Social Conflict Analysis Database (SCAD)
Version 3.3

Codebook and coding procedures

Idean Salehyan
University of North Texas

&

Cullen Hendrix
Korbel School, University of Denver

Updated September 24, 2017

*(Before using SCAD 3.X, please read below about "geographic information.")

The Social Conflict Analysis Database (SCAD) contains information on protests, riots, strikes, and other social disturbances in Africa, Latin American and the Caribbean. Whereas conflict data is generally available for large-scale events such as civil and international war, the purpose of this dataset is to compile information on other types of social and political disorder. See www.scaddata.org for further information about the project.

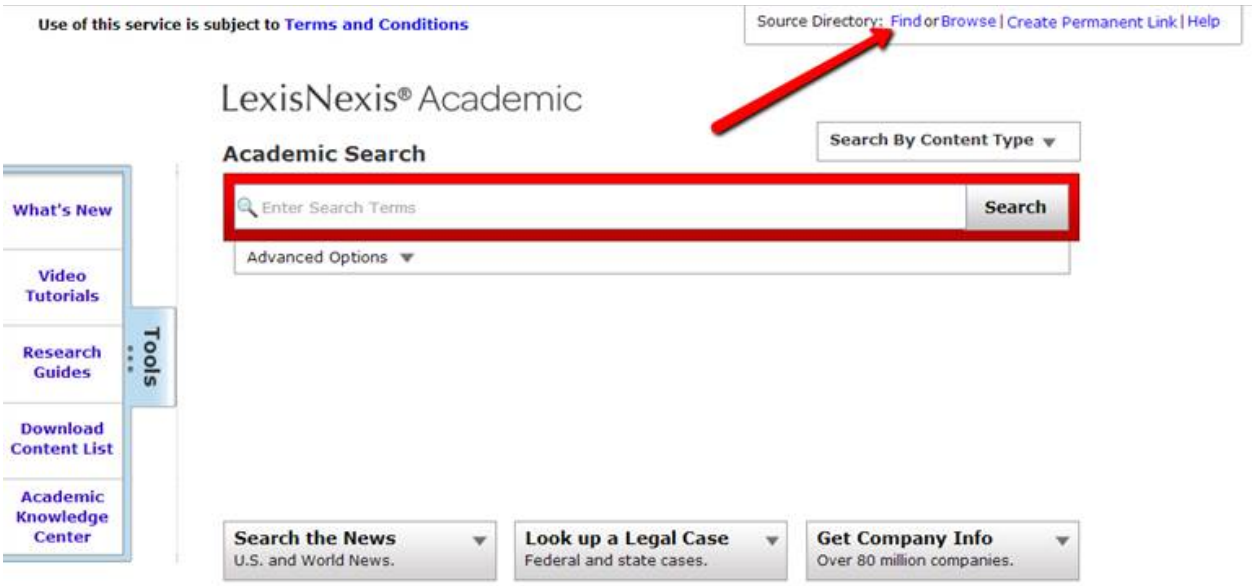
Countries with populations greater than 1,000,000 in 1990 are coded. The primary source of information for this dataset comes from searches of Associated Press and *Agence France Presse* newswires, as compiled by the Lexis-Nexis news service (http://www.lexisnexis.com/).

