

UCDP/PRIO Armed Conflict Dataset Codebook¹

Version 17.2²

Uppsala Conflict Data Program (UCDP)

www.ucdp.uu.se

International Peace Research Institute, Oslo (PRIO)

www.prio.no

When using the data, please cite Gleditsch et al. (2002), the most recent presentation of the data (Allansson, Melander and Themnér), and (when appropriate) this codebook. Please always include the version number in analyses using the dataset. When referring to the dataset, make sure to use the correct name: the UCDP/PRIO Armed Conflict Dataset.

¹ The first release of the Armed Conflict Dataset was prepared at PRIO in 2002 in close collaboration with researchers at the Department of Peace and Conflict Research at Uppsala University and the Departments of Sociology and Political Science and Geomatics at the Norwegian University of Science and Technology (NTNU). For a description of the division of labor in creating the database, see the first footnote in Gleditsch et al. (2002: 615). This footnote also lists the financial sources of support for the entire project and credits for comments and advice received along the way.

This is version 17.2 of the codebook and associated documents. We are grateful to several colleagues and external users for constructive comments. For our policy on version labeling, see Section 2.3 below, and for a complete history of earlier versions see the document called Version History. For further comments and suggestions on the data and the codebook, please communicate both to project leader in Uppsala Lotta Themnér (lotta.themner@pcr.uu.se) and to jpr@prio.no.

² In 2017, the versioning was changed. For more information on this, see section 2.3.

This version compiled and updated by:

Lotta Themnér

Uppsala Conflict Data Program (UCDP)

Earlier versions and definitions:

Peter Wallensteen

Birger Heldt

Margareta Sollenberg

Mikael Eriksson &

Stina Högladh

Uppsala Conflict Data Program (UCDP)

Håvard Strand

Håvard M. Nygård

Halvard Buhaug

Joachim Carlsen

Nils Petter Gleditsch

Håvard Hegre

Christin M. Ormhaug &

Lars Wilhelmsen

International Peace Research Institute, Oslo (PRIO)

1 Introduction

This document describes the UCDP/PRIO Armed Conflict Dataset, a joint project between the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University and the Centre for the Study of Civil War at the International Peace Research Institute in Oslo (PRIO). The dataset was first presented in Gleditsch, Wallensteen, Eriksson, Sollenberg & Strand (2002), and is available for download from <http://www.pcr.uu.se/research/ucdp/datasets/>. The project is part of the larger Uppsala Conflict Data Program, which is thoroughly described at www.ucdp.uu.se.

Both UCDP and PRIO offer a range of other datasets, compatible with the UCDP/PRIO dataset. Of special importance is the UCDP Dyadic dataset, which is based on the UCDP/PRIO Armed Conflict Dataset, but goes beneath the conflict level and focuses on different dyads within each conflict. For more information on the UCDP Dyadic dataset, and for free download visit UCDP's web page. Further compatible datasets can be found on both PRIO's and UCDP's web pages.

Version 17.2 is updated in accordance to the changes listed in Allansson, Melander and Themnér (2017), and the 2016 conflicts have been added to the dataset. The dataset will continue to be updated with new data annually.

2 Definition of Conflict

The main unit in this dataset is an "Armed Conflict" as defined by UCDP.³ This definition is presented unabridged in Section 2.1. Each conflict is listed in the database and given a unique ID code. The temporal aspect of a conflict is not addressed by this definition; hence, two conflict episodes over the same incompatibility will be assigned the same ID regardless of the time separating them. See below for further clarifications.

2.1 *Armed Conflict*

UCDP defines 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.

³For a more in-depth discussion on definitions, see <http://www.pcr.uu.se/research/ucdp/definitions/>.

- (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.⁴
- (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.⁵ 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. See Section 4 for details on the sample of countries covered by this definition.
- (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.

⁴Note 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.

⁵With 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 3.2) below for further information on these cases.

(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 (internal conflict).

For information on how these elements and definitions are connected to the different fields in the dataset more precisely, see Section 3.

2.2 *Missing Data Problems*

The missing data code is -99. However, the dataset does not include unclear conflicts where information on key variables to the definition of conflict is uncertain or missing. Key variables are those related to the incompatibility, actors and intensity. In addition, a number of events have been identified as potentially in accordance with the criteria for inclusion. These events include possible new dyads and additional years for active conflicts. Consult the list of unclear cases for further information.

The information also varies with regard to the level of precision. For the start date variables, the precision level is indicated in a separate variable, see Sections 3.16 and 3.18. Apart from that, the dataset only includes information when we are quite confident that it is correct. The bias produced by this approach is against the inclusion of conflicts in the earlier decades and in the less-developed world. An armed conflict in a developed country in the 1990s is more likely to be recorded than a conflict in a less developed country in the 1950s.

