always encompass longer time periods than one day and have different start and end dates. Since summary events refer to estimated periods of time which contain a series of events that cannot be separated from each other, they always receive a temporal precision code of 0.

Continuous events can, in theory, have an infinite time span. In the data, however, the start and end dates given will show the start and end dates of the continuous fighting in question. Continuous events always have a longer time span than one day and always have precision codes of 2 (2-6 day continuous clash) to 5 (only the exact year can be specified).

Users interested in statistically analyzing the data using calendar date as a unit of analysis are encouraged to use the end date of each event as these represent dates after which the deaths coded in the event cannot have occurred.

Appendix 2 gives further instructions and examples on how to estimate dates and how to assign temporal precision values for different events.

6. Splitting of the Events (Deathsplit variable)

Many of the sources on which the dataset is built refer to a great number of geographical locations when only reporting a total number of deaths for those places. It is thus at times unclear how many deaths were incurred in each separate location.

In the UCDP GED this is dealt with through the Deathsplit variable. If an event is constructed through a Deathsplit this is denoted by the variable attaining the value of 1.

What this means is that a total number of deaths given for multiple geographical locations has been split between these multiple locations, as evenly as possible in the "total" category if the source does not imply anything else, to create separate events with fatalities estimates for each location.

The Deathsplit variable is useful for those users that wish to analyse only the most disaggregated data available since such splits increase geographical disaggregation but at the same time induces uncertainty into the data in terms of the location of fatalities.

At times the number of deaths will not be separable into even numbers. If this happens, the remaining death[s] is/are randomly placed in one/some of the events created.[5] The deaths of spatially split events therefore represent relatively crude estimates. They differ from the deaths counted in events that do not contain a Deathsplit in the uncertainty of the data.

When an event refers to multiple locations, yet only one death resulted from the fighting, the event will not be split and instead aggregated to a higher level of geographical location. For instance, if the source refers to an attack on two villages in Gulu district during the weekend as a result of which five people were abducted and one killed, Gulu district is used as the event location.

If reliable and significant information supports splitting deaths between multiple locations according to an informed, weighted decision instead of systematic procedure, this is indicated in the dataset by the geo-comment.

7. Data Inclusion

[5] The alternative of dividing casualty rates with the number of mentioned locations might risk contributing to total fatality rates being given in the form of fractions. GIS Software is better suited for working with whole numbers.

The event dataset has a dyad and actor focus, tracing the events of all dyads and actors that have crossed the 25 death threshold in any year of the UCDP annual data.[6] The dataset includes all three types of UCDP organised violence: state-based conflict, non-state conflict and one-sided violence. All three conflict categories of the UCDP annual data are mutually exclusive and coded events will therefore also be exclusive and non-overlapping. The data series start in 1989 and events before this calendar year are not included.

Inherent biases in the first version of the dataset emanate from an exclusive focus on conflict actors and their behaviour within the three categories of conflict behaviour, and not across (See Reliability and Validity of the GED Data), as well as including only years in which a dyad or actor cross the 25 fatalities threshold.[7] A first version of the data will thus have a clear relationship to the original UCDP datasets, whilst later versions may include new conflicts and conflict actors as well as cross-behavioural references for the conflict actors.

Table 1. UCDP data included in the GED

Conflict Type	Period	Actor Inclusion	Event Inclusion	Reference
State-Based	1989-2010	All dyad-years that cross the 25 death threshold and have a stated incompatibility.	All events leading to at least one death.	UCDP/PRIO Armed Conflict Dataset Codebook Version 4-2011
Non-State	1989-2010	All dyad-years that cross the 25 death threshold.	All events leading to at least one death.	UCDP Non-State Conflict Codebook Version 2.3-2010

[6] A dyad consists of two conflicting primary parties. In state-based armed conflicts, at least one of the primary parties must be the government of a state. In interstate conflicts, both primary parties are state governments. In intrastate and extra-systemic 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. 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. Finally, one-sided violence does not refer to dyads as there is only one organized actor directing its violence against unorganized civilians.

[7] For example, PREPAK in India becomes active in the state-based category in 2008, but also has one-sided activity below the threshold in years preceding 2008. The one-sided activity before 2008 can be included in future versions of the data given that the UCDP lists are updated with more comprehensive and structured information on incompatibilities for state-based conflicts and consistency in fighting in situations of non-state conflicts and one-sided violence.

