PLOVER

Political Language Ontology for Verifiable Event Records

Event, Actor and Data Interchange Specification

Open Event Data Alliance

`http://openeventdata.org/`

DRAFT Version: 1.3

26 January 2023

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Acknowledgments

Contributors to the development of PLOVER include, in alphabetical order, Benjamin Bagozzi, Andreas Beger, John Beieler, Liz Boschee, Patrick Brandt, Andrew Halterman, Jill Irvine, Jennifer Holmes, Javier Osorio, Tom Parris, Grace Scarborough, and Philip Schrodt.

The Open Event Data Alliance is an educational and open research corporation chartered in the Commonwealth of Virginia, United States.

The PLOVER logo is based on a drawing found at
`http://www.rspb.org.uk/discoverandenjoynature/discoverandlearn/birdguide/name/r/ringedplover/`

Funding for PLOVER was initially provided in part by the U.S. National Science Foundation award SBE-1539302, "RIDIR: Modernizing Political Event Data for Big Data Social Science Research". Any opinions, findings, conclusions or recommendations in this document are those of [at least one of] the authors and do not necessarily reflect the views of the National Science Foundation, or any company or government agency employing or funding the authors or otherwise contributing to the document.

Additional work was sponsored by the Political Instability Task Force (PITF). The PITF is funded by the Central Intelligence Agency. The views expressed in this codebook are the authors' alone and do not represent the views of the US Government.

GitHub repository: `https://github.com/openeventdata/PLOVER`

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright © 2021 by the Open Event Data Alliance

Latest update: Thursday 26$^{\text{th}}$ January, 2023 (UTC)

Chapter 1

Introduction

The concept of political event data originated in the academic quantitative international relations community in the mid-1960s. While a number of projects produced some event data, often for specialized applications, eventually two coding frameworks dominated the production of general-purpose event data sets: Charles McClelland's WEIS (McClelland, 1967, 1976) and the Conflict and Peace Data Bank (COPDAB) developed by Edward Azar (Azar and Sloan, 1975; Azar, 1980, 1982). Both were created during the Cold War and assumed a "Westphalian-Clausewitzian" political world in which sovereign states reacted to each other primarily through official diplomacy and military threats. Consequently these coding systems proved less than optimal for dealing with post-Cold-War issues such as ethnic conflict, low-intensity violence, internal conflict and repression, and multilateral intervention.

During the early 2000s, the CAMEO framework—Conflict and Mediation Event Observations—was developed (Schrodt et al., 2009) to support an NSF-funded project at the University of Kansas on the study of inter-state conflict mediation, not as a general-purpose event ontology. Nonetheless, it was gradually adopted as a "next generation" coding scheme, notably for the DARPA-funded Integrated Conflict Early Warning System (ICEWS) project (O'Brien, 2010) because it corrected some of the long-recognized ambiguities in WEIS and COPDAB, and was explicitly designed both for automated coding and for the detailed coding of sub-state actors. It was continued in the widely-used public ICEWS data (`https://dataverse.harvard.edu/dataverse/icews`) coded using the BBN SERIF/ACCENT coder, with BBN doing considerable additional work on various details of the system.

As event data came into wider use in the 2010s, several problems with CAMEO became apparent, largely dealing with the complexity of the system, the absence of a standard format beyond the original date-source-target-event fields (for example, representing geographical location), and continuing interest on coding substate activities not present in the earlier systems.

To address these concerns, an informal group of academic, government and private sector producers and users of event data met and circulated drafts during the fall of 2016 to develop a new, simplified and more flexible event data specification to replace CAMEO, which became PLOVER and the basis of this document: Section 7.2 summarizes the major changes. Additional extensive work was done in 2021 as PLOVER was adopted by the Political Instability Task Force to replace CAMEO in its new coder for iDATA (the next generation ICEWS).

Because the PLOVER event categories are generally a simplification of CAMEO our expectation is that it will be relatively easy to splice existing CAMEO data sets to PLOVER equivalents by simply collapsing the two- and three-digit categories. The scaled "PLOVER scores" are also designed for splicing with time series generated from the CAMEO "Goldstein scores." The standardization

of the JSON field names—as well as adoption of JSON as the data interchange format—will allow the development of general-purpose utilities that can work with all formats, in contrast with the current proliferation of incompatible and error-prone CSV and tab-delimited formats.

Compared to the Kansas and BBN CAMEO manuals—though curiously, consistent with the public documentation for WEIS and COPAB in the ICPSR archive—at this point we have provided only general guidance on the content of the various categories, modes, and contexts. With the current state of automated natural language processing, any 2020s automated coding system will almost certainly be implemented using machine learning systems trained on a labeled set of news texts and those training cases effectively are the detailed examples. This differs from the older systems which classified events using dictionaries which were abstracted, by human developers, from the texts. We eventually hope to provide a set of training cases, possibly synthetic, that will be free of intellectual property constraints; at present a small set of cases extracted from the Kansas CAMEO manual is available on the PLOVER GitHub site `https://github.com/openeventdata/PLOVER` but it is not sufficient for training a system, at least with 2021 technologies.

1.1 Why "PLOVER"?

Plovers (*Charadriidae*) are a globally-distributed family of short-billed gregarious wading birds who spend their lives frantically poking through endless stretches of sand and muck trying to find something of interest. It is difficult to imagine a better analogy to the process of coding event data.

Chapter 2

Event, Mode, and Context

2.1 Overview

A major difference between PLOVER and the earlier widely-used event coding systems is moving the information in the hierarchical 3- and 4-digit categories of CAMEO into three components: `event-mode-context` generally corresponding to "`what-how-why`." We anticipate at least five advantages to this approach:

1. The three `what-how-why` components are now distinct, whereas various CAMEO subcategories inconsistently used the *how* and *why* to distinguish between subcategories.

2. Because a `context` can be applied to any event category and, where relevant, any `mode`, PLOVER has far more combinations of codes for describing events than the fixed hierarchy of CAMEO.

3. We are probably increasing the ability of general machine-learning classifiers—as distinct from the older customized dictionary-based parser/coders—to assign `mode` and `context` compared to their ability to assign subcategories.

4. In initial experiments, it appears this approach is *much* easier for humans to code than the hierarchical structure of CAMEO because a human coder can hold most of the relevant categories in working memory. More generally, `event-mode-context` coding uses words, not numerical codes, so coders will probably be using the parts of the brain (Broca's area) which are specialized for processing words. No known specialized cognitive facility exists for handling some 250 2-to-4-digit codes.

5. Because the words used to differentiate `mode` and `context` are generally very basic, translations of the coding protocols into languages other than English is likely to be easier than translating the subcategory descriptions found in CAMEO.

While both `mode` and `context` will usually take a single value, in some instances multiple values will be appropriate and this is allowed. Both fields are optional, and if no existing values seem appropriate, the field should be left null, though perhaps with some details provided in the JSON `comment` field, particularly when the record is generated using human coding.

2.2 Context codes that can be used with any category

The `context` field re-introduces, albeit in a greatly extended form, a concept found in the original COPDAB data (but absent from WEIS and hence CAMEO) which allows, for example, a distinction in the event record between a meeting dealing with military issues and a meeting dealing with economic issues. Human analysts naturally incorporate this information in their reading of an article. Based on some initial experiments, we believe that with contemporary text classification algorithms this should be relatively easy to implement.

Additional details and clarifications for some of the more ambiguous or complicated general contexts listed in Table 2.1 can be found in chapter 4.

