Political Instability Task Force Worldwide
Atrocities Event Data Collection Codebook
Version 1.0B1

updated 2007.11.16

This document describes the motivations, definitions, and coding procedures we expect to use to create a global dataset on atrocities, by which we broadly mean the deliberate killing of non-combatant civilians in the context of a wider political conflict. This draft codebook—prepared by Jay Ulfelder as a revision of an earlier document produced in collaboration by Jay Ulfelder and Phil Schrodt in consultation with the project's sponsors—attempts to incorporate comments made by several experts who reviewed the design and results of an experimental data-collection effort. That experimental effort was undertaken in autumn 2003 and led by Philip Schrodt, University of Kansas, under subcontract to Science Applications International Corp. (SAIC). SAIC project management is provided by Irwin Jacobs, Program Manager, and Jay Ulfelder, Technical Manager. Additional text for the codebook has been provided by Dennis Hermrick, Milos Jekic and Taylor Price, University of Kansas.

This research project is funded by the Office of Transnational Issues (OTI) in the Central Intelligence Agency's Directorate of Intelligence (DI). This document does not represent the official view of the US Government, the US Intelligence Community, or the Central Intelligence Agency, but rather the views of its individual authors.

The data described in this codebook is available at:

`http://web.ku.edu/keds/data.dir/atrocities.html`

Project Motivations

This project is intended to advance efforts to understand and anticipate atrocities, which we understand as the deliberate use of lethal violence against non-combatant civilians by actors engaged in a wider political or military conflict. Conflicts producing atrocities may be motivated by a variety of interests—not only the desire to wield state authority, but also the desire to control economic resources, to change the status of a particular ethnic or religious community, or to defeat another state in an international war.[1]

The practical objective of this project is to create a dataset representing a reasonably systematic sample of atrocities occurring worldwide in recent decades in order to: 1) enable the development of statistical models that might be used to identify countries vulnerable to the occurrence of atrocities or, if atrocities are already occurring, to an escalation in their rate or intensity; and 2) create a descriptive record that might be used by researchers with an interest in particular countries or conflicts.

To our knowledge, most existing datasets on atrocities either cast the "macro" event of a larger conflict process as the unit of observation—for example, Harff's (2003) or Fein's (1993) lists of genocides or Rummell's list of democides—or focus on "micro" events within the context of a single

[1] We do not intend to suggest that each atrocity is motivated solely or even partly by the objectives at stake in a wider conflict. We are simply noting that atrocities occur in the context of virtually every form of violent political struggle.

conflict process—for example, Ball, Kobrak, and Spirer's (1999) dataset listing instances of human rights violations in Guatemala, 1960-1996. While many of these datasets are important and useful, they limit researchers to analyzing either the onset of conflict processes involving atrocities—and often more specifically to conflict episodes that rise to the level of genocide—or to analyzing the dynamics of violence in the context of a genocide. They do not allow researchers to draw inferences about the likelihood of any deliberate violence against non-combatants in situations where none has yet occurred, or the dynamics of such violence in situations that may or may not rise to the level of genocide—both of which may of substantial concern to policy-makers, researchers, and other interested parties.

To enable the development of models of the dynamics of violence applicable to countries that may or may not experience atrocities on a massive scale, we aim to record information from press accounts of atrocities occurring in all countries of the world during our period of observation. The data set that results from this global approach will allow us to study the risk of atrocities that may or may not accumulate to the level of genocide or "democide" while also avoiding the problem of selection on the dependent variable that would contaminate any effort to pool observations from existing atrocity event data sets.

Even with massive resources, any such global effort will fail to observe many atrocities occurring in the context of particular sustained episodes of violence or genocide, a task to which strategies focused narrowly on such episodes are better suited. We also recognize that variations across countries and over time in the availability and reliability of journalistic accounts of atrocities will shape the data in ways unrelated to the dynamics we wish to study. We expect to address the problems of incomplete coverage and source bias in two ways: 1) by using a variety of press sources, including local ones (as captured in FBIS), in an effort to maximize coverage and offset the biases of individual publications within the inevitable constraints of our research budget; and 2) by explicitly addressing the issue of bias at the modeling stage, attempting either directly or indirectly to measure it and thus "control" for its effects on the subject of interest.

We are primarily interested in situations where state or non-state actors use violence against non-combatant civilians as a tactic in wider conflicts. We also recognize, however, that the unambiguous intent to kill non-combatants is often difficult to establish, and that such gross violations of human rights sometimes occur in the context of conflicts between or among non-state actors, or in situations where state authority has decayed so completely that the state vs. non-state archetype does not apply.

In our view, the ambiguity of many situations involving attacks on non-combatant civilians necessitates a flexible data collection scheme. In the approach described below, we have attempted to avoid asking coders to base the decision to include or exclude reports on the basis of their own assumptions or inferences about perpetrator motivations and identity. Instead, we have attempted to cast our net as widely as possible, recording information on all events that, according to the description offered by journalists, might be interpreted as representing the deliberate use of lethal violence against civilians as part of a political conflict.

By capturing as many potentially relevant events as possible and explicitly recording their ambiguities, we intend to allow researchers to employ various approaches when analyzing these data, ranging from narrow interpretations of what constitutes an atrocity to broader ones, or from restrictive rules about the character of reporting on these events to looser ones. To our knowledge, such an effort systematically to document atrocities worldwide has not been undertaken previously.

By focusing our research on reports of killings or series of killings, we aim to treat the use of violence against civilians in any political conflict as a variable, open to empirical study. This approach recognizes that such killings may occur before or after a large-scale conflict is conventionally understood to have begun or ended; such killings may also occur in or outside the context of a conflict process considered genocide by other researchers. We believe this strategy will allow us to study not just the occurrence but also the dynamics of mass atrocities in the widest variety of contexts.