One-Sided	1989-2010	All actor-years that cross the 25 death threshold.	All events leading to at least one death.	UCDP One-Sided Violence Codebook, Version 1.0-September 28, 2005

8. Geo-referencing of the UCDP Events

Data in the UCD GED is geo-referenced, meaning that each event is connected to a specific location which in turn is represented by a pair of x and y coordinates with the highest precision available.[8] The coordinates are fixed to the World Geodetic System of 1984 (WGS 84) and are specified in decimal degrees. Coordinates (latitude and longitude) used in the GED are based on the most precise location mentioned in the source and available in the gazetteers.

The lowest level of spatial disaggregation for an urban location is a suburb (in the NGA GNS gazetteer commonly referred to as “section of a populated place”) and the lowest level of smaller populated location is a village (in the NGA GNS – “locality”). Other features such as “mountains”, “parks” and “hills” are also used to specify geographical location. The highest level of spatial aggregation for location is a country (in the NGA GNS referred to as “independent political entity”). Between these two ends, sources may refer to larger areas including different levels of administrative divisions.

All text-based information on location is streamlined as much as possible in order to facilitate an overview of the spatial locations. For instance, all capitals are referred to as “cities” and most “populated places” and “localities” (in the NGA GNS) as towns and villages. Any unique information of added value to the preciseness of the location and its concomitant geographical coordinates is saved in the “Where” column in parenthesis (e.g. information on specific parts of a populated place, such as a city neighbourhood or district). Similarly, if the event occurs in a refugee camp close to a village and coordinates only exist for the village, then the location text will specify that the coordinates are near the village, leaving the name of the camp in parenthesis.

[8] The highest precision available in the NGA GEOnet Names Server is six decimals for each x and y coordinate.

8.1. Geo-referencing sources

The main source for the event location coordinates is the web-based gazetteer of the of the United States National Geospatial Intelligence Agency's GEOnet Names Server (abbreviated NGA GNS, available at http://earth-info.nga.mil/gns/html/index.html). When coding geographical locations the NGA GNS is always the first choice for the UCDP's coders. The NGA GNS gazetteer is probably the world's most comprehensive geographic library, covering the entire globe. Data is available either for downloading or from a web-based search engine. In addition to containing information on, for example, populated places, mountains and airports, the NGA GNS also contains information on administrative divisions. NGA GNS's coverage is, however, not perfect. This source is thus supplemented by a wide variety of sources, such as Maplandia.com, GeoHack and information provided from NGOs, governments etc. Throughout the UCDP GED the Geocomment variable contains information on what source has been used for the coordinates if the NGA GNS was not the source used.

Information used to determine administrative divisions (ADM1 and ADM2) stem from several different sources, commonly from a government's own website or reference literature that covers administrative divisions globally. What administrative divisions in a country correspond to the UCDP GED's ADM1 and ADM2 variables is stated in the Administrative Divisions list, which accompanies this codebook and the data files.

Correspondence regarding geographical coordinates, administrative divisions and any general questions or comments regarding the geographic aspects of the coding should be emailed to conflictdatabase@pcr.uu.se. Also, please report any potential errors in the dataset to the same e-mail address.

8.2. Geo-precision Variable and its Values

In order to determine the precision with which specific x and y coordinates are connected to an event location, the dataset uses a geo-precision variable. Its values also provide information on whether the geographical coordinates are estimated or not. Precise coding rules and examples of how the geo-precision values are assigned in the GED can be found in Appendix 3.

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 x and y 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 x and y coordinates;
- 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 related to a section of a country that is larger than the ADM1 (e.g. Northern Uganda), to an estimated pair of coordinates or to such locations as rivers, lakes, forests or parks which cover several administrative divisions or even countries (see Appendix 3 for precise rules);
- 6 - Event can only be related to the whole country;
- 7 - Event can only be related to an estimated pair of coordinates at sea/in water formations.