Table 2.1: General contexts
(Each context has additional details in chapter 4)

Name	Content
military	military, including military assistance; only covers official state military bodies, *excluding* paramilitary or militia organizations, law enforcement, or non-state armed actors
intelligence	gathering of information by state intelligence organizations, including more general discussions of those agencies
executive	executive agencies and bureaucracies, executives and their families, as well as personal and political issues involving non-elected heads of state
legislative	legislative debate, parliamentary coalition formation
election	elections and campaigns
political_institutions	changes to—as well as non-routine interactions between and within—formal political bodies, including political parties
pro_democracy	processes and initiatives associated with democratization
pro_autocracy	processes and initiatives associated with autocratization
economic	economic news, trade, finance, economic development, inflation/GDP/recessions, or monetary policy. *Exclude* news solely about individual companies
legal	courts and judiciary; national and international law
reparations	reparations, i.e., monetary or non-monetary compensation for past harm inflicted
human_rights	unspecified human rights in a general sense, often but not always literally including the term "human rights"
rights_freedoms	political rights and freedoms, civil liberties, press/media freedom, restrictions on civil rights, gatherings, etc.
repression	forceful repression of groups, individuals, or societies by security forces
human_security	access to water, food, housing, energy, land/property tenure (e.g. seizure of crops), etc.
gender	mentions of gender, women's issues, gender equality, etc., including abortion and issues of reproductive rights
lgbt	LGBTQ issues such as related laws, discrimination, and protests
religion_ethnicity	religious, ethnic, caste, or linguistic issues, including discrimination, violence, cooperation, or more broadly important mentions
inequality	the unequal distribution of resources and opportunities

Continued on next page

Table 2.1: General contexts
(Each context has additional details in chapter 4)

Name	Content
diplomatic	diplomacy, which can be both international and domestic
territory	international and domestic territorial disputes, protests and rebellions over territory, and forced displacements
peacekeeping	peacekeeping, typically multilateral operations
health	public health, disease outbreaks and epidemics, disease in general
asylum	discussions of seeking or granting asylum
migration	migration, refugees, and displaced people
disasters	disasters including both "natural" and accidents/spills etc.
natural_resource	natural resource extraction, trade, dependence, and pollution
environment	pollution and related environmental problems; environmental policy, environmental cooperation, and environmental protest
technology	technological development, new technology, high-tech industry
terrorism	events related to terrorist groups; anti-terrorism actions and policies
cbrn	chemical, biological, radiation, and nuclear weapons or attacks
cyber	cyber attacks and crime; events with this context are not necessarily cyber events (e.g. hacking) themselves
misinformation	misinformation, influence operations, fake news; events with this context are not necessarily misinformation themselves
crime	Any discussion of crimes or criminals, *excluding* events that have one of the more specific crime contexts like trafficking
corruption	misuse/abuse of public office, including corruption, bribery, cronyism, nepotism, abuse of public office
illegal_drugs	criminal possession, distribution, sale, or manufacturing of illegal drugs
trafficking	trafficking or smuggling of humans, drugs, animals, or other goods; both domestic and international

How these contexts are used in practice will depend on how a coder implements them. If they are applied using document-level classifiers, researchers/analysts could misinterpret the meaning of contexts. For example, an ASSAULT event with an "elections" context does not necessarily imply electoral violence. It could be an article about, for instance, violence in Afghanistan in the context of an article on US election politics. Thus, the interpretation of "context" may depend on how it's implemented in practice. Certain types of events, particularly general protests and meetings, will also have multiple contexts: this is a feature, not a bug.

Based on our past experience developing event-oriented data sets, we are almost certain to find some additional contexts as we start coding data, so this list is likely to change somewhat. In many cases, it may be possible to extract the information needed for specialized applications by simply modifying the `context` coding—which should be relatively easy—rather than modifying the event category coding, which is more difficult but was the only option in the older hierarchical systems.

2.3 Auxiliary modes

Note: these Auxiliary modes have not yet been fully implemented in any PLOVER-based coding projects to date.

PLOVER also includes four *auxiliary modes*, which provide information about whether a reported event is **historical**, **future**, **hypothetical**, or has a **negation**. PLOVER assumes that the coding engine will be able to resolve these and put that information in the `context`. Negated events can be excluded from the final event dataset. Examples:

- **historical**: "During the decolonization struggle, Angolan forces..."
- **future**: "Members of the G-7 will meet in Ottawa next month..."
- **hypothetical**: "If Russian forces were to cross the border, that would represent a major..."
- **negation**: "Thus far, fighting has not re-emerged in the tense region."

Theoretically, some event types can be represented as other event types plus an auxiliary mode: AGREE could be SUPPORT + **future** or THREATEN could be ASSAULT + **hypothetical**. In situations like these, the coder should always return the single event category, not a category+auxiliary.

We anticipate that in general—and consistent with earlier event coding schemes—it will be possible to code `mode` from the same sentence used to code the event, or possibly that sentence and one before it. `context`, in contrast, will usually be coded at the paragraph- or document-level: this differs from earlier automated coding, though probably is similar to human-coded data such as COPDAB and BCOW (Leng, 1987) where `context`-like fields were coded.

2.4 Changes in treatment of political entities

2.4.1 "actor" and "recipient" replace "source" and "target"

In order to reduce overlap with how the policy community uses the terms "source" and "target", we are using the word `actor` to refer to the entity or entities who initiated the event, and `recipient` to refer to the the entity or entities to whom the event is directed, if this is clear. `recipient` is optional for many events. We are using the term "entity" to refer to the class of potential actors and recipients.

2.4.2 Compound and Reciprocal Actors

Following the dyadic approach which originated in WEIS and COPDAB, most CAMEO-based coders dealt with compound entities—"The United States and France accused Russia..."—by generating multiple events: this example would generate two events of the form

```
USA  RUS  112
FRA  RUS  112
```

This approach, however, gets very problematic in the not-uncommon situation where an alliance is involved and is expanded to all of its constituent members: a single reference to the G20 expands to at least twenty events, and a *meeting* of the G20, generating reciprocal events, expands to 380 events, which is one of the reasons "consult" events are so frequent in CAMEO-coded data.

In PLOVER, compound entities generate a single actor or recipient, but with multiple members: the actor and recipient are a list of entities rather than a single entities. Depending on the

application, a user might expand this in a post-processing phase to multiple events with single entity codes following the earlier conventions, but the initial coding of the event uses the list.

PLOVER also uses actor lists to deal with reciprocal events. In CAMEO, a meeting "President Obama met with Japanese Prime Minister Abe at the White House" generated two events

```
JAP  USA  042
USA  JAP  043
```

where 042 and 043 are the CAMEO codes for "visit" and "host" respectively.[1] While the PLOVER CONSULT mode provides for a host/visit distinction, this is not required. In such instances, all of the entities are considered as the actor and no recipient is included.

In ASSAULT events, reciprocal violence—as distinct from one-sided violence—is handled in a similar fashion, with both entities as `actor`: this applies in any event where both sides are using force, even if one side "started it", an assessment often as not contested anyway. One-sided violence, such as assassinations or police firing on demonstrators, will have the perpetrator as the `actor` and the victims as `recipient`. The `sideAB` mode in ASSAULT is used in conjunction with the `recipient` field when there are two clear "sides" in the conflict.

2.5 Numeric conflict/cooperation scores

Researchers and analysts often want to represent events along a conflict/cooperation scale. In CAMEO, these were called "Goldstein scores" since they were an extension of the WEIS scores in Goldstein (1992) which mapped the CAMEO codes to a -10 to $+10$ scale. The "PLOVER scores" given in Table 2.5 provide comparable scaled conflict/cooperation scores for PLOVER. An initial mapping was created by taking a weighted average of the CAMEO-based Goldstein scores for each PLOVER category, with the weights being the empirical frequency of the CAMEO event type in an 18 month sample of data (October-2017 to March-2019).