Defining and Coding Atrocity Events

For purposes of this project, we define an **atrocity** as *implicitly or explicitly political, direct, and deliberate violent action resulting in the death of noncombatant civilians.*

The victims' status as noncombatant civilians is a critical element of virtually all conceptions of an atrocity. This criterion focuses our research on the use of violence against individuals not intentionally involved in an armed struggle. Following Valentino, Huth, and Balch-Lindsay (2001: 8), we define a *noncombatant civilian* as "any unarmed individual who is not a member of a professional or guerrilla military group and who does not actively participate in hostilities by intending to cause physical harm to enemy personnel or property." Further following those authors, we note that associating with combatants, feeding or sheltering them or participating in non-violent political activities in support of combatants does not transform a civilian into a combatant.

Our conception of noncombatant civilians, however, does not extend to individuals sentenced to death with due process by a state's criminal justice system, nor does it include individuals killed while engaged in acts of violent crime, such as rioting, brawling, or looting. It is also worth noting that our definition does not depend on the racial, ethnic, religious, or political status of the victims—only their status as noncombatants. In this manner, we cast our net more broadly than researchers interested in the narrower issue of genocide.

Of course, claims about the combat status of killing victims are often contested, and journalistic accounts of potential or apparent atrocities sometimes include the claims and counterclaims of interested parties. This project adopts a skeptical stance on claims about the combat status of the victims. Consistent with our broader attempt to cast the net as widely as possible, if some parties claim the victims were combatants or other legitimate military targets but those claims are contested, we include that event in our dataset and use a variable to record the fact that there were competing claims about the victims' status.

We also err on the side of inclusion by assuming that when news sources do not apply a military or paramilitary label to victims of killings, those victims were most likely non-combatant civilians and the event should be included. Again, however, we attempt to capture this ambiguity in our database with a field indicating whether the victims were explicitly described as non-combatant civilians.

Finally, departing somewhat from the Valentino, Huth, and Balch-Lindsay definition, we include as "non-combatants" individuals who may be combatants in a different context, but who at the time they were killed were unarmed and unable to defend themselves. This would include, for example, members of guerrilla groups who had come into refugee camps unarmed in order to get food or medical care, or off-duty police eating at a pizza stand. This does not include situations where the individuals were in a military or para-military situation (for example a guerrilla base or police barracks) which the attackers could reasonably have expected to be defended, even if the individuals

happened to be unarmed (e.g. sleeping) at the time of the time of the attack. If there is doubt about whether the victims qualify as "temporary noncombatants", we code the case but indicate the ambiguity in the description.

The modifiers *direct* and *violent* focus our research on situations where deeds committed by the perpetrators are the proximate cause of noncombatant civilian deaths. These characteristics distinguish atrocities from other actions considered war crimes or crimes against humanity in which individuals or organizations make decisions that put non-combatants at grave risk but do not directly kill them. Such situations include attacks on humanitarian aid missions; the use of civilian facilities, such as hospitals or schools, as "hiding places" in the context of an armed conflict; and efforts to restrict or destroy civilian supplies of food or water.

The references to action that is *deliberate* and *implicitly or explicitly political* focuses our research on violence intentionally perpetrated to political ends, as opposed to accidents, acts of nature, or acts of private or personal violence. This criterion gets to the intentions and agency of the perpetrators—aspects that are notoriously difficult to establish with certainty even in a court of law, let alone from brief journalistic accounts. To address this dilemma, we again intend to err on the side of inclusion by excluding only those events that appear to fit decisively in one of several categories in which we are not interested. [Coders: When in doubt, include the event as an atrocity but note the source of your reservations in the "Comments" field.]

- **Acts of Nature.** Any event in which deaths are not the result of human actions, e.g., natural disasters, falling trees, etc.
- **Accidents.** Any happening that is not intended or expected. [Coders: Note that apparently accidental civilian deaths resulting from military or police action should be included but coded as "collateral damage."]
- **Suicide.** A single victim kills himself or multiple victims kill themselves and no one else is harmed.
- **Private Murders.** A person or persons are killed over a personal matter or in the course of a criminal act with no apparent political intent. The key issue is whether the perpetrator or perpetrators apparently acted on behalf of a state agency or a communal or political group.

In general, something is political when it is intended (realistically or otherwise) to have an effect on some collective group that goes beyond the group engaged in the action. That collective group frequently involves a government, but it can also be an ethnic or religious group or any other distinct group of individuals. Criminal activity, in contrast, only benefits the individuals who are involved in it and is not intended to have a collective impact.

The two obvious "grey areas" in this definition are:

- When criminal groups engage in "Robin Hood" activities where they attack one group for the supposed benefit of another group (stereotypically the rich for the benefit of the poor, though in ethnic conflicts this may become very fuzzy); well-organized criminal groups may also provide indirect distribution through the provision of some social services (distribution of food; support for religious institutions) in populations that support them.

- Collective action by representatives of one social class (typically landowners or factory owners, or rather thugs and death squads hired by them) against another social class (typically workers or peasants), where there is a direct benefit only to the wealthier class.

The ambiguity occurs because in many situations, one gets a mix of collective and individual benefits, so there is no single distinguishing characteristic. So for example poor farmers in one ethnic group might attack even poorer farmers in a different ethnic group—as happened in South Africa near the end of the apartheid era—and such actions have both a collective and individual effect. Any time there is some indication that the motivation for the violence might be interpreted as being political, assume that it is, but indicate any reservations about this in the comments.

We realize that many—perhaps most—atrocities are never directly reported in public news sources. What's more, in many instances where atrocities are reported, journalists summarize a collection of apparently related atrocities in a single story, often without reference to the discrete killing events that constitute that sequence. Such summary descriptions are not directly comparable to reports of discrete killing events—they represent different units of observation—but they also contain important information about the use of violence against civilians, information that is often unavailable in any other fashion from the press.

To deal with this dilemma, we decided to record in our database information not simply on atrocities as such, but rather on two related types of atrocity events, which we label *incidents* and *campaigns*. In instances where we can only locate information on campaigns and not the discrete events comprising them, we intend to record that information in a manner that distinguishes it from the information on discrete killing events. In instances where reports provide information describing both discrete events and campaigns or linking one type of event to the other, we intend to record these connections with variables in our database, giving researchers greater flexibility in deciding how to aggregate or disaggregate the information for analysis.