9. GED Structure

The UCDP GED has a structure that includes both classic UCDP variables (found in the UCDP/PRIO Armed Conflict Dataset) as well as new variables relevant for the geo-referenced event data. Some of the new variables have precision codes for time and space, event type codes, administrative divisions, as well as unique identifiers, actor and dyad IDs, type of violence etc. Below follows an overview over the relevant variables of the GED (from left to right in the Excel sheet).[9]

Variable Name	Content	Type
RelID	Relationship ID. This identifier key shows to what year, conflict and dyad/actor the event is related. The key is constructed using the abbreviation of the country name (for instance AFG for Afghanistan), the calendar year, the type of violence, the dyad or actor ID and a counter. This variable is also a unique identifier for each event in the	Text and number

[9] Some information is not published in the dataset. For instance, the extracts of text from the source article on which the coding is based is not included due to copyright issues.

	entire dataset.	
Year	Shows the calendar year of the event	Date
Active_year	Reflects if the event has been recorded in an active year for this dyad/conflict/actor. 1= yes, 0=no. (In this first version of the dataset only active years are included)	0-1
Code_status	Clear: event fulfils all UCDP criteria for inclusion. Aggregation (non-event): see the Data Discrepancies file	String
Type_of_violence	Reflects the type of conflict: 1 State-based 2 Non-state 3 One-sided	1 – 3
Source_article	The name and date of the source material from which information on the event is gathered	Formatted string (text and date)
Source_original	The type of person, organisation, or other unit from which the information in the source stems.	Text
Conflict_ID*	Conflict ID that corresponds to the conflict name (above).	Numeric
Conflict_name	Name of the conflict to which the event belongs. For non-state conflicts and one-sided violence this corresponds to the dyad name.	String
Dyad_ID*	UCDP dyad ID code for state based dyad. UCDP conflict ID code for non-state dyad. UCDP actor ID code for the one-sided violence actor.	Numeric
Dyad_name	Name of either the dyad (state-based and non-state conflicts) or the actor (one-sided violence).	String
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
Side_A_ID*	The unique ID of Side A. From the UCDP Actor Dataset.	Numeric
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
Side_B_ID*	The unique ID of Side B. From the UCDP Actor Dataset	Numeric
Where_location	The location as specified by the source material. Streamlined in spelling and structure for easy reference.	String
Where_prec	The geo-precision code for the location reflecting the preciseness of the coordinates and eventual estimation.	1-7
Country	Name of the country in which the event takes place. Note that this variable differs from the country variable in the annual UCDP data,	String

	which registers the country of the incompatibility/actor and not the country location of the specific events.	
ISOCC	The ISO 3166-1 country code for the country in which the event took place.	String
ISONumeric	The ISO 3166-1 numeric country code for the country in which the event took place.	Numeric
GWNO	The Gleditsch & Ward country code for the country in which the event took place.	Numeric
Region	The name of the region in which the country of the event is located	String
ADM1	The name o the first order administrative division (province etc) in which the event took place.	String
ADM2	The name o the second order administrative division (district etc) in which the event took place.	String
Lat	The latitude coordinates of the location.	Numeric
Lon	The longitude coordinates of the location.	Numeric
PRIOGrid_gid	The PRIO-GRID cell in which this specific event is located. Corresponds to PRIO-GRID version 1.0	Numeric
Geocomment	A cell used to comment on any special decisions that have been made in terms of the geo-referencing.	String
Date_start	Start date of the event in the form YYYY-MM-DD.	Date
Date_end	End date of the event in the form YYYY-MM-DD.	Date
Temp_prec	How precise the information about the exact time (day) of the occurrence of the event is.	0 - 5
Event_type	Denotes event type. Differentiates between single-day events (1), summary events (2), and continuous events (3)	1 - 3
Side_A_Deaths	The estimated number of deaths for Side A	Numeric
Side_B_Deaths	The estimated number of deaths for Side B	Numeric
Civilian_Deaths	The estimated number of deaths of civilians	Numeric
Unknown_deaths	The estimated number of deaths of unknown persons	Numeric
Best_est	The best estimate of fatalities resulting from the event.	Numeric
High_est	The high estimate of fatalities resulting from the event. When there is no high estimate for an event, the high estimate cell registers the best estimate fatalities.	Numeric
Low_est	The low estimate of fatalities resulting from the event. When there is no low estimate for an event, the low estimate cell registers the best estimate fatalities.	Numeric
Deathsplit	Variable denoting whether or not the fatality estimates given have	0-1

	been artificially split between events. 1=yes, 0=no.	