Using this initial mapping and the original CAMEO-based Goldstein scores as starting points, a subject matter expert then qualitatively developed a set of PLOVER scores for each of PLOVER's event types and modes. A second subject matter expert simultaneously developed an alternative set of proposed PLOVER scores through a combination of (i) pairwise comparisons of PLOVER event type-mode combinations and (ii) a Bradley-Terry model to estimate locations of each PLOVER event type-mode on a cooperative–conflictual scale. The two separately developed sets of PLOVER scores were then reviewed by a team of five subject matter experts (including the two mentioned above) to reconcile discrepancies. Ultimately, the final PLOVER scores presented below reflect a re-scaled average of the two sets of expert-derived PLOVER scores, with a small number of qualitative score adjustments that were implemented by the larger team of experts upon discussion. By and large, these final PLOVER scores reflect the original "Goldstein scores" in ranging from -10 to $+10$, with more positive values reflecting more cooperative actions and more negative values reflecting more conflictual actions.

[1] This example optimistically assumes the coding system was clever enough to recognize the significance of the phrase "at the White House."

Table 2.2: PLOVER conflict–cooperation scores

CATEGORY: Mode	Score	CATEGORY: Mode	Score	CATEGORY: Mode	Score
AGREE	2.5	REJECT	-2	COERCE	-7
CONSULT	2.5	REJECT: Assist	-1.5	COERCE: Seize	-7
CONSULT: Visit	4	REJECT: Change	-2	COERCE: Restrict	-7
CONSULT: Third-party	1	REJECT: Yield	-2	COERCE: Ban	-7
CONSULT: Multilateral	4	REJECT: Meet	-2	COERCE: Censor	-6.5
CONSULT: Phone	1.5	THREATEN	-3.5	COERCE: Curfew	-6.5
SUPPORT	5	THREATEN: Restrict	-3	COERCE: Martial-law	-7.5
CONCEDE	4.5	THREATEN: Ban	-3.5	COERCE: Arrest	-7
COOPERATE	9.5	THREATEN: Arrest	-3.5	COERCE: Deport	-7
AID	10	THREATEN: Relations	-3.5	COERCE: Withhold	-7
RETREAT	6.5	THREATEN: Expel	-3.5	COERCE: Misinformation	-6.5
RETREAT: Withdraw	8	THREATEN: Territory	-4	COERCE: Cyber	-7
RETREAT: Release	6.5	THREATEN: Violence	-3.5	ASSAULT	-9
RETREAT: Return	5.5	PROTEST	-5	ASSAULT: Abduct	-8.5
RETREAT: Disarm	10	PROTEST: Demo	-5	ASSAULT: Beat	-7.5
RETREAT: Ceasefire	7.5	PROTEST: Riot	-7	ASSAULT: Torture	-8.5
RETREAT: Access	6.5	PROTEST: Strike	-5	ASSAULT: Execute	-9.5
RETREAT: Resign	4.5	PROTEST: Hunger	-4.5	ASSAULT: Sexual	-8.5
REQUEST	0	PROTEST: Boycott	-4.5	ASSAULT: Assassinate	-9.5
REQUEST: Assist	0.5	PROTEST: Obstruct	-4.5	ASSAULT: Destroy	-8.5
REQUEST: Change	-1	SANCTION	-6	ASSAULT: Primitive	-8
REQUEST: Yield	-1	SANCTION: Convict	-6	ASSAULT: Firearms	-9
REQUEST: Meet	0	SANCTION: Expel	-5.5	ASSAULT: Explosives	-9
ACCUSE	-3	SANCTION: Withdraw	-6	ASSAULT: Suicide-attack	-9.5
ACCUSE: Disapprove	-2	SANCTION: Discontinue	-6	ASSAULT: Aerial	-9.5
ACCUSE: Investigate	-3.5	MOBILIZE	-5.5	ASSAULT: Drone	-9.5
ACCUSE: Allege	-3.5	MOBILIZE: Troops	-6.5	ASSAULT: Heavy-weapons	-9.5
		MOBILIZE: Weapons	-6.5	ASSAULT: Crowd-control	-7.5
		MOBILIZE: Police	-5.5	ASSAULT: Cleansing	-10
		MOBILIZE: Militia	-6	ASSAULT: Massacre	-10
				ASSAULT: Unconventional	-10

Chapter 3

Event Categories

3.1 AGREE

Agree to, offer, promise, or otherwise indicate willingness or commitment to cooperate, including promises to sign or ratify agreements. Cooperative actions of types CONSULT, SUPPORT, COOPERATE, and AID reported in future tense are also taken to imply intentions and should be coded as AGREE.

3.1.1 Potential ambiguities

- As noted in Section 2.3, there's the potential for some events to fit both the definition of AGREE and SUPPORT + **future**. For example, "Russia and the United States *will sign an agreement* limiting certain kinds of weapons...". When situations like this occur, the coder should always return the single event category that fits, rather than a category+auxiliary mode.
- Future cooperative actions of type RETREAT + **future** and CONCEDE + **future** should be left as they are, i.e. they do not reduce to AGREE.

3.1.2 Requires recipient: No

3.1.3 Supplementary fields: None

3.1.4 Quad category: VERBAL COOPERATION

3.2 CONSULT

All consultations and meetings: this includes visiting and hosting visits, as well as meeting at a neutral location, and consultation by phone or other media. Because this type of political event is both frequent and easily (and safely. . .) covered in the international press, it is the largest category in most event data sets. Additional useful keywords for identifying CONSULT: "Holding talks" and "discussions", "negotiations, bargaining, or discussions". See the discussion in Section 2.4.2 on the treatment of actors in CONSULT events.

3.2.1 Potential ambiguities

References to future meetings, summits, state visits, etc. and invitations for state visits should be coded as AGREE.

3.2.2 Requires recipient: No

In CONSULT events where there is no clear distinction between whether an entity is hosting or visiting, all participants are coded in the `actor` field. In events where one side is hosting and one is visiting, the visitor will always be coded as the `actor` and the host will be the `recipient`.

3.2.3 Supplementary fields: modes

Table 3.1: CONSULT modes

Name	Content
visit	`actor` is visiting, `recipient` is hosting.
third-party	Meeting is hosted by a third party
multilateral	Meeting occurs in a multilateral context, typically an alliance or IGO
phone	Consultation occurs via phone or some other remote medium

Adapted from CAMEO.

3.2.4 Quad category: VERBAL COOPERATION

3.3 SUPPORT

Initiate, resume, improve, or expand diplomatic, non-material cooperation; express support for, commend, approve policy, action, or actor, or ratify, sign, or finalize an agreement or treaty. Use this code only for political, diplomatic, and non-material support, including recognition of newly independent states, new governments that might have come to power through unconventional means, and initiation of diplomatic ties with an entity for the first time.

SUPPORT is distinct from the CAMEO APPEAL category, where the actor simply *requested* support from the recipient.

3.3.1 Requires recipient: No

3.3.2 Potential ambiguities

The term used for this category, SUPPORT, is a somewhat ambiguous. Although it may imply a material event, but this category should only be used for verbal cooperation.

- Formal pardons and amnesties of arrested persons should be coded as CONCEDE; the actual release or exchange of prisoners should be coded as RETREAT.
- Expressions of regret or remorse for an action or situation should be coded as CONCEDE.
- Promises to sign or ratify agreements and treaties are coded as AGREE
- Military cooperation or defense should be coded as COOPERATE with a `military` *context*.