In theory, all killing campaigns are comprised of a series of discrete killing events. (It should be noted, however, that not all killing events are parts of a campaign.) We also realize, however, that not all of those events will be reported; in fact, in many instances, it's likely that none of them will. Ideally, we would be able to record the details of all mass-killings events comprising all mass-killing campaigns in all violent political conflicts during our period of observation. Realistically, we recognize that information is often unavailable on discrete mass-killings events; in many situations, this phenomenon is only reported at the level of the campaign, whether because more detailed information is unavailable or because reporters choose to describe violence in these terms for other reasons.

Data Limitations

This codebook was originally developed in 2004 and outlined a scheme that we thought we could implement based on our experiments in 2003. In the subsequent three years of actual coding, some additional limitations have emerged that make the actual data somewhat more circumscribed that what was outlined here. The key limitations:

1. We have only coded incidents involving five or more non-combatant deaths. We attempted at one point to lower this threshold to "one" and the data collection demands proved completely overwhelming, as this involved assessing every murder and ambiguous accidental death reported anywhere in the world in the international media. "Five" has no underlying theoretical justification; it merely provides a threshold above which we can confidently code all of the reported events given our available resources.

2. To date, we have coded only four international sources: Agence France Presse, Associated Press, New York Times, and Reuters. Due to limitations in our available labor, we have not attempted to code either local sources or various possible NGO, IGO and other web-based sources.
3. The "campaign" and "incidents" scheme hasn't really worked out, nor have the linked records.
4. Due to legal limitations on the Central Intelligence Agency, which is funding this collection, we cannot code incidents involving the United States as either target or perpetrator. These limitations date to the 1970s and were intended to limit the ability of the CIA to monitor (or analyze) the behavior of U.S. citizens. While open-source data collections such as this were probably not the intent of the original legislation, the CIA has interpreted those restrictions cautiously and therefore such episodes are not included in this data set.

Atrocity Event Data Records

The data are entered into an Excel spreadsheet. The following types of entries are used:

Text	Any text can be entered
Date	Date in the form dd-mm-yyyy, e.g. 05-12-1997 for December 5, 1997
Number	Numeric value
Category	Value is chosen from a fixed of choices
Boolean	True/False

Event Type and Reporting

The first variable here is used to record the type of event reported in an article, as discussed in the preceding section, "Defining and Coding Atrocities." If a single article provides information on multiple events, create a separate record for each event.

- **Incident.** An atrocity perpetrated by members of a single organization or communal group, or by members of multiple organizations or groups reportedly acting in concert, in a single locality within a 24-hour period.

  By locality, we mean a single village, town, or city. We recognize that villages, towns, and cities are not entirely comparable levels of analysis; for example, the distance between two neighborhoods in a single city may be greater than the distance between two rural villages. Even so, we believe the act of traveling from one village or town to another represents a kind of discontinuity that often requires perpetrators to commit anew to their behavior and thus represents a break point between discrete events.

  Just as traveling from one locality to another represents an important discontinuity, so does the decision to renew violence each day. Following this reasoning, we treat violence in the same locality committed by the same perpetrators but on different days—even consecutive ones—as distinct incidents. If reports do not distinctly describe events on those consecutive days, the event should be coded as a campaign rather than an incident.

  In situations where members of opposing organizations or groups engage in tit-for-tat or

reprisal killings in a single locality within a 24-hour period, those two (or more) sets of killings are treated as distinct events.

- **Campaign.** A set of reportedly related atrocities perpetrated by members of a single organization or group, or by members of multiple organizations or groups reportedly acting in concert, over a distinct period of time within a single country.

  This is, in effect, a residual category intended to capture information about atrocities when the details required to identify discrete incidents are not available in the accounts we read. Reflecting this motivation, our definition of a killing campaign is less specific than the definition of a killing event. The requirements related to the identity of victims (non-combatant civilians) and perpetrators (identified as members of a political organization or communal group) are identical, but the temporal and spatial boundaries are much looser.

  Coders should note, however, that even under the looser definition of a campaign, the event—not the report—is the unit of observation. Where a single report provides sufficient information to distinguish multiple incidents, each of those incidents should be recorded separately. If the same report also describes those incidents as part of a wider campaign, the campaign should be recorded as well, and the linkage between the incidents and campaign should be noted. In similar fashion, when a report provides information sufficient to distinguish multiple campaigns—either because the perpetrators differ or because the report mentions multiple start and end times—these campaigns should be recorded separately.

  In contrast to incidents, campaigns may span multiple localities. Coders should record all geographic information provided in the appropriate fields, and they should use the comment field to relate reported death totals to those locations to whatever extent possible.

  Campaigns have a "campaign identifier" that consists of the 3-letter UN code of the country where the campaign occurs, a hyphen and a 6-number YYYYMM date corresponding to when the campaign began. For example:

  IND-199212
  RWA-199404

  If the campaign occurs in more than one country, use the identifier where most of the deaths occurred; if multiple distinct campaigns begin in the same country and month, distinguish them with the addition of a capital letter suffix "A", "B", etc at the end of the identifier.

The second variable here is used to indicate the nature of the reporting on the event. If there are multiple reports describing a single event, use later reports to "correct" earlier ones. For example, if an initial article refers to "allegations" that soldiers killed villagers but a later article reports that same incident as fact, code the event as a fact. By the same token, if an initial article reports an event as fact but subsequent articles express doubts about whether or not an incident occurred, code the event as a rumor/allegation.

- **Eyewitness Account, Not Contested.** A journalist claims to have witnessed the event in question and the occurrence of the event is not contested.
- **Secondary Account, Not Contested.** Press reports offer a second-hand account of an event that is represented as a fact and the occurrence of the event is not contested. Press accounts of reports by government agencies, international organizations, and non-governmental

organizations (including human rights groups) describing atrocities should be coded in this category.