2.3 *Version Name Convention*

In 2017, the version name convention was changed, giving all UCDP datasets the same version number across the board. This was done so that users more easily can see which UCDP dataset corresponds with which.

This codebook corresponds to Version 17.2 of the UCDP/PRIO dataset. For every new release, substantial changes will be documented in a separate document.⁶ This should be helpful to researchers trying to replicate a particular study. We recommend that whenever this dataset is used, the version number should be cited.

The version number is a combination of a year and a number. The year refers to when the dataset is updated with new observations. If there are changes in the data between yearly updates, or if there are substantial changes in the structure of the dataset, the number behind the year is incremented.

⁶ <http://ucdp.uu.se/downloads/>

3 The Main Conflict Table

The observation (or unit) in the Main Conflict table is the conflict-year. Each conflict is listed in all years where fighting in one or more dyad(s) caused at least 25 battle-related deaths. This is in contrast to Version 2.1 (and earlier versions) of the dataset.⁷

The calendar year is the basic unit of every observation. Thus, if a conflict during the period June–September results in 30 casualties, that year will be recorded as a year of conflict. However, if the same number of casualties occurred in the period November–February and the conflict failed to reach the threshold of 25 battle-related deaths in either calendar year, neither year will be coded as in conflict. This has a number of consequences that will be discussed below. Start dates frequently refer to years prior to the first calendar year of a conflict, as the start of a conflict might be in a year with less than 25 fatalities. Small conflicts might not be included. Certain observations might be based on a single event, such as the Omagh bombing in Northern Ireland in 1998, which exceeded the minimum threshold for armed conflict.

Table 1: Definition of Variables In The Main Conflict Table

Variable	Label	Description
conflict_id	Conflict identifier	The unique identifier of all conflicts
location	Country name(s)	The name(s) of the country/countries whose government(s) have a primary claim to the issue in dispute.
side_a	Country name(s)	Identifying the country/countries of side A in a conflict. Always the government side in internal conflicts. Note that this is a primary party to the conflict.
side_a_id	Identifier of actor on side A	The unique identifier of the actor on side A. Note that the ID system employed has changed in version 17.1. In prior versions, this ID was the Gleditsch and Ward identifier for the state; this is NO LONGER THE CASE. For this, use the GWNoA field.
side_b_2nd	Name of state(s) supporting side A with troops.	Identifying the country/countries supporting side A in the conflict.
side_b	Country name(s) or Opposition actor(s)	Identifying the opposition actor or country/ countries of side B in the conflict. In an internal conflict, this includes a military opposition organization. Note that this is a primary party to

⁷ The previous formulation read ‘The observation (or unit) in the database is a conflict-year, a subconflict, or a subset of either over a period of time where no element in the definition described in Section 2 is changed. Each conflict is likely to include several observations’. This definition of the primary unit made sense in the text lists that preceded this dataset, where space was an important issue. The data structure was kept in order to stay compatible with previous versions. But while this might be advantageous to old users, it has confused a number of new users. The new annual data structure therefore replaced the old structure in Version 3.0.

		the conflict.
side_b_id	Identifier of actor on side B	The unique identifier of the actors on side B, as a comma separated list. Note that the ID system employed has changed in version 17.1. In prior versions, this ID was the Gleditsch and Ward identifier for the state; this is NO LONGER THE CASE. For this, use the GWNoB field.
side_b_2nd	Name of state(s) supporting side B with troops	Identifying the country/countries supporting side B in the conflict.
incompatibility	Incompatibility	A general coding of the conflict issue
territory_name	Name of territory	The name of the territory over which the conflict is fought, provided that the incompatibility is territory.
year	Year of observation	
intensity_level	Intensity level	The intensity level in the dyad per calendar year. Two different intensity levels are coded: minor armed conflicts and wars. See Section 3.12 for definitions of the two categories.
cumulative_intensity	Cumulative Intensity	The intensity of the conflict, taking into consideration the conflict history.
type_of_conflict	Conflict type	Four different types of conflict: extrasystemic, interstate, internal and internationalized internal. See Section 3.14 for definitions of the four types.
start_date	Date of conflict initiation	The date, as precise as possible, of the first battle-related death in the conflict.
start_prec	Precision of startdate	The level of precision for the initial start date.
start_date2	Date of fatality threshold in current episode	The date, as precise as possible, when a given episode of conflict activity reached 25 battle-related deaths in a year.
start_prec2	Precision of startdate2	The level of precision for startdate2
ep_end	Episode Ended	Codes whether this is the last year of conflict activity in this episode
ep_end_date	Date when conflict activity ended	The date, as precise as possible, when conflict activity ended.
ep_end_prec	Episode end precision	The level of precision for episode end
gwno_a	GW number of side A	GW numbers of the state on side A.
gwno_a_2nd	GW number(s) of states supporting side A	GW numbers of all countries supporting side A with troops.
gwno_b	GW number(s) of side B	GW numbers of all countries on side B, separated by semicolons.
gwno_b_2nd	GW number(s) of states supporting side B	GW numbers of all countries supporting side B with troops.
gwno_loc	GW number(s) of locations	GW numbers of all location countries, separated by semicolons.

