

UCDP Non-State Conflict Codebook

Version 18.1

Uppsala Conflict Data Program
Department of Peace and Conflict Research, Uppsala University

This version compiled and updated by Marie Allansson and Mihai Croicu (2017)

Replacing the earlier versions by Therése Pettersson (2014), Marcus Nilsson & Therése Pettersson (2013), Therése Pettersson (2012), Ralph Sundberg (2010) and Joakim Kreutz & Kristine Eck (2005)

Citation for the dataset:

Sundberg, Ralph, Kristine Eck and Joakim Kreutz, 2012, "Introducing the UCDP Non-State Conflict Dataset", *Journal of Peace Research*, March 2012, 49:351-362

(When appropriate) also cite this codebook. Always include the Version number in analyses using the dataset.

Introduction

This document describes the Non-State Conflict Dataset, a project within the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University. The UCDP Non-State conflict project has been developed with support from the Human Security Report Project, Simon Fraser University, in Vancouver, Canada.

In the development of the definition of non-state conflict, the input from Kristine Eck, Peter Wallensteen, Margareta Sollenberg, Lotta Harbom, Ralph Sundberg, Stina Högbladh, Therése Pettersson and Johan Brosché, have been instrumental. The UCDP non-state conflict project is also grateful for additional advice and feedback from Andrew Mack, Zoe Nielsen, Ole Magnus Thiesen, and others.

Case-specific information about the cases of non-state conflict is available at www.ucdp.uu.se. Questions regarding the definitions and the content of the dataset can be directed to ucdp@pcr.uu.se

Definition of Non-State conflict

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

The separate elements of the definition are operationalized as follows:

- (1) *Use of armed force*: the use of arms, resulting in deaths.
 - (1.1) *Arms*: any material means, e.g. manufactured weapons but also sticks, stones, fire, water, etc.
- (2) *25 deaths*: a minimum of 25 battle-related deaths per year
 - (2.1) *battle-related deaths*: deaths directly related to the use of armed force between the warring groups
- (3) *Organized groups*: consists of either
 - (3.1) formally organized groups: any non-governmental group of people having announced a name for their group and using armed force against another similarly formally organized group, *or*
 - (3.2) informally organized groups: any group without an announced name, but who uses armed force against another similarly organized group, where the violent activity meets the following requirement:
 - (3.2.a.) there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other
- (4) *State*: a state is
 - (4.1) an internationally recognized sovereign government controlling a specified territory, *or*
 - (4.2) an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.
- (5) *Government*: the party controlling the capital of the state

Variables in the Non-State conflict Dataset

Variable name	Content	Type
ConflictID	<p>The unique identifier of the Non-state conflict.</p> <p>In version 17.1 of the dataset, the ID system for conflicts, actors and dyads was changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit http://ucdp.uu.se/downloads/.</p>	Integer
DyadID	<p>The unique identifier of the non-state dyad (a pair of two opposing actors).</p> <p>Note that one non-state conflict has, per definition, one and only one non-state dyad. The inclusion of both dyad IDs and conflict IDs in the dataset is meant to allow easier integration of this dataset with other UCDP products such as the UCDP/PRIO Armed Conflict Dataset, the UCDP Dyadic Dataset or the UCDP GED.</p> <p>In version 17.2 of the dataset, the ID system for conflicts, actors and dyads has been changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit http://ucdp.uu.se/downloads/.</p>	Integer
Org	<p>This variable indicates the organizational level of the warring sides. The level of organization is determined according to the following categories:</p> <p>Organizational level 1 (formally organized groups): Rebel groups and other organized groups that have a high enough level of organization so as to be possible to include in the state-based armed conflict category. These include rebel groups with an announced name, as well as military factions (For organization captures fighting between highly organized rebel groups and fatalities are recorded according to the criteria set for battle-related deaths in the state-based conflict category.</p>	Integer

Organizational level 2 (informally organized groups):

Groups composed of supporters and affiliates to political parties and candidates. These are commonly not groups that are permanently organized for combat, but who at times use their organizational structures for such purposes. In addition to supporters of political parties and candidates, included in this category is also fighting between groups composed of supporters of other organizations such as the supporters of al-Ahly football team fighting against the supporters of al-Masry football team in Egypt 2012. Battle-related deaths are recorded according to section 3.2.a of the definition of non-state conflict.

Organizational level 3 (informally organized groups):

Groups that share a common identification along ethnic, clan, religious, national or tribal lines. These are not groups that are permanently organized for combat, but who at times organize themselves along said lines to engage in fighting. This level of organization captures aspects of what is commonly refer conflicts', in that conflict stand identity. Battle-related deaths are recorded according to section 3.2.a of the definition of non-state conflict

SideA

The party that constitute Side A in the conflict. For each conflict the parties are listed in alphabetical order, using the latest known names of the parties involved

Integer

SideAID

The ID of the groups that make up Side A. For conflicts with multiple actors fighting together a temporary coalition ID has been assigned.

Integer

In version 17.1 of the dataset, the ID system for conflicts, actors and dyads was changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit <http://ucdp.uu.se/downloads/>.