- **Rumor/Allegation.** Journalistic accounts report an event as a rumor or an unverified allegation. Use this only if all sources indicate that the incident was a rumor: if it is reported as true in a local source but as rumor in an international source, code the source as the local report, and the event reporting as "Contested".
- **Contested.** The occurrence of an event is contested. Note that this does not refer to arguments over the characteristics of an event, but rather to whether or not the event occurred at all.

Field Names:
Event Type [select from categories]
Campaign Identifier [use only if the event type is "Campaign"]
Event Reporting [select from categories]

Event Date

Record the reported start and end dates of each event as accurately as possible. In cases where the event comes to light later—for example, because a mass grave is discovered—record the reported date of the occurrence of the atrocity as the time of the event, not the date of the related discovery (though it may be useful for other purposes to keep a record of the timing of the discovery as well).

The coding procedures differ according to the event type. For atrocity incidents, record the event date in the Start Date fields and use 9s (i.e., 99, 99, 9999) in the End Date fields. For atrocity campaigns, record information in both sets of fields. For both event types, leave fields blank if information is not available (e.g., if report describes a series of killings in January 1998, the coder would leave the Start Day blank, record 01 in Start Month; record 1998 in Start Year; leave the End Day blank; record 01 in the End Month; and record 1998 in the End Year).

Skip fields when information on day or month is not known.

Field Names:
Start Day [dd]
Start Month [mm]
Start Year [yyyy]
End Day [dd]
End Month [mm]
End Year [yyyy]

Event Location (Country, Region, District, Locality)

Record the name of the country in which the event occurred using the three-letter UN country code and record the name of the region, district, or locality in associated text fields. If geographical information is missing, leave the field blank. If multiple named locations are identified, create multiple records with the same Link field. Distinguish between missing information on a smaller territorial unit and cases where the event spanned a larger unit. For example, leave the "Locality" field blank in cases where no city, town, or village is reported, but use multiple linked records where the event spanned several named localities in a given region.

Field Names:
Country [select from UN codes]
Region [text]
District [text]
Locality [text]

Event Location (Latitude, Longitude, and Relational Location)

The geographical location fields provide geographical coordinates (in the form of degree: minute: second, or DMS) for the locality in which the event occurred. It is not always possible to find enough information to enter data in these fields. This is due to two issues. First, some parts of the world, particularly in conflict-prone areas of Africa and the Middle East, lack readily available geo-spatial data. Second, news sources do not always provide a specific locality name. Barring these issues, however, an extraordinary number of localities are recorded, along with their geographical locations, in various online databases.

To find a geographical location, first use www.geonames.org. Most locations are listed on this site. If a particular place cannot be located in www.geonames.org, www.maplandia.com will infrequently have listings not covered by www.geonames.org. Wikipedia is also a potentially useful second source, particularly for finding the coordinates of neighborhoods or districts within larger cities. Longitude/latitude measurements, if available, are listed in the upper right-hand corner or at the very bottom of a location's Wikipedia page.

In order to achieve the most accurate longitude/latitude, always try to find the place listed in the locality field first. Sometimes, however, it is impossible to locate a particular locality, especially if it is a small village, a forest, or a road between two cities. If this occurs, move to the district field and try to capture that location. If that cannot be found, and a region (a province or department, for example) is listed, this location should be captured.

Search Results

When using www.geonames.org, often more than one place with the same name (in the same state) will appear. For example, these are the first three entries that come up on a search for "Baghdad":

```
Baghdad             Iraq, Baghdād, capital of a political entity
                    N 33° 20' 19''       E 44° 23' 38''
Muhāfazat Baghdād   Iraq, Baghdād, first-order administrative division
                    N 33° 8' 0''         E 44° 15' 0''
Qaḍā' Baghdād       Iraq, Baghdād, second-order administrative division
                    N 33° 18' 0''        E 44° 36' 0''
```

If you were trying to find the entry for Baghdad, the capital city, which would likely be listed under "locality," you would enter in the coordinates listed in the first entry. If you were trying to find the entry for Baghdad province, you would use the coordinates listed on the second line, "first-order administrative division," a term that signals that the name is a province, region, or other major administrative division. Baghdad district is the third entry. This is the district in which the city of Baghdad is located.

Neighborhoods

In some regions, particularly in Iraq, the news report will name the neighborhood or district in the city in which the event occurred. If this occurs, the neighborhood should be placed in the Locality field, and the city should be placed in the District field. If, however, the neighborhood's coordinates are not listed in any of the online resources, the coordinates for the city should be captured instead.

Multiple Locations with the Same Name
Sometimes, an event will take place in a village or city that shares its name with many other places within the same state; if this is the case, code the coordinates of the location that has the largest population.

If the news report lacks sufficient information about the province or region in which the event occurred, it may be impossible to select with certainty the correct geographical coordinates and these fields should be left blank.

Transcription issues
In some instances, various news sources will transcribe place names differently. For example, the major city of Baqubah in Iraq's Diyala province is often spelled "Baquoba" by American news sources. A search for the American spelling of this city on www.geonames.org will yield no listings. There are several tricks that can be used to find listings that may be missed because of transcription differences. First, www.geonames.org will often provide links to alternate place names spelled similarly if it cannot find any exact search results. Try clicking on these and see if they are in the same province or district as the location that is being searched for. Second, try searching on either www.maplandia.com or Wikipedia for the place name. Wikipedia in particular appears to find the desired places with differently transcribed names more often than www.geonames.org. These strategies should only be used, however, when it is clear that the issue is a difference in transcription, which will most often be the case with large cities like Baquoba. If it is possible that there may be two villages with similarly-spelled names within close proximity to each other, the geographical location fields should be left blank.

Latitude

Latitude coordinates on www.geonames.org will be listed as follows: (N or S) XX° XX' XX". On www.maplandia.com and Wikipedia, the direction (N or S) follows the degrees, minutes, and seconds. Wikipedia also sometimes lists coordinates in decimal notation. If this is the case, click on the coordinates and Wikipedia will go to a page which lists them in DMS notation. The degrees measurement should be recorded in the Degrees field, minutes in the Minutes field, and seconds in the Seconds field. The direction field will have either North or South, spelled out, as its entry.