region	Region of location	Identifying the region of the location. See Section 3.27 for the definition of the regions.
version	Version number	The current version of the dataset, i.e. 17.1

3.1 *conflict_id*

Conflict identifier.

In version 17.1 of the dataset, the ID system has been changed. To download a conversion table containing new and old IDs, visit <http://ucdp.uu.se/downloads/>.

3.2 *location*

Location is defined as the government side of a conflict, and should not be interpreted as the geographical location of the conflict.

- For internal and internationalized internal conflicts (see Section 3.14 for definition), only one country name is listed. This is the country whose government or territory is disputed. For certain conflicts, such as Kurdistan, the disputed territory is divided between different countries. Following our definition, we have coded separate conflicts for each country.
- For interstate conflict, all primary parties are listed in the Location field. Even if several governments are involved in the conflict, only countries that fulfil the inclusion criteria for primary actors are listed here. This normally means that two countries are listed, but there are four notable exceptions: In the Arab-Israeli war of 1948–49 as well as the Suez war of 1956, the war in Afghanistan in 2001 and the war in Iraq in 2003, there are more than two primary parties to the conflict.
- For extrasystemic conflicts, Location is set to be the disputed area, not the government of the colonial power. Since the Location field in these conflicts by default does not indicate members of the international system, it constitutes an exception from the definition presented in Section 2. Location is a string variable, listing the names of the countries involved. These might be fighting together or against each other. The string is split in two ways, hyphen ('-') splits the different sides in an interstate war, and comma (',') splits different countries fighting together on the same side.

3.3 *Side_a*

Side A is by definition always a primary party to the conflict. In internal conflicts, side A is always the government side, it is one of the sides in interstate conflicts and the colonial state in extrasystemic conflicts.

Side A is a string variable, and in the few cases where there are more than one primary party on side A, these are separated by a comma (',').

3.4 *Side_a_id*

Side A ID is the unique identifier of the actor on side A in the conflict.⁸

3.5 *Side_a_2nd*

Side A Secondary lists all states that enter a conflict with troops to actively support side A in the conflict. By definition, only independent states can be a secondary party in conflict. A secondary party on side A shares the position in the incompatibility with Side A in the conflict. Side A Secondary does not need to meet the 25 battle-related deaths criterion to be included in the dataset; an active troop participation is enough.

Side A Secondary is a string variable, where the different names are separated by a comma (',').

3.6 *Side_b*

Like Side A, Side B is by definition a primary party to the conflict. Side B is the opposition side of all internal and extrasystemic conflicts and the second side in an interstate conflict. Thus, side B can include both states and non-governmental opposition groups, depending on the type of conflict. When the primary party listed on Side B is an opposition group, the column lists the group name in abbreviated form. Even if the group changes its name during the course of the conflict, we record them under the same name for all years. Thus, instead of recording PRA (Popular Resistance Army) for Uganda 1982, we have recorded NRA (National Resistance Army) for all years (1982-86), even though the group only changed its name from PRA to NRA in 1983. We refer to the UCDP Actor Dataset for the full name and name history of opposition groups.⁹

Side B is a string variable, where the different names are separated by a comma (',').

3.7 *Side_b_id*

Side B ID is the unique identifier of the actor on side B in the conflict.¹⁰ The ID is taken from the UCDP Actor Dataset (<http://ucdp.uu.se/downloads/>).

⁸ In all versions up until v-4 2016, the Gleditch& Ward (2007) country codes were used as IDs. From version 17.1, all actors, both governments and non-state groups, have new IDs.

⁹ <http://ucdp.uu.se/downloads/>

¹⁰ In all versions up until v-4 2016, the Gleditch& Ward (2007) country codes were used as IDs, if the actor was a government. From version 17.1, all actors, both governments and non-state groups, have new IDs.

3.8 *Side_b_2nd*

Side B Secondary lists all states that enter a conflict with troops to actively support side B in the conflict. Only states are included as Side B Secondary. Furthermore, the states listed share the position in the incompatibility with Side B in the conflict. Side B Secondary does not need to meet the 25 battle-related deaths criterion to be included in the dataset; active troop participation is enough. Note that when there is more than one opposition organization listed in an internal conflict, the dataset does not provide information on which of these groups the state coded as Side B Secondary is supporting.