Search methodology

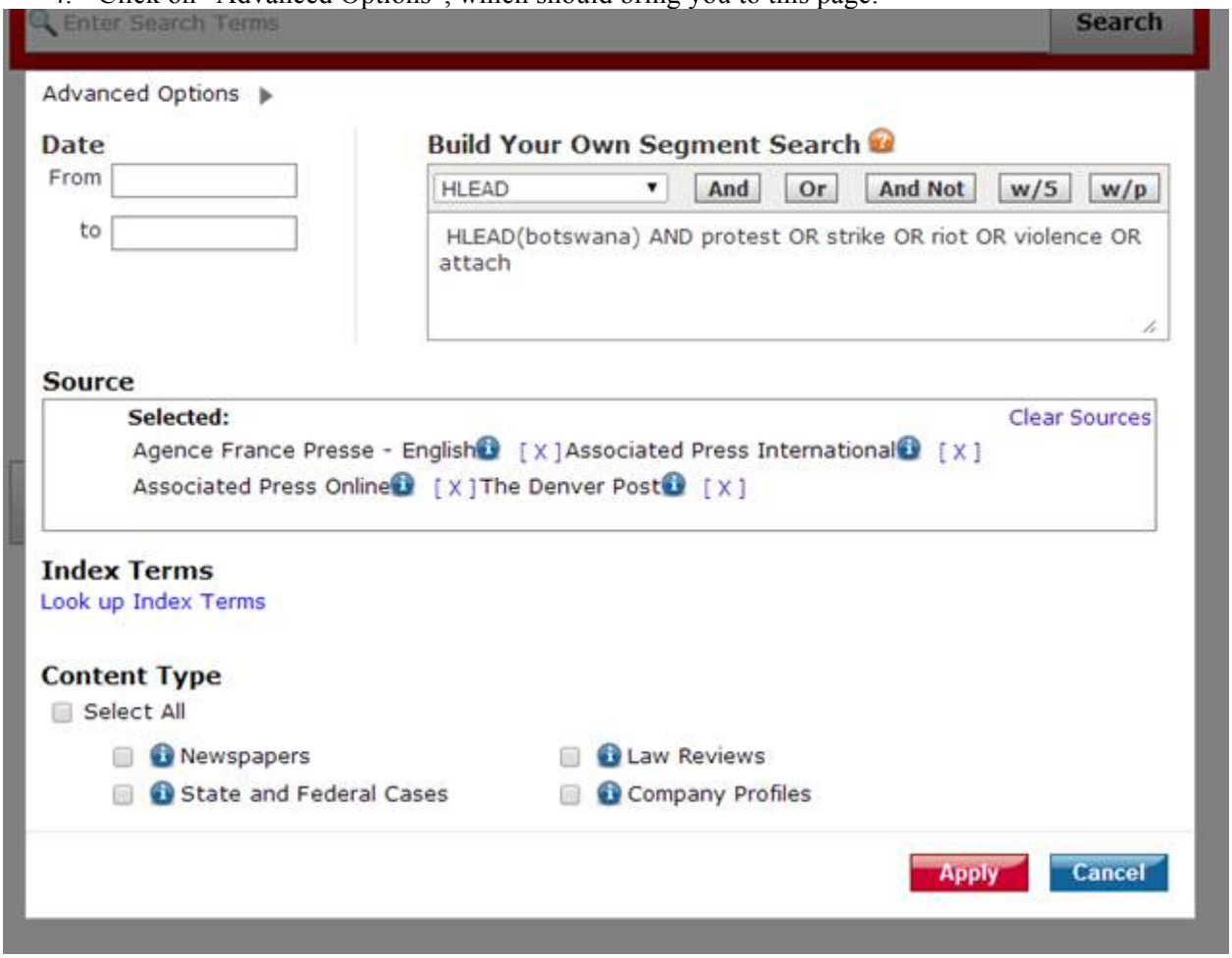
This section describes our search procedure in Lexis-Nexis. The research staff conducted news searches and coded variables data using the following steps. Users wanting more information about particular events in the database can replicate the search methodology to find the original story.

Search Procedure

1. Click on the "Browse" link at the top of the page.



2. On the next page, click "News", then "Newswires and Press Releases".
3. Select "Agence France Presse – English", "The Associated Press", "Associated Press Online," and "Associated Press International". Then click "OK – Continue".
4. Click on "Advanced Options", which should bring you to this page:



5. In the *Build Your Own Segment Search* dialogue box, enter your search terms: HLEAD(country name) AND (protest OR strike OR riot OR violence OR attack).
6. Select appropriate data ranges.

Sorting procedure

1. Begin with the oldest listed story and proceed chronologically. For each story, determine if this information is relevant to the data project. If not, discard the story.
2. During periods of civil conflict, defined by the start and end dates in the *Uppsala Armed Conflict Database* (see below for further details), code violent events associated with the civil conflicts but make sure to code these cases as ACD questionable = 1.[1]
3. Sometimes, many articles will cover a single news story or event. Group these articles together as a single event. Avoid double or triple counting a single event if there are multiple articles chronicling a story. However, use the most recently cited and/or most common estimates for disputed information, such as the number of deaths.
4. Sometimes, a single article will cover multiple events. Determine if the main actor(s), target(s), and issue(s) are different, and if so, code these as distinct events.
5. Code a new event if the actor(s), target(s), and issue(s) are different or if there is substantial time that elapses between one event and another. If the actors, targets, and issues are the same, but there is a brief lull in activity, code it as the same event, so long as there is significant "momentum" linking the time periods.

Coding procedure

This section describes the particular variables in the database.

Identification scheme
The "country code" column provides the Correlates of War (http://www.correlatesofwar.org/) numeric designation of the country. The "id" column lists the event number. These fields are combined to generate a single, unique "event id" code.

Start Date
Lists the day, month, and year in DD/MM/YYY format for when the event begins. If the exact day cannot be identified, this is the best approximation of the start date. The start year (*styr*), start month (*stmo*), start day (*stday*) are listed in separate columns.

End Date
Lists the day, month, and year in DD/MM/YYYY format for when the event ends. If the exact day cannot be identified, this is the best approximation of the end date. The end year (*eyr*), end month (*emo*), end day (*eday*) are listed in separate columns.

Duration
The number of days the event lasted.