Field Names:
Degrees [number]
Minutes [number]
Seconds [number]
Direction [text: North or South]

Longitude

Longitude coordinates on www.geonames.org will be listed as follows: (E or W) XX° XX' XX". On www.maplandia.com and Wikipedia, the direction (E or W) follows the degrees, minutes, and seconds. Wikipedia also sometimes lists coordinates in decimal notation. If this is the case, click on the coordinates and Wikipedia will go to a page which lists them in DMS notation. The degrees measurement should be recorded in the Degrees field, minutes in the Minutes field, and seconds in the Seconds field. The direction field will have either East or West, spelled out, as its entry.

Field Names:
Degrees [number]
Minutes [number]
Seconds [number]
Direction [text: East or West]

Relational Location

Often, in the case of very small villages, news reports will give the locations of the villages in relation to a larger and better known city. If this occurs, the Relational Location fields should be utilized. The Distance category is a number and should be entered in kilometers. The Direction category is also a number that should be expressed as a point on a circle:

North = 0°
Northeast = 45°
East = 90°
Southeast = 135°
South = 180°
Southwest = 225°
West = 270°
Northwest = 315°

The longitude and latitude fields should have the coordinates of the city from which the relational location is given. Therefore, if a location is listed as "40 km northeast of Baghdad," the latitude field will read: 33°20'19" North (Baghdad's latitude); the longitude field will read: 44°23'38" East (Baghdad's longitude); the distance field will read: 40 (for 40 km); and the direction field will read 45 (for 45°, the angle measurement for "northeast.")

If a relational direction is given in an article, be sure to note it in the description field, even if the actual locality has an easily-obtainable geographical location. Also note it in any free field within the Event Location section.

Field Names:
Distance (km) [number]
Direction [number]

Perpetrators

These variables record the reported identity of the event's perpetrators and their reported relationship to the state. Code all of the groups that were explicitly reported as responsible; for example, if both the police and paramilitary forces were reported as responsible, enter codes in both of these fields. In all instances, any group names and keywords indicating something about the perpetrator's identity (e.g., "ethnic militia" or "partisan thugs") should be recorded in the "Description" field.

The categories for state involvement are the following:

- **State Perpetrator.** Reports explicitly identify the perpetrators as employees of a state agency—i.e., members of the armed forces, the police, other official security forces, or any other government agency—for the state in which the event occurred.
- **Non-State, Internal, State Sanction.** Attackers are identified as members of a non-state organization or group with purported state support or state sanction.
- **Non-State, Internal, No State Sanction.** Attackers are identified as members of a non-state organization or group, and there is no allegation of state support or sanction for their actions.
- **Transnational.** Attackers are identified as members of a military or paramilitary group from across an international border.
- **Multiple Perpetrators (State).**
- **Multiple Perpetrators (State And Non-State).**
- **Multiple Perpetrators (Non-State).**
- **Unknown/Unclear/Other.** The report does not characterize the perpetrators in terms specific enough to identify them as state or non-state actors. Indicate any available information in the "Comments" and "Description" text fields.

The perpetrator identity variables are the following:

- **Military.** Attackers are identified as uniformed members of an official state military force. Keywords include "armed forces," "soldiers," "security forces."
- **Police.** Attackers are identified as law enforcement officers. Keywords include "police."
- **State Other.** Attackers are identified as state agents, but there is not sufficient information to characterize them as either military or police, or they are identified as employees of a different state agency.
- **Non-State Ideological.** Attackers are not affiliated with the state and are identified in terms of their membership in an ideological organization or their political or ideological views. Keywords include "leftists," "fascists," "radicals," as well as organization names.
- **Non-State Ethnic.** Attackers are identified in terms of their ethnicity, race, or nationality.
- **Non-State Religious.** Attackers are identified in terms of their religion.
- **Non-State Criminal.** Attackers are identified as members or associates of a criminal syndicate, drug cartel, mafia, etc. Criminal activity includes production and trade of illegal drugs, arms smuggling, kidnapping and extortion, and prostitution when this has a coercive element. If the group is clearly involved in both criminal and legal activity (for example a warlord who controls both diamonds and illegal drugs), use both the private and criminal categories.
- **Non-State Private.** Attackers are identified as part of a private army or security service reportedly working on behalf of a warlord, landowner, private corporation, or association of any such individuals or interests (e.g., a cartel of landowners).
- **Unknown/Unclear/Other.**

Field Names: enter text identifying the group where this is relevant
Perp State Role
Perp State Military
Perp State Police
Perp State Other
Perp Non-State Ideological
Perp Non-State Ethnic
Perp Non-State Religious
Perp Non-State Criminal
Perp Non-State Private
Perp Unknown/Unclear/Other

Victims

The first two fields under this heading are used to indicate whether or not reporting explicitly identifies some or all of the victims of an event as noncombatants, and whether or not that status was contested.

The categories for the first status variable are as follows:

- **Noncombatant Status Asserted.** At least one report of a given event identifies some or all fatalities as noncombatant civilians.
- **Noncombatant Status Not Asserted.** No reports explicitly identify victims as noncombatants, but neither do they identify those victims as combatants. (If any reports identify the victims as combatants and no reports identify them as noncombatants, the event should not be coded.) In the latter situation, victims are often labeled in some fashion that implies they were mostly or exclusively noncombatants (e.g., "villagers," "passengers," "men, women, and children").

The categories for the second status variable are as follows:

- **Noncombatant Status Contested.** At least one report of a given event contains an assertion that the victims were combatants.
- **Noncombatant Status Not Contested.** No reports of a given event contain an assertion that the victims were combatants.