Side B Secondary is a string variable, where the different names are separated by a comma (',').

3.9 *Incompatibility*

As a country can experience several simultaneous conflicts, it is essential to differentiate between them. As described in Section 2.1, UCDP collects information on conflicts where the incompatibility, i.e. the general incompatible positions, concerns either government, territory or both. Earlier versions of the UCDP/PRIO dataset only contained two incompatibility categories, territory and government. Conflicts that concerned both territory and government were assigned to their primary incompatibility. From Version 4-2007 the incompatibility is coded in three categories:

1. Territory
2. Government
3. Government and Territory

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.

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.

3.10 *Territory*

If the incompatibility is territory, the disputed territory will be listed here. In case the two sides use different names for the disputed territory, the name listed is the one used by the opposition organisation. One reason for this is that this is most often the name that the general public recognises. Another reason is that there are cases where the disputed territories do not have an official name. This is the case in e.g. conflict ID 421, a territorial conflict in north-eastern India. The rebel organisation NDFB is fighting for independence for

a territory that is a part of the Assam region, and does not have an official, separate name. Thus, we use the rebel's name for the territory: Bodoland.

3.11 Year

The year of observation.

3.12 Intensity_level

The intensity variable is coded in two categories:¹¹

1. Minor: between 25 and 999 battle-related deaths in a given year.
2. War: at least 1,000 battle-related deaths in a given year.

3.13 Cumulative_intensity

This variable takes into account the temporal dimension of the conflict. It is a dummy variable that codes whether the conflict since the onset has exceeded 1,000 battle-related deaths. For conflicts with a history prior to 1946, it does not take into account the fatalities incurred in preceding years. A conflict is coded as 0 as long as it has not over time resulted in more than 1,000 battle-related deaths. Once a conflict reaches this threshold, it is coded as 1.

3.14 Type_of_conflict

We define four types of conflict:

1. Extrasystemic armed conflict occurs between a state and a non-state group outside its own territory. (In the COW project, extrasystemic war is subdivided into colonial war and imperial war, but this distinction is not used here.) These conflicts are by definition territorial, since the government side is fighting to retain control of a territory outside the state system.
2. Interstate armed conflict occurs between two or more states.
3. Internal armed conflict occurs between the government of a state and one or more internal opposition group(s) without intervention from other states.
4. Internationalized internal armed conflict occurs between the government of a state and one or more internal opposition group(s) with intervention from other states (secondary parties) on one or both sides.

¹¹ In earlier versions of the UCDP/PRIO dataset, the intensity variable contained three categories: minor, intermediate and war. The intermediate category was defined as “more than 25 battle-related deaths per year and a total conflict history of more than 1000 battle-related deaths, but fewer than 1000 per year.” Thus, the variable included a temporal dimension into the intensity coding. However, as many users incorrectly interpreted the variable as ordinal, it was decided that the intermediate category should be represented by a separate dummy variable denoting cumulative intensity.

3.15 *Start_date*

The date of the first battle-related death recorded in the conflict is coded as the Startdate in the dataset. The date is set after the conflict fulfils all criteria required in the definition of an armed conflict, except for the number of deaths. In some cases, the initial fatality occurs in a year prior to the first year of activity. For instance, in the conflict in Ethiopia over the territory Eritrea (conflict ID 78), the first battle-related deaths occurred in September 1961. During the remaining months of 1961, the conflict did not reach the required 25 battle-related deaths threshold and the conflict is thus coded as inactive in 1961. 25 battle-related deaths were not recorded until three years later.¹²

3.16 *Start_prec*¹³

The Startdate is coded as precisely as possible. For certain conflicts we can pinpoint the start of the armed conflict down to a single event, taking place on a specific day. For other conflicts, this is not possible, due to lack of precise information. The Startprec (start precision) is coded to highlight the level of certainty for the date set in the Startdate variable.

1. Day, month and year are precisely coded; we have good information on the event.
2. Day is assigned; month and year are precisely coded. The assigned date can either be one of several events that can be classified as the first; it can be the last day in a period when several fatalities have been reported jointly or it can be an event that different sources claim occurred on different dates.
3. Day is unknown; month (or a period of 30 days, not necessarily a calendar month) and year are precisely coded. The day is known to be in a given month or 30 day-period, but we are missing information on an exact date. Day is then set to the last day of the period.
4. Month is assigned; year is coded precisely.
5. Day and month are unknown, year is coded precisely. Day and month are set as precisely as possible. For example, if an event is known to have taken place sometime between January and August, the date is coded as 31 August of the coded year, with a precision of 5.
6. Year is assigned. There is a wide disagreement between different sources, so that not even year can be coded precisely. The start year is assigned based on subjective judgment. Day and month are set as precisely as possible. For example, if an event is known to have taken place sometime between January and August in an assigned year, the date is coded as 31 August, with a precision of 6.