3.3.3 Supplementary fields: None

3.3.4 Quad category: VERBAL COOPERATION

3.4 CONCEDE

This covers verbal concessions which have no immediate material consequences, including the promise of future concessions such easing of administrative or legal restrictions on persons and organizations, removing curfews, suspending protests, and declarations (but not implementations) of ceasefires and withdrawals from territory.

CONCEDE, like the verbal components CAMEO/WEIS predecessor YIELD, is inherently problematic since many concessions deal with promises that certain things will *not* happen, or will happen in the distant future (e.g. many policy changes). So, for example, the lifting of a curfew is, effectively, a promise that people will not be arrested for violating the curfew, which itself is not an event. We're treating such concessions as verbal rather than material even though sometimes they have material consequences, e.g. people coming out in the streets after a curfew is lifted, provide they believe the entity lifting the curfew actually has done so.

3.4.1 Requires recipient: No

3.4.2 Potential ambiguities

- Formal pardons and amnesties of arrested persons should be coded as CONCEDE; the actual release or exchange of prisoners should be coded as RETREAT.
- Expressions of regret or remorse for an action or situation should be coded as CONCEDE.
- Future concessions, i.e. CONCEDE + **future** should be left as they are, i.e. they do not reduce to AGREE.
- Announcements of future retreats should be coded as CONCEDE. E.g., "announced...would withdraw troops" is RETREAT + FUTURE and should be coded as CONCEDE.

3.4.3 Supplementary fields: None

3.4.4 Quad category: VERBAL COOPERATION

3.5 COOPERATE

Initiate, resume, improve, or expand *mutual* material cooperation or exchange, including

- Initiate, resume, improve, or expand economic exchange or cooperation.
- Military exchanges such as joint military games and maneuvers.
- Cooperation on judicial matters, such as extraditions and war crimes.
- Voluntary exchanges or sharing of intelligence and other significant information.

3.5.1 Potential ambiguities

- COOPERATE is distinguished from AID because the activity is generally understood to directly benefit both parties, whereas AID is understood to primarily benefit only the recipient.
- Promises, offers, or agreement for future cooperation should be coded as AGREE.

3.5.2 Requires recipient: Yes

3.5.3 Supplementary fields: None

3.5.4 Quad category: MATERIAL COOPERATION

3.6 AID

All provisions of providing material aid whose material benefits primarily accrue to the recipient. Examples include:

- Monetary aid and financial guarantees, grants, gifts and credit, including reparations.
- Military and police assistance including arms and personnel.
- Humanitarian aid such as emergency assistance.
- Asylum, both to persons in its territories (territorial asylum) and diplomatic asylum on the premises of an embassy.

3.6.1 Requires recipient: Yes

3.6.2 Potential ambiguities

- While reparations or voluntary settlements should be coded as AID, court-ordered payments should be coded as SANCTION events with the winner and court as actor and guilty party as recipient.
- Broad promises of future AID should be coded as AGREE.
- Debt forgiveness and loan cancellations, e.g. "The Paris Club has agreed to forgive $200 of Nigeria's $350 million in loan obligations, . . . " should be coded as AID.
- Hostage rescues, e.g. "Pakistan's military rescued 10 hostages from the border region with Afghanistan yesterday" or "Police in Barcelona freed 10 human trafficking victims during an operation yesterday" should be coded as AID.

3.6.3 Supplementary fields: None

3.6.4 Quad category: MATERIAL COOPERATION

3.7 RETREAT

RETREAT covers any events—not just military "retreat" from territory—which have an immediate (not simply promised) material consequences, such as the release of prisoners and hostages, repatriation of refugees, the return of confiscated property, allowing the entry of observers, peacekeepers, or humanitarian workers, disarming, observing a ceasefire or otherwise ending active conflicts, and, of course, a military retreat from, or ceding, territory. RETREAT also covers resignations of government officials.

3.7.1 Requires recipient: No

3.7.2 Potential ambiguities

- Announcements of future retreats should be coded as CONCEDE. E.g., "announced...would withdraw troops" is RETREAT + FUTURE and should be coded as CONCEDE.
- Debt forgiveness and loan cancellations, e.g. "The Paris Club has agreed to forgive $200 of Nigeria's $350 million in loan obligations, . . . " should be coded as AID.
- Hostage rescues, e.g. "Pakistan's military rescued 10 hostages from the border region with Afghanistan yesterday" or "Police in Barcelona freed 10 human trafficking victims during an operation yesterday" should be coded as AID.

3.7.3 Supplementary fields: modes

Table 3.2: RETREAT modes

Name	Content
withdraw	Retreat from territory or withdraw forces from an area
release	Release captives
return	Return property
disarm	Disarm militarily or give up weapons
ceasefire	Implement ceasefire
access	Allow third party (e.g., observers, peacekeepers, humanitarian workers) access
resign	Official resignation

3.7.4 Quad category: MATERIAL COOPERATION

3.8 REQUEST

All requests, demands, and orders. Requests, demands, and orders are less forceful than threats and potentially carry less serious repercussions. Coding will need to rely primarily on the language used by reporters to make this distinction. All requests are verbal acts.

3.8.1 Requires recipient: No

3.8.2 Potential ambiguities

- This category only applies to verbal demands: demands that take the form of demonstrations, protests, etc. are coded as PROTEST.
- When one or more parties to a conflict call for ending the conflict, that is taken to be an expression of intent on the part of the `actor` and is thus coded as AGREE.
- Withdrawing a demand should be coded as CONCEDE.

3.8.3 Supplementary fields: modes

Table 3.3: REQUEST modes. These modes are shared with REJECT.

Name	Content
assist	Any form of exchange, relations, or assistance
change	Any changes in policy, government, or institutions that are not concessions
yield	Release of prisoners, ending sanctions, easing curfews and boycotts, ceasefires
meet	Meetings and negotiations

3.8.4 Quad category: VERBAL CONFLICT

3.9 ACCUSE

- Express disapprovals, objections, and complaints; condemn, decry a policy or an action; criticize, defame, denigrate responsible parties.
- Accuse, allege, or charge, both judicially and informally
- Sue or bring to court
- All investigations, including those of historical cases. Examples include investigations of criminal activity (theft, killing, etc.) and corruption, human rights abuses, war crime, and violations of basic freedoms, military activities such as violations of ceasefire, seizures, and invasions.

3.9.1 Requires recipient: No

3.9.2 Potential ambiguities

Candidates for ACCUSE-allege + context: "misinformation" events could also be COERCE-misinformation, depending on the wording of the relevant text.

3.9.3 Supplementary fields: modes

Table 3.4: ACCUSE modes

Name	Content
disapprove	Express disapproval; condemn; complain
investigate	Any investigation, including commissions, grand juries, judicial or political
allege	Formally or informally accuse; sue, indict, or charge; bring to trial

3.9.4 Quad category: VERBAL CONFLICT

3.10 REJECT

All rejections and refusals.

3.10.1 Requires recipient: No

3.10.2 Potential ambiguities

Withdrawal of military aid or other assistance is coded as SANCTION.

3.10.3 Supplementary fields: modes

Table 3.5: REJECT modes. These modes are shared with DEMAND.

Name	Content
assist	Any form of exchange, relations, or assistance
change	Any changes in policy, government, or institutions that are not concessions
yield	Release of prisoners, ending sanctions, easing curfews and boycotts, ceasefires
meet	Meetings and negotiations

3.10.4 Quad category: VERBAL CONFLICT

3.11 THREATEN

All threats, coercive or forceful warnings with serious potential repercussions. Threats are generally verbal acts except for purely symbolic material actions such as having an unarmed group place a flag on some territory.