The remaining variables and associated fields are used to record information about the identity of the victims of an atrocity or atrocities. Because identity categories are often overlapping, we have chosen to record this information through a series of dummy variables rather than a single categorical variable. Those dummy variables center on the following categories:

- **Political.** Victims are identified as members a particular political party, political organization, social movement organization, issue-oriented non-governmental organization, or in more general terms as sharing certain political beliefs or sympathizing with particular political or paramilitary groups.
- **Ethnic/National/Citizenship.** Victims are identified as members of a particular ethnic, racial, or national group, a particular tribe or clan, or citizens of particular state.
- **Religious.** Victims are identified as members of a particular religion, sect, or spiritual movement.
- **Socio-Economic.** Victims are identified as members of a particular socio-economic class or occupational group (e.g., peasants, landowners, squatters, workers).

- **Unarmed Combatants.** Victims are identified as individuals who under other circumstances would have been combatants, but were unarmed and unable to defend themselves at the time they were killed (e.g. unarmed members of guerilla groups; off-duty police or security personnel).
- **Random/Unknown/Unclear/Other.** This field should be used for instances of terrorist attacks where the victims appear to have been chosen at random; where victims are described as coming "from all walks of life" or similar terms; where victims are described in terms that do not relate to any of the above categories (e.g., "women and children"); or where no description is given. If this field is used, use the "Comments" and "Description" text fields to record any potentially relevant information about victim identity or the source of any ambiguity.

Coders should record all categories explicitly mentioned in any reports of a given event, but coders should not record any categories not explicitly mentioned in those reports, even if those additional categories are widely known. For example, if a report identifies victims as "Kosovar Muslims," a coder would record both a religious and an ethnic/national identity; but if a report simply identifies the victims as "Kosovars," a coder would only record an ethnic/national identity. Note also that these fields should not be used to capture information about individual victims if that information is not apparently intended as representative of a larger set of the victims (e.g., details provided in an article that describes the different interests and occupations of individual victims). Finally, coders should record any information that led her to conclude something about the identity of the victims (e.g., "tribal violence" or "sectarian killings") in the "Description" field.

Field Names:
Victim Noncombatant Status Asserted [choose the appropriate category, record relevant text in associated field]
Victim Noncombatant Status Contested [choose the appropriate category, record relevant text in associated field]

For victim identities, enter text providing information on the identity when relevant
Victim Identity Political
Victim Identity Ethnic
Victim Identity Religion
Victim Identity Socio-Economic
Victim Identity Unarmed combatant
Victim Identity Random/Unclear/Other

Casualties

These fields are used to record information about the number of deaths or injuries associated with a particular event and whether or not those numbers were reportedly contested. For events described in multiple articles, use the highest available figure or estimate, unless a more recent article offers a figure or estimate that is explicitly presented as a correction of previously reported figures, in which case the purportedly correct figure should be used instead.

The categories for death and injury scales are as follows:

- 1-4.
- 5-24. Keywords include "several", "a few", "a number of"
- 25-49.
- 50-99. Keywords include "tens", "dozens", "scores".
- 100-999. Keywords include "hundreds".
- 1,000-9,999. Keywords include "thousands".
- 10,000+. Keywords include "tens of thousands".
- No information.

When a phrase such as "up to 30", "about 30", or "more than 30" is used, base the category on the reported number ("30" in this example) but use the "Range Reported" ambiguity code.

The categories for the ambiguity variable are as follows:

- **Single Number Reported.** Indicates that the report or reports provided a single whole number.
- **Range Reported.** Includes cases where a numeric range is reported, or where a descriptive word or phrase is used to suggest a number, e.g., "dozens" or "hundreds".
- **Multiple Numbers/Ranges Reported.** Indicates that the report or reports provided different death counts or estimates.
- **No Numbers/Ranges Reported.** Indicates that the relevant report or reports provided no indication of the number of deaths.

The categories for the contested variable are as follows:

- **Not Contested.** Reports include death counts (as numbers or ranges) and do not explicitly indicate that any group is contesting those counts.
- **Contested.** Reports include different death counts (as numbers or ranges) that are explicitly described as competing or contesting claims.
- **No Numbers/Ranges Reported.**

To measure the scale of violence, we are also interested in recording information on victims injured as well as killed. The variables for injured victims follow the same coding rules as those for deaths.

Field Names:
Deaths Number [record number if given]
Deaths Scale [select from categories]
Injured Number [record number if given]
Injured Scale [select from categories]
Deaths Ambiguity [select from categories]
Deaths Contested [select from categories]

Modes of Violence

These variables capture information about the nature of the violence involved in an atrocity or atrocities. One focuses on the nature of the interaction between the perpetrators and victims, while the other focuses on the weapon or weapons reportedly used.

The categories for organization variable are as follows:

- **Riot/Pogrom.** Group of perpetrators engages in apparently disorganized or rudimentarily organized and relatively spontaneous attacks. Keywords include riot, pogrom, rampage.
- **Attack/Massacre.** Single or multiple perpetrators engage in planned attack. Keywords include massacre, assault.
- **Unclear/Other.** Use associated text field to record any potentially relevant keywords.

In ascending order of lethality, the categories for the weapons variable are as follows. If multiple types of weapons are used, record the most lethal. The associated text field should be used to record more detailed information about the weapons used when such information is available.

- **Primitive Weapons.** Perpetrators directly inflict violence on victims at close quarters using primitive weapons such as machetes, spears, clubs, or knives.
- **Firearms.** Perpetrators directly inflict violence on victims at close quarters using small firearms, such as pistols, rifles, light automatic weapons, or grenades.
- **Explosives.** Perpetrators plant an explosive device that is expected to detonate without killing the perpetrators.
- **Suicide Bombing.** Perpetrators detonate explosive device that also takes their lives, apparently by design.
- **Heavy Weapons.** Perpetrators use crew-served weapons such as heavy machine guns, artillery, tanks, or aircraft.
- **WMD.** Perpetrators use chemical, biological, or radiological weapon(s) to carry out a mass killing.
- **Unclear/Other.** Use associated text field to record any potentially relevant keywords.