¹² In earlier versions of the dataset the Startdate was assigned a new value whenever there was a complete change on side B. Departing from the new Startdate users could break a conflict up into different periods. However, with the introduction of episodes in the UCDP/PRIO dataset (see Section 7) and with the launching of the new UCDP Dyadic dataset this became redundant.

¹³ The format for coding start dates has been slightly changed in 2014, and then again in 2017, in an attempt to provide more detailed information.

7. Year is missing. No reliable information on the start year is available. Day and month are set as precisely as possible for the first year of recorded activity. For example, if an event is known to have taken place sometime between January and August in an assigned year, the date is coded as 31 August, with a precision of 7.

3.17 *Start_date2*

Startdate2 gives the date, as precise as possible, when *a given episode of conflict activity* reached 25 battle-related deaths. Thus, for each episode of a conflict, a new Startdate2 is coded. In case precise information is lacking, Startdate2 is by default set to 31 December.

An episode is defined as continuous conflict activity. Consequently, a new episode is coded whenever a conflict restarts after one or more year(s) of inactivity. The introduction of episodes within conflicts enables users to distinguish between different phases in the conflict and, potentially, code these as separate conflicts.

3.18 *Start_prec2*

For Version 17.1 of the dataset, the categories for this variable are similar to those of Startprec.

3.19 *Ep_End*

EpEnd is a dummy variable that codes whether the conflict is inactive the following year and an episode of the conflict thus ends. If the conflict is inactive the following year(s), this variable is coded as 1. If not, a 0 is coded. For the latest year in the dataset, it is unknown whether the conflict will be recorded as active or inactive in the following year, and the variable is always given the code -99.

3.20 *Ep_End_Date*

This variable is only coded in years where EpEnd has the value 1. If a conflict year is followed by at least one year of conflict inactivity, the EpEndDate variable lists, as precise as possible, the date that violence stopped.

Previously, this variable has been taken directly from UCDP's Conflict Termination Dataset.¹⁴ However, as UCDP has finalized in its event coding exercise, resulting in the UCDP GED Dataset (<http://ucdp.uu.se/downloads/>), the dates are taken from the latter.

¹⁴ For a presentation of the data, see Kreutz, Joakim (2010) How and When Armed Conflicts End: Introducing the UCDP Conflict Termination Dataset. *Journal of Peace Research* 47(2):243-250. The dataset can be found at <http://ucdp.uu.se/downloads/>. For coding rules and more detailed definition of these variables please see the Conflict Termination Dataset's code book.

3.21 *Ep_End_Prec*

The enddate is coded as precisely as possible. For certain conflicts we can pinpoint the termination of the armed conflict down to a single event, taking place on a specific day. For other conflicts, this is not possible, due to lack of precise information. The Endprec (end precision) is coded to highlight the level of certainty for the date set in the Enddate variable.

1= Day, month and year are precisely coded; we have good information on the event.

2= Day is assigned; month and year are precisely coded. The assigned date can either be one of several events that can be classified as the last; it can be the last day in a period when several fatalities have been reported jointly or it can be an event that different sources claim occurred on different dates.

3= Day is unknown; month (or a period of 30 days, not necessarily a calendar month) and year are precisely coded. The day is known to be in a given month or 30 day-period, but we are missing information on an exact date. Day is then set to the last day of the period.

4= Month is assigned; year is coded precisely.

5= Day and month are unknown, year is coded precisely. Day and month are set as precisely as possible. For example, if an event is known to have taken place sometime between January and August, the date is coded as 31 August of the coded year.

As with the EpEndDate variable (3.19), this variable was previously taken directly from UCDP's Conflict Termination Dataset. However, with the launch of the UCDP GED dataset, both EpEnd and EpEndDate is taken from this dataset.

3.22 *GWNo_A*

To facilitate analytical use of the dataset, country codes for the active state(s) on side A are listed in this field. The country codes are taken from Gleditsch & Ward (2007). In line with Section **Error! Reference source not found.**, in internal conflicts GWNoA lists the country code of the government; in interstate conflicts it arbitrarily lists one of the states and in extrasystemic conflicts it lists the code for the colonial state. This variable may contain more than one country.

GWNo A is a string variable, where the numbers are separated by a comma (',').

3.23 *GWNo_A_2nd*

GWNo A Secondary lists the country codes for the states coded in Side A Secondary (Section **Error! Reference source not found.**).