Etype (categorical)

[1] Prior to the 2014 update, conflict-related events such as battles and clashes between rebels and armed forces/militias were excluded from coding; for supplemental instructions for merging SCAD with UCDP event data for event-level analysis, see the appendix.

Indicates the type of event according to the following coding scheme:

1 = **Organized Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct "other" group or government authorities. In this event, clear leadership or organization(s) can be identified.

2 = **Spontaneous Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct "other" group or government authorities. In this event, clear leadership or organization cannot be identified.

3 = **Organized Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct "other" group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) can be identified.

4 = **Spontaneous Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct "other" group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) cannot be identified.

5 = **General Strike.** Members of an organization or union engage in a total abandonment of workplaces and public facilities.

6 = **Limited Strike.** Members of an organization or union engage in the abandonment of workplaces in limited sectors or industries.

7 = **Pro-Government Violence (Repression)**: Distinct violent event waged primarily by government authorities, or by groups acting in explicit support of government authority, targeting individual, or "collective individual," members of an alleged opposition group or movement. Note that this event is initiated by the government or pro-government actors. See code for repression, below.

8 = **Anti-Government Violence**: Distinct violent event waged primarily by a non-state group against government authorities or symbols of government authorities (e.g., transportation or other infrastructures). As distinguished from riots, the anti-government actor must have a semi-permanent or permanent militant wing or organization.

9 = **Extra-government Violence**: Distinct violent event waged primarily by a non-state group targeting individual, or "collective individual," members of an alleged oppositional group or movement. As distinguished from riots, at least one actor must have a semi-permanent or permanent militant wing or organization. Government authorities are not listed as actors or targets.

10 = **Intra-government Violence:** Distinct violent event between two armed factions associated with different elements within the government. These include violence between two legally constituted armed units (e.g. clashes between police and military) or between unofficial militias associated with particular governmental leaders. This code includes events such as military coups.

Escalation (categorical)
Did the nature of the event change during its duration?

0 = **No Escalation.**

1 = **Organized Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct "other" group or government authorities. In this event, clear leadership or organization(s) can be identified.

2 = **Spontaneous Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct "other" group or government authorities. In this event, clear leadership or organization cannot be identified.

3 = **Organized Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct "other" group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) can be identified.

4 = **Spontaneous Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct "other" group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) cannot be identified.

5 = **General Strike.** Members of an organization or union engage in a total abandonment of workplaces and public facilities.

6 = **Limited Strike.** Members of an organization or union engage in the abandonment of workplaces in limited sectors or industries.

7 = **Pro-Government Violence (Repression)**: Distinct violent event waged primarily by government authorities, or by groups acting in explicit support of government authority, targeting individual, or "collective individual," members of an alleged opposition group or movement. Note that this event is initiated by the government or pro-government actors. See code for repression, below.

8 = **Anti-Government Violence**: Distinct violent event waged primarily by a non-state group against government authorities or symbols of government authorities (e.g., transportation or other infrastructures). As distinguished from riots, the anti-government actor must have a semi-permanent or permanent militant wing or organization.

9 = **Extra-government Violence**: Distinct violent event waged primarily by a non-state group targeting individual, or "collective individual," members of an alleged oppositional group or movement. As distinguished from riots, at least one actor must have a semi-permanent or permanent militant wing or organization. Government authorities are not listed as actors or targets.

10 = **Intra-government Violence:** Distinct violent event between two armed factions associated with different elements within the government. These include violence between two legally constituted armed units (e.g. clashes between police and military) or between unofficial militias associated with particular governmental leaders. This code includes events such as military coups.

Actor 1 (text)
Records the social, political, or identity group (i.e. actor) directly involved in the event.

Actor 2 (text)
Records the social, political, or identity group (i.e. actor) directly involved in the event.

Actor 3 (text)
Records the social, political, or identity group (i.e. actor) directly involved in the event.

Target 1 (text)
Records the social, political, or identity group directly targeted by the event.

Target 2 (text)
Records the social, political, or identity group directly targeted by the event.

Cgovtarget (dichotomous)
Was the central government the target of the event?
Yes = 1, No=0

Rgovtarget (dichotomous)
Was a regional, provincial or local government the target of the event?
Yes = 1, No = 0

Npart (categorical)
Total number of participants in the event.
1 = less than 10
2 = 10 to 100

3 = 101 to 1,000
4 = 1,001 to 10,000
5 = 10,001 to 100,000
6 = 100,001 to 1,000,000
7 = over 1,000,000
-99 = unknown

Ndeath (cardinal numbers)
Records the best estimate of the number of persons killed in the event. If multiple estimates are given, use the mean number of reported deaths or the most recent figures. If the exact number is not given, use the following codes:
-99 = unknown
-88 = unknown but probably small (less than 10)
-77 = unknown but probably large (10 or more)