3.11.1 Requires recipient: No

3.11.2 Supplementary fields: mode

Note that in THREATEN the mode is the *content* of the threat, rather than how it has been expressed.

Table 3.6: THREATEN modes

Name	Content
restrict	Restrict movement of people or goods, including boycotts, strikes, blockades, and curfews
ban	Threaten to ban political activities of particular parties or individuals
arrest	Arrest, detain, imprison
relations	Threaten to suspend relations, talks
	such as speech, expression, and assembly
expel	Expel diplomats, peacekeepers, NGOs
territory	Threaten to occupy, seize control of the whole or part of a territory
violence	Threaten violence

3.11.3 Quad category: VERBAL CONFLICT

3.12 PROTEST

All civilian demonstrations and other collective actions carried out as protests against the recipient: Dissent collectively, publicly show negative feelings or opinions; rally, gather to protest a policy, action, or actor(s).

3.12.1 Requires recipient: No

3.12.2 Supplementary fields:

mode: Mode of protest: see Table 3.7

event_loc: Location[s] of event

Table 3.7: PROTEST modes

Name	Content
demo	Organized demonstration. Distinct, continuous, and largely peaceful action directed toward members of a distinct 'other' group or government authorities
riot	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
strike	Members of an organization or union engage in the abandonment of work-places, either within specific sectors/industries or across sectors/industries
hunger	Hunger strike
boycott	The boycott of an activity, person, country, or organization via the withdrawal of commercial or social relations
obstruct	Obstruct passage or access to a particular locale

Adapted from Salehyan and Hendix, *Social Conflict Analysis Database* (SCAD) Version 3.2: `https://www.strausscenter.org/images/codebooks/SCAD_32_Codebook.pdf`

3.12.3 Quad category: MATERIAL CONFLICT

3.13 SANCTION

All reductions in existing, routine, or cooperative relations. Note that this is not confined to formal "sanctions"—SANCTION was just the best word we could find for WEIS and CAMEO's "REDUCE RELATIONS"

3.13.1 Requires recipient: Yes

3.13.2 Potential ambiguities

- Convictions that result in imprisonment should also e coded as COERCE 'arrest'.
- Expulsions or deportations of individuals—typically a legal matter—are coded as COERCE.
- Cancellation of meetings are REJECT and therefore verbal conflict.
- While reparations or voluntary settlements should be coded as AID, court-ordered payments should be coded as SANCTION events with the winner and court as actor and guilty party as recipient.

3.13.3 Supplementary fields: mode

Table 3.8: SANCTION modes

Name	Content
convict	Find an entity or behavior guilty, unconstitutional, or liable in a court of law or legally constituted tribunal. This includes not only formal convictions but also instances where an entity is discussed as being found guilty and sentenced or sanctioned via court-order, though court sentences that include imprisonment will also receive a COERCE 'arrest' mode, (see ambiguities above)
expel	Permanently or temporarily expel an entity from a group, organization, political party, or country. This excludes individual deportations, which are coded as a mode under COERCE
withdraw	Withdraw oneself or one's non-military resources (e.g., aid, observers, diplomats, peacekeepers) from a group, mediation activity, organization, or country. This excludes official resignations from an occupation or elected position, which are coded as a mode under RETREAT
discontinue	Curtail, decrease, break, or terminate diplomatic, commercial, or material exchanges in manners not specified above. International (political or economic) sanctions are coded as 'discontinue', although they can occasionally receive additional SANCTION modes depending on the nature of the sanctions

3.13.4 Quad category: MATERIAL CONFLICT

3.14 MOBILIZE

All military or police moves that fall short of the actual use of force.

This category is different from ASSAULT, which refers to actual uses of force, while military posturing falls short of actual use of force and is typically a demonstration of military capabilities and readiness. MOBILIZE is also distinct from THREAT in that the latter is typically verbal, and does not involve any activity that is undertaken to demonstrate military power.

`actor` entities are not necessarily militaries affiliated with states: they can be any organized armed groups (for example militias or gangs).

The `recipient` are entities against whom the `actor` mobilizes its military capabilities in a threatening manner if that is clear, but a group may mobilize with no specific entity stated.

3.14.1 Potential ambiguities

Joint military operations are coded as COOPERATE but single-country exercises should be coded as MOBILIZE.

Events that involved "mobilizing supporters to demonstrate" should be coded as PROTEST.

If a document reports a MOBILIZE event in the context of an ASSAULT, only the ASSAULT should be coded. For example: "military units in the area were activated and returned fire on the rebels" should not report a separate MOBILIZE event.

Mobilizing police or military forces for disaster relief should be coded as AID.

3.14.2 Requires recipient: No

3.14.3 Supplementary fields: modes

Table 3.9: MOBILIZE modes

Name	Content
troops	Mobilize armed personnel or units
weapons	Increase readiness of weapons systems (can occur with a **cyber** context)
police	Mobilize or increase readiness of police or security units
militia	Mobilize or increase readiness of any non-state entity with significant military capability

Adapted from CAMEO cue category 15

3.14.4 Quad category: MATERIAL CONFLICT

3.15 COERCE

Repression, restrictions on rights, or coercive uses of power falling short of violence.

3.15.1 Requires recipient: No

Most cases of COERCE have a clear intended `recipient`, but occasionally, for example in shutting off internet access, the entities intended to be affected by the action are so broad as to be unclear.

3.15.2 Potential ambiguities

- Candidates for COERCE-misinformation could also be ACCUSE-allege + context: "misinformation" events, depending on the wording of the relevant text.
- The "cyber" and "misinformation" modes both have corresponding contexts with the same name. The modes capture specific instances of "cyber" or "misinformation" actions, e.g. a report of a website denial of service attack or a social media disinformation campaign. The contexts are broader, and also capture other actions related to "cyber" or "misinformation" events, like for example the arrest of a person for hacking or cybercrime, or a threat by the recipient of a misinformation campaign to the responsible actor to cease it.

3.15.3 Supplementary fields:

Table 3.10: COERCE modes

Name	Content
seize	Execute search, confiscate property, raid
restrict	Impose restrictions on political freedoms or movement, including cordoning off areas
ban	Ban individuals or organizations
censor	Censor, ban or restrict access to publications or other information
curfew	Impose curfew
martial-law	Impose state of emergency or martial law
arrest	Arrest, detain
deport	Expel or deport individuals
withhold	Withhold public goods/services, e.g. shut off power/internet/water/ utilities or withhold food/medical supplies
misinformation	deception/manipulation/misinformation (will also automatically add "misinformation" context)
cyber	cyber attacks and crime, including hacking, security breaches, website attacks, etc. (will also automatically add "cyber" context)

Adapted from CAMEO cue category 17

3.15.4 Quad category: MATERIAL CONFLICT

3.16 ASSAULT

ASSAULT events are deliberate actions which can potentially result in substantial physical harm.

3.16.1 Requires recipient: No

In ASSAULT events where the violence is two-sided, all participants are coded in the `actor` field except when a "Side A/Side B" can be distinguished per the conventions of the Correlates of War project, in which case the `sideAB` mode is added to any other relevant modes. In one-sided violence, the perpetrator is coded as the `actor` and the victim as the `recipient`.

3.16.2 Potential ambiguities

- General or ambiguous mentions of beatings/assaults without any mention of specific tools or methods only receive the ASSAULT 'beat' mode.
- Mentions of beatings/assaults that use fists, kicks, punches, and similar methods also only receive the ASSAULT 'beat' mode.
- Mentions of beatings/assualts that use blunt instruments such as bats, clubs, and so on receive both ASSAULT 'beat' and ASSAULT 'primitive' modes.
- Mentions of assaults, murders, or clashes that involve knives, machetes, fire, rocks and similar instruments only receive the ASSAULT "primitive" mode.