GWNo A Secondary is a string variable, where the numbers are separated by a comma (',').

3.24 *GWNo_B*

GWNoB lists the country codes for states coded in Side B. GWNoB is only coded in interstate conflicts, since that is the only time a state is active on side B. (see Section 3.6)

GWNoB is a string variable, where the numbers are separated by a comma (',').

3.25 *GWNo_B_2nd*

GWNoB2nd lists the country codes for the states coded in Side B Secondary (see Section 3.8).

GWNoB2nd is a string variable, where the numbers are separated by a comma (',').

3.26 *GWNo_Loc*

This field contains the country code(s) for the state(s) listed in the Location variable. Thus, it lists the country codes for the primary party/parties in the conflict.

GWNoLoc is a string variable, where the numbers are separated by a comma (',').

3.27 *Region*

Region of Location. This variable groups the various conflicts into five geographical categories, dependent on the location of the conflict. Table 2 summarizes these categories:

Table 2: Regions by GWNo

Region	Name	StartGWNo	EndGWNo
1	Europe	200	395
2	Middle East	630	698, not 651
3	Asia	700	990
4	Africa	400	626 and 651
5	Americas	2	165

3.28 *Version*

See Section 2.3.

4 System Membership Description

The definition of a state is crucial to the UCDP/PRIO conflict list, and subsequently also for the dyadic dataset.

State: A state is:

- a) an internationally recognized sovereign government controlling a specified territory,
or

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

The conflict definition, and hence also the dyad definition, is based upon participating governmental actors of independent states, and the classification of a given conflict rests heavily upon the status of the different actors. In order to define the total population of states we use the Gleditsch & Ward (1999) systems membership definition, which is based on the Correlates of War project. Gleditsch & Ward include countries with a population of more than 250,000 that have 'a relatively autonomous administration over some territory', and is 'considered a distinct entity by local actors or the state it is dependent on' (Gleditsch & Ward, 1999: 398).

The Gleditsch & Ward definition differs from our original definition in two main ways. First, for Gleditsch & Ward a state is considered to be a new entity if it is 'considered a distinct entity by local actors' while the conflict definition only recognizes it as a new entity when the 'sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory'. Second, the Gleditsch & Ward states all have a population of more than 250,000 while our conflict definition does not prescribe a specific size of the population. In a few cases, countries with less than 250,000 inhabitants experience conflict or are active as secondary parties in a conflict. In those cases we use the Gleditsch & Ward tentative list of microstates. For example, the microstate Tonga was part of the multinational coalition fighting in Iraq in 2004.

The UCDP/PRIO dataset includes one political entity that lacks a country code (GWN0A is set at -99) as it is excluded from both the Gleditsch & Ward system membership table and the tentative list of microstates: Hyderabad (Hyderabad vs. CPI and Hyderabad vs. India). In 1947 when England granted India independence, a cabinet memorandum declared that the Princely States were free to decide to either join India or Pakistan before August 1947 - or devise their own sovereign political system for self-governance. Hyderabad declared itself independent on 15 August 1947. The state was populated by approximately 15,000,000 inhabitants. The system membership ended when Hyderabad was annexed by India in September 1948, after Winston Churchill had proposed to the United Nations to consider it as an independent state.

In a couple of cases, the dataset has a different sovereignty date for some countries that have experienced conflict:

- Croatia

Gleditsch & Ward: 25 June 1991 (declares independence)

UCDP/PRIO: 27 April 1992 (the new constitution of Yugoslavia)

- Georgia:

Gleditsch & Ward: 6 September 1991 (the USSR's recognition of Georgia's declaration of independence, on 9 April 1991)

UCDP/PRIO: 21 December 1991 (USSR formally ceased to exist)

Comment: While Gleditsch & Ward claim that the USSR recognised Georgia's independence on 6 September, this seems to be incorrect. On 6 September the Soviet Union recognized the Baltic States' independence, but not Georgia's. Instead, the Soviet Union in early September 1991 refused to discuss recognition of the declaration of independence.

Gleditsch & Ward use a slightly modified version of the COW numbering system. The table below presents the system membership table that we base our data tables on. These data are based on Gleditsch & Ward's 2013 version of the list of independent state, which is update through 31 December 2012.