Repress (categorical)
Did the government use repression or violence against participants in the event?
0 = no repression used
1 = non-lethal repression used (e.g. tear gas, arrests, etc)
2 = lethal repression used (there must be deaths reported)

Elocal (text)
Identifies the name of the locality where the event occurred.
[See below for more information on geocoding SCAD]

Locnum (categorical)
Coding of the event locality
1 = Capital city
2 = Other major urban area (population greater than 100,000)
3 = Rural (including small towns, villages with population less than 100,000)
4 = Multiple urban areas
5 = Multiple rural areas
6 = Province/region listed, exact location unknown
7 = Nationwide. Effects several cities and rural areas
-99 = location unknown

Issue 1 (categorical)
What was the first issue that was mentioned as the source of the tension/disorder?
1 = elections
2 = economy, jobs
3 = food, water, subsistence
4 = environmental degradation
5 = ethnic discrimination, ethnic issues
6 = religious discrimination, religious issues
7 = education
8 = foreign affairs/relations
9 = domestic war, violence, terrorism
10 = human rights, democracy
11 = pro-government
12 = economic resources/assets
13 = other

14 = unknown, not-specified

Issue 2 (categorical)
What was the second issue, if any, that was mentioned as the source of the tension/disorder

1 = elections
2 = economy, jobs
3 = food, water, subsistence
4 = environmental degradation
5 = ethnic discrimination, ethnic issues
6 = religious discrimination, religious issues
7 = education
8 = foreign affairs/relations
9 = domestic war, violence, terrorism
10 = human rights, democracy
11 = pro-government
12 = economic resources/assets
13 = other
14 = unknown, not-specified

Issue 3 (categorical)
What was the third issue, if any, that was mentioned as the source of the tension/disorder

1 = elections
2 = economy, jobs
3 = food, water, subsistence
4 = environmental degradation
5 = ethnic discrimination, ethnic issues
6 = religious discrimination, religious issues
7 = education
8 = foreign affairs/relations
9 = domestic war, violence, terrorism
10 = human rights, democracy
11 = pro-government
12 = economic resources/assets
13 = other
14 = unknown, not-specified

Issuenote (text)
Include a very brief description of the event.

Nsource (categorical)
Which news agency provided information on the event?
AFP, AP, Both

Notes (text)
This provides additional information about the event. It lists any irregularities in the case, difficulties in coding, discrepancies between sources, etc.

Female_Event (categorical)
This column notes whether women and girls play a significant role in the event in question. This is coded as one if: a. the main actor(s) in the event organize on the basis of female gender identity; b. the target(s) of the event are explicitly female; or c. the motivating issue centers on matters of specific concern to women (e.g.

reproductive rights). Events coded as zero may or may not involve women as participants, as women may play important roles in social organizations and protests otherwise listed. For this column to be coded as one, the main focus of the event, or the identity of the participants, must be explicitly related to women.
1 = yes, 0 = no

LGBTQ_Issue (categorical)
This column notes whether members of the Lesbian, Gay, Bisexual, Transgender, or Queer/Questioning (LGBTQ) community play a significant role in the event in question. This is coded as 1 if: a. the main actor(s) in the event organize on the basis of LGBTQ identity; b. the target(s) of the event are explicitly LGBTQ; or c. the motivating issue centers on matters of specific concern to the LGBTQ community (e.g. same-sex marriage). Events coded as zero may or may not involve LGBTQ individuals as participants, as they may play important roles in social organizations and protests otherwise listed. For this column to be coded as one, the main focus of the event, or the identity of the participants, must be explicitly related LGBTQ persons.
1 = yes, 0 = no

Relationship with the Uppsala University Armed Conflict Database

Some events listed in the database serve as "placeholder rows" for armed conflicts, as listed by the Uppsala University Armed Conflicts Database (ACD). For these events, a negative event id code is given to distinguish them from SCAD events, indicating that our researchers did not compile this information. The start and end dates of ACD conflicts are given, along with brief summary information, using the coding classifications given above. For these events, the etype is given as -9, again to distinguish them from SCAD event types. For most purposes, users will find it acceptable to treat ACD armed conflicts as its own category of events within the SCAD framework. However, in the "notes" field, the ACD conflict id is provided as an aid to researchers who would like to merge the datasets or find additional information.

Until the 2014 update, SCAD researchers excluded events that are clearly part of the civil conflict dynamic, such as when specific rebel actors are named. In some cases, however, it was difficult to determine whether or not the event pertained to the ACD conflict. The column for "ACD questionable" assigned these ambiguous cases a code of 1. Beginning with the 2014 update, SCAD updates became more frequent than UCDP updates, and as such the compatibility of codings could not be preserved. Since 2014, armed conflict events have appeared in some number in SCAD, and all are coded with "ACD questionable" of 1.

Geographic Information – Important!