* Conflict, dyad and actor IDs all correspond to the IDs used by the UCDP in the datasets that deal with the respective type of organized violence. To interpret these codes one should refer to the datasets on state-based armed conflict, non-state conflict, one-sided violence and the UCDP Actor dataset. The release of the first version of UCDP GED (December 2011) does not coincide with the release of updated versions of UCDP datasets on organized violence, and some inconsistencies between the datasets will occur. This is especially so in terms of the actor codes for actors that partake in an alliance when fighting one or more parties. Such alliances are given unique numeric IDs. These IDs and what alliance they correspond to are given in the Alliances document that accompanies the UCDP GED and this codebook.

10. Reliability and Validity of the GED data

The first version of UCDP GED contains some restrictions and biases in the data that UCDP aspires to remedy in future, updated versions of the dataset. Remaining aware of these challenges concerning data reliability and validity is important both for current users and future UCDP updates of the data.

First, the UCDP GED is biased towards escalation because it only includes dyads, conflicts and actors that at some point reached the 25 death threshold in one of the three categories of organised violence. The UCDP has chosen to apply this restriction for the first version of the dataset in order to secure high quality and comparability of data. Second, the event dataset is biased towards uniform behaviour of conflict actors, as the first version does not allow for analysis of behaviour of actors across the three conflict types. Third, the first version of the dataset only starts recording events from 1989. This time period will likely not be expanded backwards in time, due to the lack of quality source information before 1989.

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 types, 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 event type 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.

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 and daily reports or more summary-like reports covering larger areas. This may lead to variations across the data. 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, which decreases the geographical precision in large.

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. Each interpretation and specific rule for an unclear location phrase in a certain country is also added to the “UCDP Administrative Divisions” file, available in the UCDP website.

11. Procedures to assure high quality data

The data that make up the UCDP GED are, before being published, checked and double-checked in several stages to ensure that the dataset maintains a high quality. In the first stage each coder is asked to check for consistency regarding, to mention but a few aspects, conflict and dyad names and ID:s, dates and temporal precision, streamlining of names, and the integrity of the fatality estimates. In a second stage a project manager runs a series of additional checks, as well as those already carried out by the coder his/herself. Thirdly, automated scripts are run to check for, for instance, the uniqueness of identifying codes, that actor and dyad ID:s are correct, that fatality estimates correspond with the aggregated datasets, and that dates are coded consistently. In a final stage the coder and the project manager jointly visualize the data in Spatialkey and check the consistency and accuracy of ADM1:s and

ADM2:s, as well as conducting random controls of specific locations. Errors are corrected in dialogue with the project manager to ensure consistency.

APPENDIX 1: Event Type Coding Rules

This document presents both general and specific rules that guide the coding of event type and presents some concrete examples for each of the three categories of events.

Event Type 1 – Single-day events

1. All events occurring in one calendar day and in one specified location:

    a. “Shelling on Tuesday on Sarajevo city”;
    b. “5 rebels died yesterday in Gulu district”.

2. All events that appear to be confined to one calendar day, representing one incident of armed violence, though the exact calendar day when the event took place is unclear (only a time period):

    a. “During a battle/clash/ambush last month”;
    b. “During an offensive last week” - offensive could be interpreted as a single-day event if it refers to one attack and as a summary or continuous event if it refers to a military operation which lasts longer than one day. In order to determine the event type look at the source. If the original source does not point to any direction an offensive should be interpreted as a single-day event;
    c. “Mass grave” - if nothing else is specified in the source, mass graves are understood as resulting from one incident, often confined to one calendar day, and is thus coded as a single-day events.

Event Type 2 – Summary Events

1. All events referring to deaths resulting from separate activities of fighting which span longer than one day in one specified location:

    a. “13 killed in battles last week in Monrovia town”;

2. All events which span longer than one day and have unclear information about the continuity and/or temporal duration of fighting:

    a. “6 rebels died during fighting last weekend” – if the continuity of the event is not certain, the event is interpreted as summary event;
    b. “6 rebels died during clashes in Kitgum town” - if time period of the event is not specified, dates are estimated for 1 week and the event is coded as a summary event.

