**Atrocity Crime Events (ACE) Operational Codebook**

**Helena Hinkkainen**

**Adrian Gallagher**

# Table of Contents

[Table of Contents 1](#_Toc167196472)

[Introduction to ACE 2](#_Toc167196473)

[Overview 2](#_Toc167196474)

[Variables 3](#_Toc167196475)

[Methodology 5](#_Toc167196476)

[Event ID and date 6](#_Toc167196477)

[ACE events 7](#_Toc167196478)

[Actors 7](#_Toc167196479)

[Victims 8](#_Toc167196480)

[Location 8](#_Toc167196481)

[Sources 10](#_Toc167196482)

# Introduction to ACE

## Overview

The purpose of this codebook is to firstly explain the purpose of the project “Atrocity Crime Events (ACE)”and the need for the data the project is collecting. The introductory part also outlines the variables covered in the dataset as well as the methodology used in the data collection. The rest of the codebook provides further detail of all the variables coded.

The overall ACE project is motivated by the need to better understand the behaviour of non-state armed groups in perpetrating atrocity crimes such as crimes against humanity, ethnic cleansing and war crimes. Atrocity perpetration in the past used to be mainly the domain of the state apparatus, however, non-state armed groups have increasingly been involved in such crimes. While there are many data collection efforts related to non-state actor behaviour, to the best of our knowledge, none of these existing datasets cover the multitude of acts associated with specific atrocity crimes. Some of the more fine-grained event data collections such as the Political Instability Task Force (PITF) Worldwide Atrocities Dataset focuses on killings similarly to the Uppsala Georeferenced Events Dataset (UCDP GED). The Armed Conflict Location & Event Data Project (ACLED) codes some event types that are deemed atrocity events in international law (Rome Statute), however, multiple event types from the Rome Statute are not captured by any of these datasets. Some country specific data collections also exist in relation to human rights violations etc. but the cases in this study and the focus on atrocity events has not been done previously. The aim of this project is thus to complement the existing data collections by focusing on atrocity event types that we do not have systematic data about.

This data collection is based on six countries Central African Republic, the Democratic Republic of Congo, Iraq, Nigeria, Syria, and Somalia with a focus on non-state actor perpetrated atrocity events. Since the purpose of the project is to build an event dataset of various different types of atrocities perpetrated by non-state actors in these countries, the countries were selected based on UN reports of such events in these countries as well as on evaluation of data availability. The country selection was done in this way because the purpose of this data collection is not to compare atrocities in other contexts or to predict in which other countries they may or may not happen, but to rather examine under what conditions in these countries do non-state armed groups perpetrate such atrocities.

The focus, therefore, is fine-grained event data collection of different types of atrocity crimes in these countries. Mass atrocities are defined as: ‘Instances of “large-scale, systematic violence against civilian populations.” Although the term mass atrocities has no formal legal definition, it usually refers to genocide […] crimes against humanity, war crimes, and ethnic cleansing.’ (Strauss 2016, 29-30). Since the terms large scale and systematic continue to be debated, we use a broad criterion of inclusion here by focusing on atrocity crimes rather than ‘mass’ atrocities strictly speaking. These event types are derived from the Rome Statute. More specifically, the unit of observation in ACE is the event where a named state or non-state actor is involved on a given day in a specific location. The coders for each country aim to code each individual event with the best precision regarding the type of event, location, perpetrator and victims. The coders searched through and read reports for 12 different types of atrocity crimes, however, only 8 types of atrocity crimes were present in the cases included in the data. These event types are:  Enslavement, Mutilation, Imprisonment, Torture, Rape, Humanitarian Attacks, Peacekeeping Attacks and Sexual Slavery. In addition to these, the coders searched for reports on Forced Pregnancy, Enforced Prostitution, Enforced Sterilisation and Forced Abortion, but no reports were found that our coders were able to include for the purposes of this dataset. While ACE covers altogether 8 event types, not all the event types occur in all of the cases included in the dataset.

While the overall aim of the project behind this data collection is to analyse the correlates of atrocity crime perpetration by non-state actors in these countries, the event data coding necessitates that all atrocity events are coded, even if perpetrated by, for example, state actors. Only in doing so will we be able to analyse whether non-state actor atrocity crime perpetration is influenced by other actors using similar tactics.