Field Names:
Organization of Violence [select category]
Weapons [select category]

Perpetrator Intent

For an event to qualify as an atrocity according to the conceptual framework underpinning this project, perpetrators must have intended to kill noncombatant civilians. As noted earlier, however, intent is difficult to establish, and we generally aim to finesse this problem by assuming intent unless otherwise stated in the journalistic accounts.

Nevertheless, we expect many reports will contain explicit or implicit information about perpetrator intent, and we wish to take advantage of this information where it is available. To do so, we have created a variable that give coders an opportunity to record whether or not intent was asserted; if it was, whether that assertion was disputed. This variable allows us to narrow or broaden the sample of events used in any analysis, depending on how conservative we want to be in our approach to the question of what constitutes an atrocity.

Note that in all instances, the statements in question may be made by the journalist or by a source the journalist quotes or cites.

- **Intent Asserted And Not Disputed.** At least one report states that the perpetrators intended to kill the victims in question and no other reports dispute that assertion or offer conflicting accounts.
- **Noncombatants Not Intentionally Targeted.** Description of event states that the action by the perpetrators was deliberate, but that the intent was not to kill noncombatants. This will generally be used when the killings are the result of collateral damage, or when the perpetrators were not aware that the victims were noncombatants.
- **Intent Asserted/Conflicting Accounts.** A single report or multiple reports offer competing accounts of whether or not the perpetrators intended to kill the victims in question. For example, some reports might say that fire was directed into the crowd while others stated that the intent was to fire warning shots over the crowd.
- **Intent Apparent But Not Stated.** Description of event makes evident that perpetrators intended to kill victims in question, but intent is not specifically stated in the report(s). This would apply if actions were undertaken which a reasonable person would assume were intended to cause deaths—for example, firing live ammunition directly into a crowd or setting an explosive device—but the report does not explicitly state that there was intent.
- **No Information/Unclear/Other.** This situation is most likely to arise when the perpetrators of the violence are not identified. For example, a report stating that "Sixteen villagers were killed by gunfire in the village of Noreste on Tuesday" would most likely land in this category.

Additional comments on the extent of the specificity/ambiguity of the report can be recorded in the "Comments" field.

A second variable is used to record expressions of regret or apology made by the perpetrators or their leaders or commanders. This variable is designed to help us distinguish cases where noncombatants appear to represent so-called "collateral damage" in the course of combat or conflict. As usual, clarifying information should be recorded in the "Comments" field.

- **Perpetrator Regrets.** Reports include expressions of regret or apology made by the perpetrators or their commanders within one month of the event's occurrence.
- **Perpetrator Approves.** Reports indicate that the perpetrator approves of the killings; these statements are made by the perpetrators or their commanders within one month of the event's occurrence.
- **No Information/Unclear/Other.** No expressions of regret, apology or approval are reported within a month of the event's occurrence.

A third variable gives coders an opportunity to indicate whether or not reports state or imply that noncombatant civilian deaths represent "collateral damage" incurred in the course of a military or police action aimed at military or criminal targets. While collateral damage is always tragic and the indiscriminant use of force may qualify under certain circumstances as a war crime, we do not consider collateral damage equivalent to an atrocity. Ideally, we would use this characteristic as a criterion in deciding whether or not to record an event in our data set. We suspect, however, that the reporting on this issue is not always consistent and thorough, and coder subjectivity may color judgments about what constitutes collateral damage. To avoid excluding events of potential interest, then, we have decided to include such events while recording any indication that the civilian deaths

were collateral damage in this separate field. For events coded as collateral damage, coders should provide some information about the apparently intended target in the "Description" field.

- **Collateral Damage.** Noncombatant deaths occurred in the context of a military or police action apparently aimed at military or criminal targets, and any civilian deaths were apparently inadvertent.
- **Not Collateral Damage.** No information given that suggests civilian deaths were collateral damage.
- **Unclear/Other.**

Field Names:
Intent [select category]
Regrets [select category]
Collateral Damage [select category]

Related Tactics

These categories are used to indicate other behaviors that may accompany the atrocity and show up in the reports. Note that these are coded only if they are mentioned in the context of the atrocity; separate incidents where the tactic was used but cannot be linked to an atrocity are not coded. The following behaviors are coded:

- **Food/Aid as a Weapon.** Withholding of food aid, deliberate destruction of crops, attacks on food-aid convoys, interference with the delivery of medical or other humanitarian aid, etc. Use this only when the interference with food production is a specific deliberate action; do not use if starvation deaths result from famine that is due to the general disruption causes by warfare or dislocation.
- **Scorched Earth Tactics.** This includes wide-scale burning of houses and other buildings, destruction of infrastructure such as roads, power lines, and water treatment systems, and other tactics used to make an area uninhabitable.
- **Human Shields.** Moving noncombatants onto battlefield as protection, hiding in hospitals, hiding among crowds, etc.
- **Rape.** Include abduction for use as "sex slaves".
- **Targeted Assassinations.** An "assassination" occurs when an individual has been identified as a leader, rather than simply because they are a member of some group.
- **Mass Detentions.** Holding noncombatants in camps, use of noncombatants for forced labor, etc. The distinction between a "mass detention" and a "kidnapping" is that mass detentions are cases where tens or more people are rounded up in some short period of time and held in prisons or camps in some quasi-legalistic fashion.
- **Sieges/Closures.** These are situations where the freedom of movement of non-combatants is restricted to a limited geographical area, but they are not forcefully transported to the area, and usually the inhabitants maintain internal control of the area they are restricted to (unlike a prison or concentration camp).
- **Kidnappings.** This includes the abduction (of children or otherwise) for use as fighters. Note that the "human shield", "sex slave", and "mass detention" categories also involve

forceful abduction: the "kidnapping" category should be used only when the abduction could not be categorized as being motivated by any of these other categories.

- **Disappearances.** Noncombatants disappear without evidence of death or abduction (e.g., Argentina or Chile). This will also occur when there are reports of additional possible victims—for example following an attack on a demonstration—who cannot be accounted for among the known dead.
- **Other Tactics.** Any other coercive tactics that would involve harm to person, for example torture; give details in the "Description" field.