Table 3 List of System Members Since 1946

StateNum	StateAb	StateName	StartYear	EndYear
2	USA	United States of America	1946	2012
20	CAN	Canada	1946	2012
31	BHM	Bahamas	1973	2012
40	CUB	Cuba	1946	2012
41	HAI	Haiti	1946	2012
42	DOM	Dominican Republic	1946	2012
51	JAM	Jamaica	1962	2012
52	TRI	Trinidad and Tobago	1962	2012
53	BAR	Barbados	1966	2012
70	MEX	Mexico	1946	2012
80	BLZ	Belize	1981	2012
90	GUA	Guatemala	1946	2012
91	HON	Honduras	1946	2012
92	SAL	El Salvador	1946	2012
93	NIC	Nicaragua	1946	2012
94	COS	Costa Rica	1946	2012
95	PAN	Panama	1946	2012
100	COL	Colombia	1946	2012
101	VEN	Venezuela	1946	2012

StateNum	StateAb	StateName	StartYear	EndYear
110	GUY	Guyana	1966	2012
115	SUR	Surinam	1975	2012
130	ECU	Ecuador	1946	2012
135	PER	Peru	1946	2012
140	BRA	Brazil	1946	2012
145	BOL	Bolivia	1946	2012
150	PAR	Paraguay	1946	2012
155	CHL	Chile	1946	2012
160	ARG	Argentina	1946	2012
165	URU	Uruguay	1946	2012
200	UKG	United Kingdom	1946	2012
205	IRE	Ireland	1946	2012
210	NTH	Netherlands	1946	2012
211	BEL	Belgium	1946	2012
212	LUX	Luxembourg	1946	2012
220	FRN	France	1946	2012
225	SWZ	Switzerland	1946	2012
230	SPN	Spain	1946	2012
235	POR	Portugal	1946	2012
260	GFR	German Federal Republic	1949	2012
265	GDR	German Democratic Republic	1949	1990
290	POL	Poland	1946	2012
305	AUS	Austria	1946	2012
310	HUN	Hungary	1946	2012
315	CZE	Czechoslovakia	1946	1992
316	CZR	Czech Republic	1993	2012
317	SLO	Slovakia	1993	2012
325	ITA	Italy/Sardinia	1946	2012
338	MLT	Malta	1964	2012
339	ALB	Albania	1946	2012
340	SER	Serbia	2006	2012
341	MNG	Montenegro	2006	2012
343	MAC	Macedonia (FRY)	1991	2012
344	CRO	Croatia	1991	2012
345	YUG	Yugoslavia (Serbia)	1946	2006
346	BOS	Bosnia-Herzegovina	1992	2012
347	KOS	Kosovo	2008	2012
349	SLV	Slovenia	1992	2012
350	GRC	Greece	1946	2012
352	CYP	Cyprus	1960	2012

StateNum	StateAb	StateName	StartYear	EndYear
355	BUL	Bulgaria	1946	2012
359	MLD	Moldova	1991	2012
360	RUM	Rumania	1946	2012
365	RUS	Russia (Soviet Union)	1946	2012
366	EST	Estonia	1991	2012
367	LAT	Latvia	1991	2012
368	LIT	Lithuania	1991	2012
369	UKR	Ukraine	1991	2012
370	BLR	Belarus (Byelorussia)	1991	2012
371	ARM	Armenia	1991	2012
372	GRG	Georgia	1991	2012
373	AZE	Azerbaijan	1991	2012
375	FIN	Finland	1946	2012
380	SWD	Sweden	1946	2012
385	NOR	Norway	1946	2012
390	DEN	Denmark	1946	2012
395	ICE	Iceland	1946	2012
402	CAP	Cape Verde	1975	2012
404	GNB	Guinea-Bissau	1974	2012
411	EQG	Equatorial Guinea	1968	2012
420	GAM	Gambia	1965	2012
432	MLI	Mali	1960	2012
433	SEN	Senegal	1960	2012
434	BEN	Benin	1960	2012
435	MAA	Mauritania	1960	2012
436	NIR	Niger	1960	2012
437	CDI	Cote D'Ivoire	1960	2012
438	GUI	Guinea	1958	2012
439	BFO	Burkina Faso (Upper Volta)	1960	2012
450	LBR	Liberia	1946	2012
451	SIE	Sierra Leone	1961	2012
452	GHA	Ghana	1957	2012
461	TOG	Togo	1960	2012
471	CAO	Cameroon	1960	2012
475	NIG	Nigeria	1960	2012
481	GAB	Gabon	1960	2012
482	CEN	Central African Republic	1960	2012
483	CHA	Chad	1960	2012
484	CON	Congo	1960	2012
490	DRC	Congo, Democratic Republic of	1960	2012