For the SCAD 3.0 update and new versions, we provide geographic information on event locations in the SCAD dataset. This version of the data is organized differently from previous versions of SCAD. Previously, each event was a unique case or "row" in the dataset. In the 3.0 version, each row corresponds to a particular event location. Therefore, an event which takes place in multiple locations is given the same event ID for each row, but individual locations are listed separately.

For example, event 6150020 is a riot in Algeria which takes place in three cities: Algiers, Constantine and Batna. The *elocal* column (event location) indicates the three cities in aggregate form, "Algiers, Constantine, and Batna." However, note that in SCAD 3.0 event 6150020 is assigned to additional rows, one for each city in which the riot takes place. All three locations have the same event id number, indicating that they are part of a single SCAD event. The *ilocal* column (individual location) separately lists each location. The *sublocal* column provides a numeric count of the individual locality, 1st location, 2nd location, 3rd location, etc.

Users not interested in longitude and latitude coordinates or additional geographic information are advised to drop or filter out all rows where *sublocal* does not equal 1 (i.e. only take the first entry), and ignore information in the *ilocal*, *longitude* and *latitude* columns. Failure to do so will result in over-counting the same event.

> It is important to note that all other corresponding information pertains to the event as a whole and not for that individual location. All other informational variables (e.g. deaths, repression, magnitude) are identically duplicated for each location and may or may not represent what occurred in that particular town or village. For instance, in the Algeria example above (6150020), the event as a whole killed two people; it is not the case that two people died in each of the three cities. Therefore, in calculating statistics such as the number of deaths or the number of participants, users should take care to use unique event ids rather than individual event locations. Doing otherwise would artificially create multiple counts of the same information.

Longitude and Latitude

GeoSCAD estimates, with as much precision possible, the *latitude* and *longitude* coordinates for each event location. We consulted Yahoo! Maps, www.geonames.org, www.fallingrain.com, and www.itouchmap.com/latlong to pinpoint event locations.

For major cities and towns, identifying the X/Y coordinates was relatively straightforward. However, smaller towns, villages, refugee camps, etc, sometimes did not appear in gazetteers and required additional location cues—e.g. "10km south of Bujumbura"—and the given coordinates are rough approximations. At other times, the news report did not mention a specific location, but rather, the district or region was given (these correspond to *locnum* = 6 in the SCAD file). In such cases, we placed the point at the approximate center of the region. Point locations that are estimated rather than precise are flagged in the *location_precision* column. Care should be taken when performing spatial analyses of such cases.

For several events, the location is listed as being "nationwide" or "unknown" (locnum 7, -99). These events are assigned at or near the centroid, or geographic center point, of the country. For the sake of clarity, for location "unknown" events, the point was offset from "nationwide" by subtracting 0.5 from the longitude and latitude, with the exception of very small countries where smaller increments were used.

When further clarifications are needed, additional information about the location and geocoding procure are noted in the *geo_comments* column.

Finally, a very small share of events could not be georeferenced because the location could not be found or the description was too vague. Users may choose to drop such cases or assign them to points as they see fit.

ADDITIONAL CODING PROTOCOLS – SCAD 2.0
September 22, 2011

This note raises issues and concerns with the coding protocols listed in the Political Disorder in Africa Codebook. In particular, it is important to further clarify event types and logical issues in the coding procedures. These are not hard-and-fast rules, but should serve as a guide for coders.

1. **Degree of organization.** Sometimes, it is unclear if the event is an organized or spontaneous protest or riot. As a general rule, if the main actor (Actor1) field contains a proper noun this is considered to be an organized event. For instance, if actor 1 is listed as "students", this is not a proper noun and is probably not an organized actor (though not always so). If, however, actor 1 is the "Muslim Students Association", this is an organized actor since a proper noun can be given. When applicable, provide the organizational name in the actor and target fields rather than general descriptors.

2. **Correspondence between event types, actors/targets and other variables.** Many of the activities require that the government be a target, or indicate that the government cannot be a target. Protests, riots, and strikes may target governmental or non-governmental entities. However, the following event types imply something about the actor and/or target:

Etype 7 (pro-government violence): the target of the action cannot be the government; (cgovtarget and rgovtarget must be coded 0).

Etype 8 (anti-government violence): the target of the action must be the government (cgovtarget and/or rgovtarget must be coded 1).

Etype 9 (extra-government violence): the target of the action cannot be the government (cgovtarget andrgovtarget must be coded 0); moreover, no government entity can be listed in the actor fields.

Etype 10 (intra-government violence): the target and the actor must be government entities (cgovtarget or rgovtarget must be coded 1).