Field Names:
Food/Aid [Boolean]
Scorched Earth [Boolean]
Human Shields [Boolean]
Rape [Boolean]
Assassinations [Boolean]
Detentions [Boolean]
Siege/Closure [Boolean]
Kidnapping [Boolean]
Disappearance [Boolean]
Other Tactics [Boolean]

Description

This field should contain a short description (typically two or three sentences) describing the event in narrative English. The description should resemble the lead for a newspaper article, describing the who, what, where, when, and why of the event in question, to the extent possible from the available information. The description should indicate whether the event's status as an atrocity is contested and should provide information about any wider conflict to which the event is apparently linked. Include any keywords that you think might be useful to an analyst searching the database. Information about the credibility of the source reporting the event, and any uncertainty that the coder had in coding various data fields should be recorded in the Comments field rather than in the description.

Field Names:
Description [text]

Link

This field is used only when you are recording multiple reports of the same event that differ on some of the fields in the data (for example, the number of individuals killed; the perpetrators, death counts for a specific village, and so forth). Note that in a situation where there is an extended period of violence—for example Cambodia 1975-77, Rwanda 1995, former Yugoslavia 1991-1995—there may be a very large number of linked records.

The link field is a unique identifier consisting of the form AAAmmdd-ss (e.g., PAS0828-02), where

AAA	coder initials
mmdd	numerical month and day when case was first coded
ss	two digit serial number assigned by the coder.

The purpose of the link is solely to create a unique identifier to an incident so that multiple reports can be combined later and manipulated in a relational database. The actual content doesn't matter so long as this is unique, but keeping to this format should insure that we don't get duplicates. We may also assign some standard identifiers for cases of large-scale mass violence such as Cambodia and Yugoslavia.
If you are using only a single report of an incident, leave this field blank; the system will fill it in automatically. Linked records should only be used when the information in multiple reports cannot be accommodated in a single record. In these cases, just create a single record and document the multiple sources in the "Comments" field.

Field Names:
Link [text]

Data Source

We will be using three different types of sources; these are differentiated by the Source Type field. Attribute the story to the most general source you found it in (with the priority being International, then IGO/NGO, then Local); in other words, if an event is reported in both Reuters and a local newspaper, base the coding on the Reuters report (but the local report can be used to determine whether elements of the reporting are disputed).

- **International.** These are major international news gathering organizations (e.g., Agence France Presse, Associated Press, Christian Science Monitor, New York Times, Reuters, United Press International). To be attributed to an international source, the incident must be reported as fact, rather than as a rumor or allegation (see earlier comments under "Rumor/Allegation").

  Note: Under the 2004-2007 coding contracts, only international sources have been coded. The remaining categories were included for completeness in the event that the data set were extended.
- **IGO/NGO.** These are intended to be reports from major IGOs such as UNHCR, UNICEF, and the International Committee of the Red Cross/Red Crescent, and major NGOs such as Amnesty International, Human Rights Watch, and Doctors without Borders. "Major" means that the group has been in existence for some time and works in multiple areas around the world. Assume that any UN-affiliated IGO is "major." If there is a question about the status of an NGO, assume it is local, even if it works in more than one country.
- **Local.** This is everything else, and includes regional news agencies (e.g. Xinhua, Al-Jazeera), local newspapers, radio and television services, and local NGOs. The country-specific sources will typically be picked up from FBIS, which specializes in providing this type of coverage. Many regional agencies are available in NEXIS and Factiva. Note that some of these sources—particularly those affiliated with governments, militant groups or political parties—may be rather dubious but code them anyway; just make sure to provide information about the source itself.

Primary and Secondary Sources

The primary source is the source of the story that the information has been coded from. A secondary source occurs when the primary source is basing the story on a report from another source, rather than doing original reporting: for example Reuters quoting the NGO Doctors without Borders or

quoting a local newspaper. If the story consists of both original reporting and quotes from other sources, then code it as being only a primary source story, but indicate the other sources in the "Comments" field. If multiple secondary sources are used, code the one with the highest priority (that is, IGO/NGO sources have priority over local sources.)

Contesting Sources

The "contesting sources" field is intended to give some indication of the extent to which the report is contested. For example, in some situations government sources—which are "local" in our typology—will contest every report from international sources. However, sometimes the international sources (e.g. Reuters and AFP) may themselves disagree. In the "contesting sources" field, indicate the source with the highest priority that has contested any significant part of the report—that is, one of the "contested" fields has been checked in the report. If there is only a single report, or the reports are in sufficient agreement that none of the "contested" fields has been checked, this field is blank.

Citation

Record the following bibliographic information for each source:

Newspapers: Name, date and page

News Agencies and transcripts of electronic media: Date and any story identification number; these may or may not be available depending on the source (for Factiva, use the 25-character identification number that follows the headline)

Sources obtained from FBIS: Always include the FBIS story identification number in addition to information on the original source; as long as we have this, it is very easy to retrieve the original story

Field Names:
Primary Source Type [select from list]
Primary Source [select from list if International or IGO/NGO; otherwise enter identifying information]
Secondary Source Type [select from list]
Secondary Source [select from list if International or IGO/NGO; otherwise enter identifying information]
Contesting Source Type [select from list]
Contesting Source [select from list if International or IGO/NGO; otherwise enter identifying information]
Citation [text]

Comments

This is a text field that can be used to record any comments relevant to the sources and coding, for example source of the original lead identifying the event, multiple reports of the event that agree on all of the particulars, any uncertainty in the assignment of categories (e.g. identity of perpetrators or date) and so forth. Comments are used to assess the quality and reliability of the reports, where the description is used to provide information about the event itself.

Field Names:
Comments [text]

Coder

Enter your initials or other information that will allow us to identify who did the coding.

Field Names:
Coder [text]

Revision History

Version	Date	Changes
0.8	September 2005	Original public version
0.9	July 2006	Assorted changes to reflect actual implementation of the coding, particularly the use of free text rather than categories in many fields
1.0	November 2007	Geocoding fields added