Event Type 3 – Continuous events

1. Continuous events are those events which consist of more than one calendar day of continuous and interrelated fighting. These are coded with moderation and require solid evidence and support for determining the continuity of fighting.

    a. “A weekend offensive”;
    b. “During 5 day offensive”;
    c. “Fighting throughout the last 2 days in X village”

APPENDIX 2: Temporal Precision Coding 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 one week, counting backwards from the day of reporting or specified event end date:

    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 0 – Summary Events

1. Since summary events consist of several events which are not separable from each other and do not have clear temporal duration, it is impossible to determine their temporal precision. Therefore, such events always receive a temporal precision code of 0. Summary events always span for more than one day and have different start and end dates. The dates, however, are estimated according to the rules set below.

Temporal Precision 1 – Daily Precision of Time

1. If the exact date of a single-day 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 single-day or continuous event types 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 single-day or continuous event types 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 single-day or continuous event types 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 single-day or continuous event types 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. Coordinates for those locations which represent a higher level of aggregation than suburb, village or town, but lower than ADM2 (such as sub-counties, parishes etc.) should be used only if they are available in the gazetteers. If coordinates for such locations are not available, they should be aggregated to the lowest available administrative division. For instance, if coordinates for Pader parish are not available, but it is known that the parish is in Kilak county (ADM2), then coordinates for Kilak county should be used. Some exceptions from this rule exist, and are specified in the UCDP Administrative Divisions file.

4. 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 suburb, town or village should be used only if it is known that the event location is within or close to (within 25 km) that suburb, town or village. If information about the locations' proximity to that suburb, 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.

5. If the source refers to a certain location (e.g. river, forest, lake, park, mountains etc.) which lies in several ADM2s, ADM1s or even countries and it is known in which ADM2, ADM1 or country the event took place, the centroid point coordinates of that location should be used only if they are within that ADM2, ADM1 or country in which the event took place. If this is not the case, the location should be aggregated to the lowest available administrative division (in which the event took place) while keeping information about the location of the event in "Where" and adding a geocomment. For example, if the event took place in the Falls National Park in Gulu district, but the centroid point coordinates of the Falls National Park are in the neighbouring Pader district, then Gulu district coordinates should be used and "Falls National Park" should be kept in the parentheses in "Where".

6. If the coordinates for certain event locations are not available in the gazetteers, these locations can be represented by a lower level geographical locations (see specific geo-precision rules below). For example, Northern Uganda could be represented by the Gulu district (ADM1). Information about representation locations and their coordinates is saved and can be found in the "UCDP Administrative Divisions" file.

7. 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. Information on these types of estimates as well as changes of structure in ADM1 over time is provided in the "UCDP Administrative Divisions" file.

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. “Suburb”, “outskirt” – outskirts and suburbs are treated as independent geographical entities which means that if coordinates for certain suburbs and outskirts are available they should be used instead of town coordinates;
4. “Bujumbura” – if the location name only refers to Bujumbura, which can represent a town, ADM2 or ADM1, and there are other entries for the specific conflict or actor that instead specify “Bujumbura province”, then the location is coded as “Bujumbura town”.

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. Information about the proximity to that location is added to the parenthesis in “Where”.