Also, if actor 1 is listed as a government entity (e.g. police, soldiers), repression must = 0. The state cannot repress itself. Note that **etype7** can only get a repression code > 0 if this actor is not an official state entity. For instance, supporters of the ruling party may engage in **etype7** and be repressed by the police, as long as actor1 (and actor2, etc) are not official state agents. If the state engages in repression and is attacked by another state entity, this becomes intra-government violence.

3. **Same versus different event.** Sometimes it is difficult to ascertain whether particular news reports are covering different events or the same event. As a general rule, an incident is a single event if the following conditions apply: a. the actors and targets are the same, b. the issue is the same, c. the action takes place in a short time interval.

Note that if an event occurs in different locations, these are not different incidents as long as the actors and issues are the same; there are codes for multiple event locations. What constitutes a "short" time interval is a judgment call and may depend on the case at hand or type of event. If two events are separated by several weeks or months of inactivity, you may choose to code them as separate events. If, however, the issues and actors are the same, but there was a brief lull of a few days, you may code the incident as a single event. There is no hard rule about how brief the interval is, but look for events that have a given "momentum" of action.

4. Estimates of participants, estimates of deaths. Sometimes news reports will differ on the number of participants and the number of deaths. For the number of participants, use the -99 code judiciously and only if there is no truly reported number of participants. Otherwise, use the most widely reported and/or most recent reports.

For the number of deaths, sometimes the estimated death counts will vary widely. Use the -99 code if the number of deaths is not given at all (e.g. "In rioting yesterday, there were reports of deaths…"). Use the -77 and -88 codes if there is some indication of the number of deaths, but there is no precision in the reporting (e.g. "Several dozen were killed in attacks yesterday…").

At times there will be "educated guesses" about the number of killings and conflicting reports, but the true number cannot be confirmed. For instance, in the Kenyan electoral violence (2007-2008) some sources reported around 1,000 dead, others reported around 1,500 dead. In these cases, it is difficult to know with certainty how many actually died, but the death estimates give specific numbers or ranges. Rather than indeterminate (-77, -88) codes, use the most widely cited and most current deaths figures.

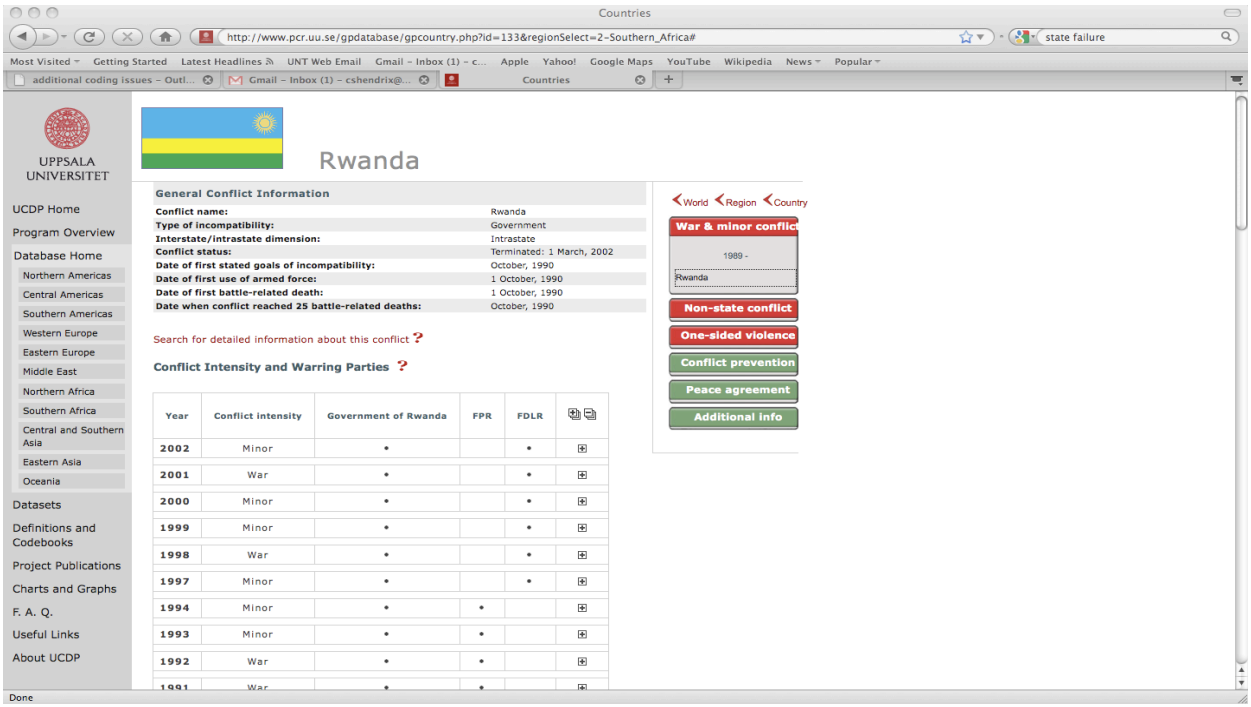
5. Clarification of coding rule for civil war/conflict cases. For SCAD releases earlier than 3.0, during periods of civil conflict, defined by the start and end dates in the Uppsala Conflict Database, we did not code violent events associated with the civil conflicts. If Actor1, Actor2, or Actor3 is the government, a quasi-governmental organization (i.e., paramilitary organization, like the Hutu militias), or the rebel organization, and these acts are directly linked to the civil conflict, do not code the event. If the Actor is the government or a quasi-governmental actor, we will not code etype 7 *unless the targets and issues are completely irrelevant to the civil conflict.* Also, do not code violent attacks by the rebels against the government or any other target (etype8 and etype9), if this is considered to be part of the civil war dynamic.