3.16.3 Supplementary fields:

mode: Mode of violence: see Table 3.11

dead: number killed (integer or code)

injured: number injured (integer or code)

size: used when total casualties are reported, combining dead and wounded

event_loc: Location of event

3.16.4 Quad category: MATERIAL CONFLICT

Table 3.11: ASSAULT modes

Name	Content
abduct	Abduct, kidnap, hijack
beat	Physically assault by striking individuals or groups with one-off or repeated blows, typically administered manually or by blunt instrument
torture	Torture
execute	Judicially-sanctioned execution
sexual	Sexual violence
assassinate	Targeted assassinations (both successful and attempted) with any weapon
destroy	Destroy property
primitive	Primitive weapons: fire, edged weapons, rocks, farm implements
firearms	Rifles, pistols, light machine guns
explosives	Any explosive not incorporated in a heavy weapon: mines, IEDS, car bombs
suicide-attack	Individual and vehicular suicide attacks
aerial	Manned aerial vehicles, e.g. aircraft, helicopters
drone	Unmanned aerial vehicles (UAVs), drones
heavy-weapons	Artillery, rocket launchers, armored vehicles, tanks, and similar weapons
crowd-control	Weapons and tactics intended to be less lethal crowd control, including tear gas, water cannons, firing weapons in the air, lathi charges, etc.
cleansing	Mass expulsions or deportations, ethnic cleansing
massacre	Instances of mass killing or massacres
unconventional	Chemical, biological, radiation, and nuclear weapons
sideAB	Two-sided violence: `actor` and `recipient` are "Side A" and "Side B"

Adapted from Political Instability Task Force Atrocities Database:
`http://eventdata.parusanalytics.com/data.dir/atrocities.html`.

Chapter 4

Contexts

Additional details for the general context tags.

4.1 Military

Discussions of the military or armed forces, including navy, army, marines, air force, and similar military branches. This includes not only discussions of active military operations, but also broader discussions of civil-military relations, military assistance, military culture, and military expenditures. However, this context field only pertains to state militaries. As such, discussions of government-aligned militias, domestic law enforcement, and armed non-state actors (e.g., rebel groups) should not be coded in this context category.

4.2 Intelligence

The covert gathering of information by state intelligence organizations, including through tactics of espionage, overhead reconnaissance, and communications interception. Accusations and suspicions of covert intelligence gathering, as well as more general discussions of state intelligence organizations, should also be coded. However, this context field excludes (1) discussions of the use of clandestine methods to spread misinformation or influence operations, which should be coded under the 'misinformation' context category and (2) discussions of artificial intelligence that are unrelated to state intelligence organizations' intelligence gathering activities.

4.3 Executive

Politics and interactions involving the executive branch, the prime minister, or the non-elected executive of a country. This can include discussions of background context on (e.g.,) an executive's legal challenges, corruption concerns, family, or lifestyle. This context field should also code discussions of executive-branch rivalries or interactions, including instances where an executive or dictator chooses to appoint, demote, or remove a representative from their cabinet, inner circle, or a bureaucratic agency under their purview. Executive relations with other institutions (e.g., foreign heads of state, legislatures, judiciaries, or the military) should generally be coded as 'executive', but will also often receive additional context tags.

4.4 Legislative

All regular and irregular legislative (or parliamentary) politics and interactions. This includes discussions of individual legislative representatives, legislative debates, legislative lawmaking, parliamentary coalition formation, corruption in the legislature, legislative oversight, legislative campaigns and elections. Legislative relations with other institutions (e.g., foreign heads of state, executives, judiciaries, or the military) should generally be coded as 'legislative', but will also often receive additional context tags. Discussions of legislatures or parliaments within authoritarian countries should still be coded under 'legislative'.

4.5 Election

Discussions of elections and political campaigns. This context field includes not only details on specific elections, campaigns, and electoral candidates, but also discussions of domestic and international election monitors. Discussions of concerns over whether elections are free and fair should also be coded as 'elections', in addition (depending on the context) to 'rights_freedoms' and either 'pro_democracy' or 'pro_autocracy'. Campaign fundraising, election debates, charges of electoral fraud, and discussions of voting in (or polling for) an upcoming election should also be coded as 'election'.

4.6 Political Institutions

Changes in the functioning of formal political institutions and parties in a country, including non-routine interactions between different bodies of government, or with political parties. This context should capture all significant changes, regardless of the direction the changes produce, i.e., whether they move the nature of government towards more democratic or more authoritarian. Instead, the latter should be reflected with additional 'pro_democracy' or 'pro_autocracy' contexts.

The 'political_institutions' context works together with the 'pro_democracy' and 'pro_autocracy' contexts to capture the broader concepts of democratic backsliding, autocratization, as well as democratization.

- Includes substantive (legal or illegal) constitutional changes that a clearly motivated to either improve or harm political competition or checks on executive authority, e.g. voting rights strengthening or weakening.
- Can include pseudo-democratic bodies that function under constraints, e.g. in Russia or Iran, not only very democratic countries.
- Excludes routine interactions without broader implications for the state of democracy or functioning of major formal institutions in a country.
- Can include events that describe extra-systemic steps by major actors (institutions), e.g. autocoups or declarations of states of emergency as a means to resolve political or institutional conflict (Tunisia 2021).
- Includes violations of norms regarding appointments, dismissals, pardons, interactions or relationships between parts of government (Boris Johnson UK family appointments to House of Lords in 2021).

4.7 Pro-democracy

Includes processes and initiatives associated with democratization, including actions taken to promote democracy by domestic or international actors. These processes, actions, or initiatives can be gradual or abrupt, and can include, for example, 1) commitments by governments to hold free and fair elections, 2) commitments to enact, follow, or reinforce democratic institutions, or 3) the provision of verbal or material support to pro-democracy actors.

4.8 Pro-autocracy

Includes processes and initiatives associated with authoritarian consolidation, autocratization, dedemocratization, or democratic backsliding, including actions taken to undermine democracy by domestic or international actors. These processes, actions, and initiatives can be gradual or abrupt, and can include, for example, 1) efforts to dissolve or undermine democratic institutions or norms by government officials, 2) the cancellation or suspension of elections, 3) coups, or 4) refusals to step down from power when mandated to do so by existing democratic institutions.

4.9 Economic

Coverage of international and domestic economic issues, including but not limited to trade, finance, economic development, economic stability, economic crises, inflation, unemployment, GDP, and business-government relations in areas of political or economic concern (e.g., a government accusing a company of monopolistic practices, or a government expropriating a company's assets). Political interactions that focus on economic outcomes (such as diplomatic negotiations over a future trade agreement) should also be coded as 'economic' (in addition to other context fields, depending on the specific context). Business news coverage should not be coded as having an 'economic' context unless that business news is discussed in light of a broader economic issue or concern. This means that news reports that solely discuss company annual reports, acquisitions, and similarly mundane business activities should not automatically be coded as 'economic.'

4.10 Legal

Discussions of activities and responsibilities associated with domestic or international courts and judiciaries; discussions of national and international laws; or discussions of associated professions (e.g., lawyers). Negotiations over international law should be coded under this context field. Common international legal activities that should be coded under this category would be International Criminal Court investigations, international tribunals, maritime law, the Geneva Convention, the World Trade Organization, and similar conventions on the use of chemical and biological weapons. Domestically, this context field should code discussions of (e.g.) legal proceedings against politicians, the role of the courts in society or politics, or judicial oversight. However, discussions of domestic or international crime events should not automatically be coded as 'legal' unless detailed discussion of associated laws or court processes is included.