StateNum	StateAb	StateName	StartYear	EndYear
		(Zaire)		
500	UGA	Uganda	1962	2012
501	KEN	Kenya	1963	2012
510	TAZ	Tanzania/Tanganyika	1961	2012
511	ZAN	Zanzibar	1963	1964
516	BUI	Burundi	1962	2012
517	RWA	Rwanda	1962	2012
520	SOM	Somalia	1960	2012
522	DJI	Djibouti	1977	2012
530	ETH	Ethiopia	1946	2012
531	ERI	Eritrea	1993	2012
540	ANG	Angola	1975	2012
541	MZM	Mozambique	1975	2012
551	ZAM	Zambia	1964	2012
552	ZIM	Zimbabwe (Rhodesia)	1965	2012
553	MAW	Malawi	1964	2012
560	SAF	South Africa	1946	2012
565	NAM	Namibia	1990	2012
570	LES	Lesotho	1966	2012
571	BOT	Botswana	1966	2012
572	SWA	Swaziland	1968	2012
580	MAG	Madagascar (Malagasy)	1960	2012
581	COM	Comoros	1975	2012
590	MAS	Mauritius	1968	2012
600	MOR	Morocco	1956	2012
615	ALG	Algeria	1962	2012
616	TUN	Tunisia	1956	2012
620	LIB	Libya	1951	2012
625	SUD	Sudan	1956	2012
626	SSD	South Sudan	2011	2012
630	IRN	Iran (Persia)	1946	2012
640	TUR	Turkey/Ottoman Empire	1946	2012
645	IRQ	Iraq	1946	2012
651	EGY	Egypt	1946	2012
652	SYR	Syria	1946	2012
660	LEB	Lebanon	1946	2012
663	JOR	Jordan	1946	2012
666	ISR	Israel	1948	2012
670	SAU	Saudi Arabia	1946	2012
678	YEM	Yemen (Arab Republic of Yemen)	1946	2012

StateNum	StateAb	StateName	StartYear	EndYear
680	YPR	Yemen, People's Republic of	1967	1990
690	KUW	Kuwait	1961	2012
692	BAH	Bahrain	1971	2012
694	QAT	Qatar	1971	2012
696	UAE	United Arab Emirates	1971	2012
698	OMA	Oman	1946	2012
700	AFG	Afghanistan	1946	2012
701	TKM	Turkmenistan	1991	2012
702	TAJ	Tajikistan	1991	2012
703	KYR	Kyrgyz Republic	1991	2012
704	UZB	Uzbekistan	1991	2012
705	KZK	Kazakhstan	1991	2012
710	CHN	China	1946	2012
711	TBT	Tibet	1946	1950
712	MON	Mongolia	1946	2012
713	TAW	Taiwan	1949	2012
731	PRK	North Korea	1948	2012
732	ROK	South Korea	1948	2012
740	JPN	Japan	1946	2012
750	IND	India	1947	2012
760	BHU	Bhutan	1949	2012
770	PAK	Pakistan	1947	2012
771	BNG	Bangladesh	1971	2012
775	MYA	Myanmar (Burma)	1948	2012
780	SRI	Sri Lanka	1948	2012
781	MAD	Maldives	1965	2012
790	NEP	Nepal	1946	2012
800	THI	Thailand	1946	2012
811	CAM	Cambodia (Kampuchea)	1953	2012
812	LAO	Laos	1954	2012
816	DRV	Vietnam, Democratic Republic of	1954	2012
817	RVN	Vietnam, Republic of	1954	1975
820	MAL	Malaysia	1957	2012
830	SIN	Singapore	1965	2012
835	BRU	Brunei	1984	2012
840	PHI	Philippines	1946	2012
850	INS	Indonesia	1946	2012
860	ETM	East Timor	2002	2012
900	AUL	Australia	1946	2012
910	PNG	Papua New Guinea	1975	2012

StateNum	StateAb	StateName	StartYear	EndYear
920	NEW	New Zealand	1946	2012
940	SOL	Solomon Islands	1978	2012
950	FJI	Fiji	1970	2012

5 References

- Allansson, Marie; Erik Melander & Lotta Themnér (2017) Organized violence, 1989-2016. *Journal of Peace Research* 54(4).
- Gleditsch, Kristian S. & Michael D. Ward (1999) Interstate System Membership: A Revised List of the Independent States since 1816. *International Interactions* 25: 393–413.
- Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand (2002) Armed Conflict 1946–2001: A New Dataset. *Journal of Peace Research* 39(5): 615–637.
- Gleditsch, Kristian S. Skrede & Michael D. Ward (2013) System membership case description list.
- Held, Birger (1993) Armed Conflicts over Government and Territory 1989–91 in Heldt, Birger, *Armed Conflict 1990–91*. Uppsala: Uppsala University.
- Kreutz, Joakim (2010) How and When Armed Conflicts End: Introducing the UCDP Conflict Termination Dataset. *Journal of Peace Research* 47(2):243-250.
- Themnér, Lotta & Peter Wallensteen (2014) Armed Conflict, 1946-2013 *Journal of Peace Research* 51(4).