## Variables

*Table 1: ACE variable columns*

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
|  |  |
| **maid** | An individual atrocity event numeric identifier by country ISO code and date |
|  |  |
|  |  |
| **eyear** | The year in which an event took place |
|  |  |
| **emonth** | The month in which an event took place |
|  |  |
| **eday** | The day in which an event took place |
| **startdate** | When the exact timing of the event is unknown, the startdate of the event if coded |
|  |  |
| **eprec** | A numeric code indicating the level of certainty of the date coded for the event |
|  |  |
| **clarity** | A numeric code about the clarity of the actual event description |
| **comments** | A text field for the coder to write comments about the event and/or coding |
| **etype** | A numeric variable coding the type of event |
|  |  |
| **mname** | A description of the name of the actor perpetrating the atrocity event |
|  |  |
| **mnotes** | A text field for the coder to write comments about the actors |
| **nvictim** | A numerical variable coding the number victims in a given event |
| **vkilled** | A numerical variable coding the number victims killed in a given event |
| **vinjured** | A numerical variable coding the number victims injured in a given event |
| **vnotes** | A text field for the coder to write comments about the victims |
| **country** | The name of the country in which the event took place |
| **iso3** | A numeric code for each individual country |
| **admin0** | GADM country ID |
|  |  |
|  |  |
| **admin1** | GADM province ID, first order administrative sub-unit of the country |
|  |  |
| **name1** | Province name, first order administrative sub-unit of the country |
|  |  |
|  |  |
| **admin2** | GADM district ID, second order administrative sub-unit of the country |
|  |  |
| **name2** | District name, second order administrative sub-unit of the country |
| **latitude** | The latitude of the location |
|  |  |
| **longitude** | The longitude of the location |
|  |  |
| **gprec** | A numeric code indicating the level of certainty of the location  coded for the event |
| **nsource** | The source of the event report |
|  |  |
| **dsource** | Publication date of the source report |
| **description** | A short description of the event |

## 

## Methodology

The data collection methodology is based on coding news reports extracted from LexisNexis. The extraction of news reports from LexisNexis has been narrowed down by using specific search terms for each event type, including the countries in this project. The focus is primarily on English language sources and where necessary, the geography filter is used to narrow down results based on the location of the event. Since atrocity events can happen outside active conflicts, we do not narrow the search based on any given time periods. The initial search of Lexis Nexis suggests that most of the reports fall withing the last 30-40 years, however, some earlier reports may exist.

In preparation for the extraction, all the atrocity event category searches were attempted and based on the number of hits, some event categories were deemed not feasible for the purposed of the coding. In addition, some categories, such as attacks, killings and battles are not included as already existing data on these can be merged with the current dataset and the number of hits for such events would overwhelm the coders whose focus is on atrocity events that no systematic data is yet available. While we do not explicitly code killings, we do attempt to code whether each event led to any fatalities or injuries.

Once a set of news reports have been identified from Lexis Nexis, the coders skim through the reports based on headlines/short descriptions and select to read through the ones that seem to constitute an event (as opposed to, for example, reports about UN meetings to discuss atrocities etc.). The coders then write a short description of the event on the dataset and code the rest of the variables in the dataset with best precision possible.

The coding of the events is based on this codebook and is conducted by six human coders, each specialising in one of the countries in question. However, in order to ensure intercoder reliability, the coders double coded approximately 30 reports each week across the cases.

# 

# Event ID and date

**ATROCITY EVENT ID**

**(maid)** Numeric Variable

A unique identifier for each event in the ACE dataset.

The identifier consists of a numeric ISO country code followed by a 12-digit ACE ID system:

ACE covers 6 countries with the following 3 digit ISO 3166-1 numeric codes:

1. The Central African Republic ISO code: 140
2. The Democratic Republic of Congo ISO code: 180
3. Iraq ISO code: 368
4. Nigeria ISO code: 566
5. Somalia ISO code: 706
6. Syria ISO code: 760

* First 8 numbers – date recorded “ddmmyyyy”.
* Last 4 numbers – sequential case number for the given day (0001, 0002 etc). This is “0001” unless there is more than one case occurring on the same date. For example, an event occurring in Nigeria on 13th April 1999 would be numbered as “566130419990001”. An additional event recorded in Nigeria for the same day would be “566130419990002”.

**ATROCITY EVENT TIMING**

**(eyear)** Numeric variable

This variable codes the year in which the event occurred. Coded as 4-digits, -99 if unknown.

**(emonth)** Numeric variable

This variable codes the month in which the event occurred. Coded as 2-digits, -99 if unknown.

**(eday)** Numeric variable

This variable codes the day when the event started. Coded as 2-digits, -99 if unknown.

**(startdate)** Numeric variable

Where no eday or emonth or eyear can be found for an event but a time frame is given, specifying a year, this year or starting year of a range can be used in the start date in order to code the event. For example, emonth and eday may be “00” and the event started in 2013, so startdate would be “20130000”. Coded as yyyymmdd. This variable field is left blank if preceding variables are coded.

**(eprec)** Numeric variable

How precise the information is about the date of an event. Coded as:

1: exact date of event is known;

2: only the week of the event is known

3: only the month of the event is known

4: the date of the event is known only within a range longer than one month.

**(clarity)** Numeric variable

How clear is the information about a specific event?