4.11 Reparations

Monetary or non-monetary payments to compensate victims, groups, or family members for past harms inflicted—typically by a national or foreign government. Discussions where such reparations

are simply promised or demanded should also be coded under this context field. Reparations may be made in the contexts of post-conflict justice, but not all post-conflict justice activities should be coded as reparations. Examples of non-monetary reparations include verbal or formal acknowledgments or apologies of past government transgressions or formal pardons of previously convicted persons that are made with explicit reference to reparations.

4.12 Human Rights

General discussions of human rights and human rights violations. This context field is intended to capture human rights contexts that may lack details concerning specific human rights violations. As such, this field will apply in most instances where the term human rights appears, and especially so when human rights is discussed in an unspecified manner. Examples of such discussions would be articles outlining 1) the global human rights regime, 2) human rights organizations and their activities (e.g., Amnesty International), 3) international meetings/treaties/negotiations over human rights, or 4) an international organization or actor's efforts to combat human rights abuse in a general sense.

4.13 Rights Freedoms

Discussions of specific political rights and freedoms or their violation. This can include actual discussions of such freedoms or violations, as well as discussions of specific laws, monitoring organizations, or protests associated with these freedoms/violations. Included here should be freedom of (foreign and domestic) movement, freedom of speech, freedom of assembly, civil liberties, electoral self-determination, freedom of the press, freedom of religion, worker's rights, and women's rights. Note that this field will often receive joint codings with 'human_rights', 'gender', 'lbgt', 'asylum', 'migration', 'pro_democracy' and/or 'pro_autocracy' depending on the specific 'rights_freedoms' context.

4.14 Repression

The use of force by security forces to control individuals or groups. The latter individuals or groups will typically be unarmed and include (e.g.) individual citizens, social or ethnic groups, societies on the whole, businesses, journalists, or NGOs. The use of force against rebels should be excluded from 'repression' as it denotes civil war or civil conflict rather than repression. The use of force against alleged criminals, terrorists, gangs, or associated groups should only be coded as 'repression' when (i) the use of force is characterized as excessive or illegal and/or (ii) the criminal- or terrorist-identity of the recipient is disputed. Security forces (i.e., the actors initiating repression) are usually state-based (e.g., the army, police, or secret police) or state-affiliated (e.g., pro-government militias, paramilitaries, or death squads), but can occasional be comprised of non-state armed forces (if the latter forces control a particular territory or region). Specific examples of repression include disappearances, extrajudicial killings & punishment, political imprisonments, the use of excessive force, forced displacements, and torture. As such, stories receiving a 'repression' context coding will often also have a 'human_rights' and/or 'rights_freedoms' context coding.

4.15 Human Security

The protection of individual, group, and societal access to water, food, housing, energy, and land/property tenure. Discussions of 'human_security' along each of these dimensions should generally make a linkage to access; whereas (e.g.,) discussions of a drought without any mention of its human/social implications should generally only be coded as 'disaster'. Mentions of "food (in)security" or "water (in)security" should generally be coded as 'human_security'. Likewise, the 'human_security' context field should include not only instances where a direct threat to human security arises, but also discussions of (potential) vulnerabilities to such insecurities, and domestic or international efforts to address such vulnerabilities (e.g., the United Nations Development Programme).

4.16 Gender

Discussions of gender, women's issues, gender equality, women's rights, reproductive rights, and women in politics. Discussions of gender-based discrimination (or femicide), as well as laws developed both for and against gender-based discrimination, should also be coded. Discussions of gender-oriented organizations, groups, or protests should likewise be coded as 'gender.'

4.17 LGBT

Discussions of lesbian, gay, bisexual, transgender, queer, intersex or asexual groups, movements, or individuals. LGBTQIA discrimination and associated laws should be coded under 'lgbt'. Discussions of LGBTQIA-oriented organizations, advocacy, or protests should also be coded as 'lgbt.'

4.18 Religion Ethnicity

General or specific discussions of religions, ethnicities, castes, or linguistic groups. In addition to general discussions of religion, ethnicity, caste, or language, this can include discussions of discrimination (or violence) against particular ethnic, religious, caste, or linguistic groups; or instances of discrimination (or violence) perpetrated by particular ethnic, religious, caste, or linguistic groups. In addition, this context fields should also code more cooperative interactions among ethnic, religious, caste, or linguistic groups, as well as instances where such groups play a major role in politics, advocacy, reparations, or protest.

4.19 Inequality

Discussions of inequality in a general sense, as well as discussions of more specific forms of inequality such as economic inequality, gender inequality, racial inequality, and social inequality. The latter forms of inequality can encompass (e.g.) the uneven distribution of income or wealth in a society, policies that exclude certain groups from political participation, as well as unequal access to education, healthcare, or government services. Discussions of societal discrimination—such as gender-based, ethnic-based, religion-based, race-based, or age-based discrimination—should also be coded as 'inequality.' Accordingly, many 'inequality' stories will often also receive context codings for 'lgbt', 'gender', or 'religion_ethnicity.' Lastly, note that 'inequality' can include discussions of not only (within-country) societal inequality, but also global or cross-country inequality.

4.20 Diplomatic

Formal diplomatic relations, negotiations, and interactions between nation-states or between actors within particular nation-states. This includes bilateral diplomatic institutions and actors such as embassies, consulates, charge d'affaires and ambassadors, and the services that they provide. This context field also includes formal diplomatic negotiations between diplomatic representatives of nation-states and non-state actors (e.g., rebel organizations) over topics such as treaties, agreements, or alliances. Discussions of countries' ministries of foreign affairs (e.g., the State Department in the U.S.) or their ministers/secretaries should also be coded under this context field.

4.21 Territory

Discussions of territory in a political context. This can pertain to protests or rebellions over territorial autonomy, independence, or succession. This context field should also include discussions of territorial disputes (including border disputes) between nation-states as well as international treaties and/or mediation developed to address these disputes. Discussions of territorial displacement of certain groups by government and non-governmental actors should also be coded as 'territory', as should discussions of territorial sovereignty. However, general mentions of land or geography should not be coded as 'territory' if those mentions do not include a clear political dimension.

4.22 Peacekeeping

Activities involving the monitoring or maintaining of ongoing peace-processes, typically in the aftermath of an international or domestic armed conflict. Such activities most commonly involve international observers drawn from multilateral operations directed by (e.g.,) the United Nations, NATO, or European Union; but may at times also be drawn from individual nation-states. Peacekeeping can include the active stationing of international forces in a post-conflict area, such as is commonly the case with UN Peacekeepers (i.e., Blue Helmets), as well as multilateral non-military assistance for activities such as peace agreement implementation, post-conflict power-sharing, and confidence building measures.

4.23 Health

Issues and institutions related to public health, disease outbreaks and epidemics, disease in general. This context field should include both actual instances of health crises and disease outbreaks and efforts to improve public health and disease preparedness. Discussions of healthcare systems or key domestic and international public health institutions (e.g., the WHO) should be coded as 'health'. However, this context field should not be used for reports of individual accidents and accidental deaths, such as those arising from (e.g.) traffic accidents or skiing accidents.

4.24 Asylum

The granting or pursuit of national or political asylum. Such asylum is often pursued by, and provided to, political refugees. Discussions of asylum seekers, as well as national or international national asylum treaties or institutions, should also be coded under this context field. However, general discussions of refugees should not automatically be coded as 'asylum' unless some reference to asylum is made.