Examples

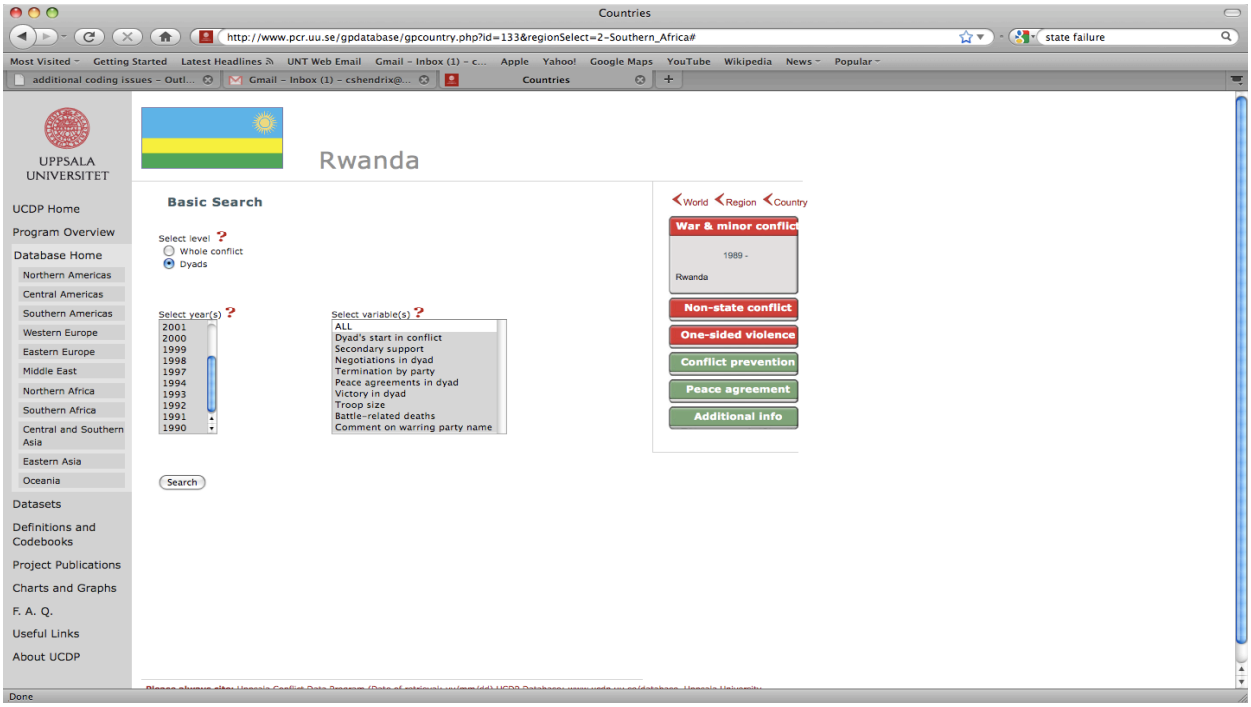
- Violence against civilians perpetrated by government actors, quasi-government actors, and rebel organizations, such as the Rwandan genocide; these events would not be coded, so long as the civil war is ongoing.
- Pro-government violence (repression) against street vendors in the capital city, while the government is engaged in a civil conflict in the countryside; this event would be coded as government repression, even though a conflict is ongoing.

As of SCAD 3.0, we are not excluding these events from coding. Events that may be related to armed conflicts as cataloged in the Armed Conflict Database are marked acd_questionable = 1.