The clarity of the event is ranked as below:

* 1. (high) for events where the reporting allows the coder to identify the event in full. That is, events where the individual happening is described by the original source in a sufficiently detailed way as to identify individual incidents, i.e. separate activities of mutilations in a single location: Example of such reporting: “2 men were mutilated by Boko Haram fighters when they stormed a city hall in Kaduna on the 24th of March 2012.”
  2. (lower) for events where an aggregation of information was already made by the source material that is impossible to undo in the coding process. Such events are described by the original source only as aggregates (totals) of multiple separate activities spanning over a longer period than a single, clearly defined day. Examples of such reporting: “The Syrian government has announced that 18 civilians have been tortured in the last week alone by ISIS forces operating in Northern Syria by the Turkish border when trying flee the conflict.”

**(comments)** String variable

Any comments about the event/coding specificity/difficulty.

# ACE events

**(etype)** Numeric variable

Event type. Coded as:

Enslavement = 6

Imprisonment = 8

Torture = 9

Rape = 10

Sexual slavery = 11

Humanitarian attacks = 24

Peacekeeping attacks = 25

Mutilation = 31

# Actors

**ATROCITY PERPETRATING ACTORS**

**(mname)** String variable

The name of the atrocity perpetrator/group.

**(mnotes)**

Any notes/comments about the perpetrator/s and difficulty of coding the actor.

# Victims

**(nvictim)** Numeric variable

A variable recording the number of victims reported in the event. 0 if none and -99 if not known.

**(vkilled)** Numeric variable

A variable recording the number of victims reported killed in the event. 0 if none and -99 if not known.

**(vinjured)** Numeric variable

A variable recording the number of victims reported injured in the event. 0 if none and -99 if not known.

**(vnotes)** String variable

Any notes describing the victims.

# Location

**(country)** String variable

Name of the country the event took place. Coded as:

Central African Republic

Democratic Republic of Congo

Iraq

Nigeria

Somalia

Syria

**(iso3)** Numeric variable

Coded using the 3 digit ISO 3166-1 numeric codes:

1. The Central African Republic ISO code: 140
2. The Democratic Republic of Congo ISO code: 180
3. Iraq ISO code: 368
4. Nigeria ISO code: 566
5. Somalia ISO code: 706
6. Syria ISO code: 760

**(admin0)** Numeric variable

GADM country ID. Coded as:

1. The Central African Republic GADM country ID: 46
2. The Democratic Republic of Congo GADM country ID: 63
3. Iraq ISO GADM country ID: 108
4. Nigeria GADM country ID: 163
5. Somalia GADM country ID: 210
6. Syria GADM country ID: 224

**(admin1)** Numeric variable

GADM province ID, first order administrative sub-unit of the country.

**(name1)** String variable

Province name, first order administrative sub-unit of the country.

**(admin2)** Numeric variable

GADM district ID, second order administrative sub-unit of the country. This variable is left blank if unknown.

**(name2)** String variable

District name, second order administrative sub-unit of the country. This variable is left blank if unknown.

**(latitude)** Numeric variable

Latitude (in decimal degrees) of the event. This variable is left blank if unknown.

The coordinates are fixed to the World Geodetic System of 1984 (WGS 84), EPSG SRID 4326. These coordinates are specified in decimal degrees with a precision up to 6 decimal figures (e.g. 75.920211).

In general, for both latitude and longitude, when specific coordinates are not available, the event is coded to the smallest subnational unit available (normally the coordinates are for the centroid of this unit). For example, in most cases this would mean the 2nd order administrative unit, in some cases 1st order if 2nd order administrative unit are not identified. In rare cases, leave blank if even the 1st order admin unit cannot be identified. If that is the case, the **gprec** variable below is coded as 5.

**(longitude)** Numeric variable

Longitude (in decimal degrees) of the event. This variable is left blank if unknown.

The coordinates are fixed to the World Geodetic System of 1984 (WGS 84), EPSG SRID 4326. These coordinates are specified in decimal degrees with a precision up to 6 decimal figures (e.g. 75.920211).

**(gprec)** Numeric variable

This variable captures the precision of the geographical location of the event. Coded as follows:

1. event occurred in city/village/town and lat/long is for that location
2. event occurred in city/village/town and no lat/long could be found, so coordinates are for centroid of smallest subnational administrative region identified
3. event did not occur in city/village/town, so coordinates are for centroid of smallest subnational administrative region identified
4. no 2nd order or smaller region could be identified, so coordinates are for center of 1st order administrative region
5. no 1st order administrative region could be identified for the location of the attack, so latitude and longitude are unknown

# Sources

**(nsouce)** String variable

Name of the source.

**(dsource)** Numeric variable

The date the source was published. Written in a ddmmyyyy format.

**(description)** String variable

A short description about the event.