1. “Near/in the vicinity of/adjacent to/just outside/around Kitgum town” – coordinates for Kitgum town and “Kitgum town (near/in the vicinity of/adjacent to/just outside/around)” in “Where”;
2. “Pietermaritzburg area” – coordinates for Pietermaritzburg town and “Pietermaritzburg town area” in “Where”;
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 together with geo-precision of 2;
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 while keeping information about the distance in the parentheses in “Where” (“Uppsala town (17 km from)”);
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 in the gazetteer, then coordinates for Bujumbura city will be used and the name of Gishingano village saved in parentheses;
7. “Niuland village near Dimapur town” - if coordinates for Niuland village are not available, but coordinates for Dimapur town exist, the latter are used while adding “near Dimapur town” in the parentheses and a geo-comment;
8. “Dungu territory in DRC” – third level administrative divisions (ADM3) receive 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 in the gazetteers, 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 the battle takes place “over” a certain ADM2, coordinates for that ADM2 will be used;
4. If the event takes place in a certain location, but the coordinates for that location in the gazetteers are not available or the location is not specific and needs an estimate (for example, “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), but the ADM2 of that location is known, then coordinates for ADM2 should be used together with geo-precision value of 3 and a geo-comment. For example, in case of “Pader village in Gaia county (ADM2) Gulu district (ADM1)”, where gazetteers do not have coordinates for Pader village, Pader village will be kept in “Where”, but Gaia county coordinates should 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 the battle takes place “over” a certain ADM1, coordinates for that ADM1 are used;
3. “Bujumbura” – if the location name does not inform on the level of aggregation for a place that can be a town, ADM2 or ADM1, and other entries for the same conflict does not either do this (see Geo-precision 1 Rule 4), then these places are understood as the highest possible level of aggregation. Thus, Bujumbura would be coded as Bujumbura province (ADM1).
4. If the ADM2 in which the event took place in unclear (e.g. different sources refer to different ADM2s in which the same event took place), the location is aggregated to the ADM1 level;
5. If the event takes place in certain location, but the coordinates for that location in the gazetteers are not available or the location is not specific and needs an estimate (for example, “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) and ADM2 of that location is not known, then coordinates for ADM1 is should be used together with geo-precision value of 4 and a geo-comment. For example, in case “Between Pader and Kitgum in Gulu district (ADM1)” information about the location will be kept in “Where”, but Gulu district coordinates and geo-precision value of 4 will be used.

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, coordinates of an ADM1 within that part of the country are used as a representation of that area together with geo-precision value of 5 Information about which ADM1 is used as a representation as well as its coordinates is saved in the “UCDP Administrative Divisions” file.

2. If a pair of coordinates is estimated using Google Earth. For example, if the location is on the border between two countries and the location of such point is more or less 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” in “Where” with the coordinates for a selected point on the border between Uganda and Sudan estimated in Google Earth;

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 event takes place in certain location, but the coordinates for that location in the gazetteers are not available or the location is not specific and needs an estimate (for example, “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) and both ADM2 and ADM1 of that location are not known, but it is known in which part of the country (larger than ADM1) that location is, then coordinates for that part should be used together with geo-precision of 5 and a geo-comment. For example, “along Aswa river in Northern Uganda” will be coded as “Aswa river” in “Where”, coordinates for Northern Uganda and geo-precision of 5;

5. If the location information refers to lakes, rivers, mountains, forests, parks and other places which lie in several administrative divisions/countries and does not specify administrative division/country, then centroid point coordinates for that location are used together with the country of those coordinates.

Geo-precision 6

If the location information refers to an entire country, centroid point coordinates of that country (GNS NGA coordinates for “Independent Political Entity”) are used. Also, if the location is not provided/is unclear/refers to several locations which can not 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.

Geo-precision 7

If the event takes place over water, 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.

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 using Google Earth.

References

ACLED ”Armed Conflict Location and Event Dataset (ACLED) Codebook”, Version 1-2006, Clionadh Raleigh, Håvard Hegre, Joachim Carlsen and Christian Ormhaug, Centre for the Study of Civil War, International Peace Research Institute, Oslo (PRIO)

Chojnacki, Sven (2006) “Anything New or More of the Same? Wars and Military Interventions in the International Systemm 1946-2003”, *Global Society* 20(1): 25-46.

EDACS “Event Data Project on Conflict and Security (EDACS) in Areas of Limited Statehood” September 12, 2007, Sven Chojnacki and Nils Metternich, Free University of Berlin

UCDP/PRIO Armed Conflict Dataset Codebook, Version 4-2011, Themnér, Lotta, and Peter Wallensteen in collaboration with Halvard Buhaug, Håvard Strand, Joachim Carlsen, , Department of Peace and Conflict Research, Uppsala University

UCDP One-Sided Violence Codebook, Version 1.0 – September 28, 2005, Joakim Kreutz and Kristine Eck in collaboration with Peter Wallensteen, Lotta Harbom, Stina Högbladh and Margareta Sollenberg, Department of Peace and Conflict Research, Uppsala University

UCDP Non-State Conflict Codebook, Version 2.2 – 2010, Ralph Sundberg, Department of Peace and Conflict Research, Uppsala University

Uppsala Conflict Data Program (2007) “Appendix 2b. Definitions, Sources and Methods for the Conflict Data” in *SIPRI Yearbook 2007, Armaments, Disarmament and International Security* pp. 91-93