SideAComponents ¹	For conflicts with multiple actors fighting together, the string actors' separate Actor IDs, are separated by a comma.	
SideB	The party that constitute Side B in the conflict. For each conflict the parties are listed in alphabetical order, using the latest known names of the parties involved	Integer
SideBID	The ID of the groups that make up Side B. For conflicts with multiple actors fighting together a temporary coalition ID has been assigned. In version 17.1 of the dataset, the ID system for conflicts, actors and dyads was changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit http://ucdp.uu.se/downloads/ .	Integer
SideBComponents	For conflicts with multiple actors fighting together, the actors' separate Actor IDs, are separated by a comma.	string
StartDate	The first time there is a recorded event in a given dyad that results in at least one fatality. This date is the same for all years in which the conflict has been active, regardless of whether the conflict has been active in several episodes or not.	
StartPrec	The Startdate is coded as precisely as possible. For certain conflicts we can pinpoint the start of the conflict down to a single event, taking place on a specific day. Integer 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; there is good information on the event. 2. Day is assigned; month and year are precisely coded. This precision score is assigned if the first event which causes at least one fatality takes place within a period of 2-6 days. 3. Day is unknown; month (or a period of 30 days, not necessarily a calendar month) and year are precisely	

¹ In the 18.1 version of the non-state dataset, Syria data is highly preliminary. Instead of coalitions, the datasets uses so-called operation rooms as parties. Components of operation rooms are listed as IDs where possible, but names are not available in the dataset.

coded. The day is known to be in a given period of 30 days. The date is set to the last date of the period.

4. Month is assigned, year is precisely coded. The date is set to the last day of the assigned month.

5. Day and month are assigned, year is precisely coded. Day and month are set as precisely as possible. For example, if an event is known to have taken place between March and July, the date is set to 31 July with precision score 5.

This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.

StartDate2	StartDate2 gives the date, as precise as possible, when a given episode of conflict activity reached 25 battle-related deaths.	Date
StartPrec2	Precision scores calculated as per StartPrec above	integer
<p>This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.</p>		
EpEnd	EpEnd is a binary 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.	Integer
EpEndDate	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 last date of recorded combat.	Date
EpEndPrec	Precision scores calculated as per StartPrec above	integer
<p>This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.</p>		
Year	The year of observation	Integer

FatBest	<p>The best fatality estimate for the given conflict-year.</p> <p>This is an automatic aggregation (summing) of all the Best figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.</p>	Integer
FatLow	<p>The low fatality estimate for the given conflict-year.</p> <p>This is an automatic aggregation (summing) of all the Low figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.</p>	Integer
FatHigh	<p>The high fatality estimate for the given conflict-year.</p> <p>This is an automatic aggregation (summing) of all the High figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.</p>	Integer
Location	<p>The countries where fighting took place in the conflict-year. Comma-separated if multiple.</p> <p>WARNING: This variable SHOULD NEVER be used for any geographical or spatial analyses of conflict as the distribution of violence as well as the relative magnitude of violence by country is not captured. In effect a country is listed here if even one dead in the given conflict has occurred in that country.</p> <p>In fact, UCDP provides much better geographic coverage of conflict (including distribution of violence for each conflict and each country) in the UCDP Georeferenced Event Dataset (GED).</p> <p>Like the UCDP Non-State Conflict Dataset, GED is global and covers the same period (1989-2017).</p>	string
GWNo	<p>The Gleditsch and Ward code for the countries where fighting took place in the conflict-year. Comma-separated if multiple.</p> <p>WARNING: This variable SHOULD NEVER be used for any geographical or spatial analyses of conflict as the distribution of violence as well as the relative magnitude of violence by country is not captured. In effect a country is listed here if even one dead in the given conflict has occurred in that country.</p> <p>In fact, UCDP provides much better geographic coverage of conflict (including distribution of violence for each conflict and each country) in the UCDP Georeferenced Event Dataset (GED).</p> <p>Like the UCDP Non-State Conflict Dataset, GED is global and covers the same period (1989-2017).</p>	string

Region	The continents (regions) where fighting took place:	string
Version	The version of the dataset: 18.1	float

Data Collection Methods:

This dataset is the result of:

1. an automatic filtering and aggregation of the UCDP Georeferenced Event Dataset from incident/event level to the conflict/dyad-year level.
2. information gathering and coding of a number of extra variables at the aggregate conflict or actor level (such as organization type).

The original reporting underlying the dataset is collected from three sets of sources:

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

The process is done in a "two-pass" system, first by consulting newswire sources for the entire globe then by consulting local/specialized sources based on information obtained from the first pass.

A detailed description of the process including a detailed description of the passes and the search-strings employed is provided in the UCDP GED Codebook version 18.1, Section 4.

Format availability:

The data is available in CSV (respecting the RFC 4180 specification), Excel (XLSX), Rdata (3.x version) and STATA 14 format.

The data is available for machine-to-machine interaction through a public API. Documentation for how to use the API is available at <http://ucdp.uu.se/apidocs>.

A note on UCDP id changes

In version 17.1 of all UCDP datasets, the ID system for conflicts, actors and dyads was changed in order to make them unique across all UCDP core datasets and all UCDP types of violence. This will allow easier aggregation and disaggregation of data as well as simplify data management for users, especially when combining multiple UCDP products together.

For example, a non-state conflict ID will no longer be able to have the same ID as a (different) state-based conflict ID or as a case of one-sided violence (which was possible before).

Further, actor/side IDs representing a government have been decoupled from their corresponding Gleditsch and Ward country codes (GWNo). This will allow, in the future, the identification of regime changes within a country etc. Gleditsch and Ward country codes for state sides/actors will continue to be provided in separate columns from their sides ID (i.e. side_a_id is different from gwno_a but represents the same entity).

This means all IDs used in this version of the dataset are no longer compatible with those in older versions of UCDP products. Further, this means that external products relying on UCDP IDs for data management tasks will have to be adapted to work with the new ID systems.

A conversion table between the new and old ID systems is available here: <http://ucdp.uu.se/downloads/>.

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 18.1 of the UCDP Non-state Conflict Dataset. For every new release, substantial changes will be documented in a separate document. 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.

References

Gleditsch, Kristian S. & Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25: 393-413