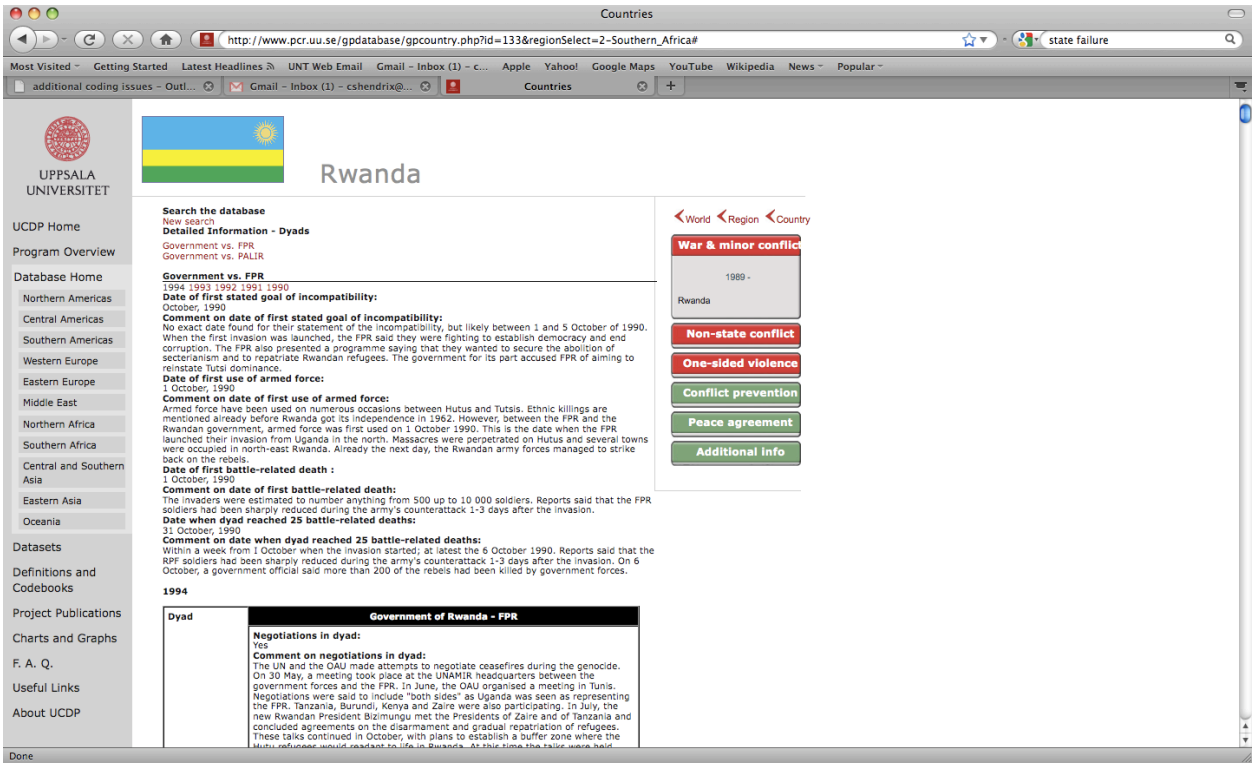
6. Finding precise UCDP start and end dates for each actor. From the War & Minor Conflict page, click on "Search for detailed information about the conflict."



At the Basic Search page, select “Dyads” in the toggle box, then select “ALL” years and “ALL” information



Once you’ve arrived at the detailed information page, you can find exact start and end dates for each party to the conflict.



Before you post final spreadsheets, it is important that they are "cleaned" to make sure there are no errors, inconsistencies, and typos. Being careful and meticulous now will save us all a lot of time and hassle later on down the road. Take the following steps before posting files.

1. Make sure that the start and end dates are valid. The end date should be after the start date – if not, there is an error. Checking this is relatively easy, just enter a new (temporary) column, formatted as a number, after the end date and subtract the two dates (see excel help files on how to calculate the number of days between dates). This number should not be negative. Also, it is a good idea to check if your start dates are in chronological order (can also do this automatically). For the most part the start date column will proceed in order. It could be that the reporting was retrospective and so the start dates are not in order, but these will be relatively rare and you should be able to identify such cases easily.
2. Check place names. Using the filter option in excel, you should make sure that everything with a **locnum** = 1 is indeed the national capital. Also, you can use the filter option to look through all of the listed place names. Make sure that all of your place names are identically spelled. These place spellings should be consistent (e.g Dar-es-Salaam and Dar-es-Salam are not consistent).
3. Double check **cgov** and **rgov** targets. Look through all of the listed targets and if a government actor is listed, **cgovtarget** and/or **rgovtarget** should be coded as 1.
4. Look for illogical codings. In particular, pro-government violence initiated by a state actor (police or military) cannot be repressed. If the police is repressing and the army steps in (for example) this is intra-government violence. Pro-government violence can only get a repression code if the actor is not an official state entity (e.g. the police can repress a pro-government mob. However, the police cannot repress itself). Also, if lethal repression is used (**repression** = 2) then the death count, by definition, cannot be zero.
5. Look for double coding of an event. Remember that if the actor, target, and issue are all the same, and the events roughly occur during the same time period (within a few days), this should be the

same event. An easy way to do this is to look and see if there are two events that occur on the same day. If yes, then double-check to make sure these are separate events and not the same.