4.25 Migration

Discussions of international or domestic migration, as well as discussions of the direct drivers or consequences of migration. This context field accordingly includes discussions of immigration and emigration, forced displacement, internally-displaced persons, international organizations seeking to manage migration, and refugees. Discussions of migrant activities or challenges should also be coded under 'migration', and would include themes such as discrimination against migrant communities, migrant remittances, and policies or laws focusing upon (e.g.) migrant pathways to citizenship or immigration.

4.26 Disasters

Discussions of slow and rapid onset natural disasters such as droughts, hurricanes, cyclones, earthquakes, flooding, landslides, and heat waves. This context field should code news stories related not only to the disasters themselves, but also their effects on social outcomes such as migration. Accidents (e.g., chemical plant explosions) and oil or chemical spills should also be coded under this context field.

4.27 Natural Resource

Discussions of natural resource extraction, management, disputes, depletion, consumption, pricing, and trade. Environmental concerns over natural resources should also be coded as 'natural_resource', in addition to as 'environment'. Common types of natural resources are minerals, rare-earth metals, fossil fuels (e.g., oil, gas, and coal), agriculture, forests/lumber, diamonds, sand, and water.

4.28 Environment

Environmental problems & solutions, environmental politics, and environmental activism. Regarding environmental problems, this context field should code environmental disasters as well as broader environmental problems or concerns such as general mentions of pollution, climate change, water stress, deforestation, or soil degradation. Environmental solutions encompass policies and technologies designed to address environmental problems, such as renewable energies or the UN Framework Convention on Climate Change. Environmental politics encompasses any (ir)regular political processes associated with environmental policymaking, lobbying, negotiation, regulation, or cooperation at the international or domestic levels. Environmental activism includes not only environmental protest and direct actions, but also broader efforts towards the conservation or preservation of particular lands or animal species, as well as discussions of specific domestic and transnational environmental advocacy organizations.

4.29 Technology

Discussions of technological development, technological innovation, new technologies, or high-tech industry. Examples of high-tech industries include those based in aerospace, semiconductors, nuclear energy, electronics, and quantum computing. Note that this context field can thereby include military technologies, cyber security technologies, and misinformation technologies, meaning that

many articles with this context will likely receive multiple other context tags. National policies and international agreements designed to harmonize or foster technologies (or their future development) should also be coded as 'technology'.

4.30 Terrorism

Events related to terrorist groups, terrorist actions, terrorist funding, or terrorist sponsors; as well as anti-terrorism activities and policies. Terrorism is defined as any activity that aims to harm, kill, or kidnap civilians or unarmed government employees in an effort to intimidate, coerce, and/or provoke fear in societies, governments, or groups therein.

4.31 CBRN

Discussions of chemical, biological, radiological and nuclear security defense. This context field should code all discussions related to chemical, biological, radiological, or nuclear weapons or attacks, as well as discussions of security or preparedness for such attacks. Discussions of the development of such weapons or capabilities, or suspicions of such development, should be coded. International organizations or treaties designed to oversee or prevent such weapons or attacks should also be coded under this context field. However, general discussions of (e.g.,) nuclear energy or nuclear disasters should not be coded under this context field, and should instead be coded under the 'environment' and/or 'disaster' context fields depending on the circumstances.

4.32 Cyber

Discussions of cyber attacks, cyber security, cyber warfare and cyber crime. This context field should include any discussions of actual cyber crime/warfare events, as well as suspicions or allegations of those events. Discussions of laws, organizations, or initiatives designed to combat these threats (via cyber security or otherwise) should also be coded as 'cyber'. Cyber terrorism should be coded as 'cyber', but will also typically receive a 'terrorism' context coding. Electronic or information warfare will often receive a 'cyber' context tag, as well as a 'misinformation' context tag.

4.33 Misinformation

Discussions of misinformation, disinformation influence operations, or fake news; as well as national and international efforts to combat these tactics. Allegations or suspicions of misinformation and related forms of information operations/warfare should be coded as 'misinformation.' Note however that this context field pertains to news stories that discuss the use of (or activities related to) misinformation-type tactics, as opposed to whether or not the news story being coded is itself likely to be misinformation.

4.34 Crime

Discussions of (international or domestic) criminal activities, as well as discussions of the perpetration of specific crimes, criminal convictions, or criminals themselves. General discussions of transnational or domestic criminal organizations (e.g., the mafia) or high profile criminal fugitives

should also be coded under 'crime.' International coordination and cooperation to craft, enact, or enforce laws against criminal actions should also be coded under this context field. Note that several additional context fields—such as 'trafficking,' 'corruption,' and 'illegal_drugs'—code more specific criminal actions. War crimes should generally be excluded from this context field, and coded instead under 'human_rights' or 'rights_freedoms' depending on the specific context.

4.35 Corruption

Instances of political, legal, and business corruption. This can include allegations or suspicions of such corruption. Coded forms of corruption for this context field include (e.g.) misuse/abuse of public office for personal (private) gain, extortion, embezzlement, bribery, cronyism, nepotism, and insider trading. The passage, enforcement, or development of international or domestic anti-corruption laws should also be coded as 'corruption'. However, the abuse of government power via tactics of repression or police brutality should not be coded as 'corruption', and should instead be coded as 'human_rights' 'pro_autocracy,' 'rights_freedoms,' and/or 'elections' depending on the specific context.

4.36 Illegal Drugs

The criminal possession, distribution, sale, or manufacturing of illegal drugs. International meetings and treaties concerning efforts to combat illegal drug use, production, or transport should also be coded as 'illegal_drugs.' Discussions of criminal organizations with mention of their involvement in the drug trade or drug cartels should also be coded. Discussions or implementations of the legalization of illegal drugs by governments should also be coded as 'illegal_drugs' as should discussions of the use of drug profits for funding by rebel or terrorist organizations.

4.37 Trafficking

The domestic or international trafficking or smuggling of illicit goods. This includes international trafficking in (e.g.,) illegal drugs, diamonds and gemstones, gasoline, cigarettes, antiquities, lumber, humans, weapons, or wildlife; as well as the smuggling or sale of such goods across subnational political boundaries to evade specific taxes or regulations. In addition to active discussions of trafficking itself, any discussions of law enforcement efforts to disrupt trafficking should be coded as 'trafficking.' Similarly, discussions of any international laws, meetings, or organizations designed to combat trafficking should be coded as 'trafficking.' Finally, discussions of negative outcomes that are directly tied to trafficking within sending and receiving states (e.g., deforestation or increases in prostitution) should likewise be coded under this context field.

Chapter 5

Country Names and Sector Codes

CAMEO employed a hierarchical entity coding structure based on 3-character coding elements which allowed nearly unlimited complexity and, depending on the exact coding system, could be resolved down to the identity of individual groups or individuals. As with the event codes, typically only the first two or three of these elements were used. ICEWS modified this somewhat, while preserving most of the sub-state differentiations as "sectors"—the terminology we've adopted here over the CAMEO/IDEA "agents" terminology—but also provided a very substantial amount of complexity at the sub-sector level.

As with the events, PLOVER seeks to pare this down to the most commonly used countries and sectors, while retaining the possibility of more specific information. In place of the pages of entity and agent specifications found in the CAMEO manual, PLOVER has four rules:

1. The country name—for both an event's `actor` and `recipient`—is either a country designation for independent states, an entity designation for a very small number of non-independent territories, or a standardized identifier for international non-state actors such as intergovernmental organizations; see Table 5.1 below.

2. The sector code is a 3-character code that best describes an `actor/recipient` role. The primary role is provided in an initial field for each `actor/recipient`. Any other applicable sector codes are listed in a secondary field.

3. Identifiers for individual persons or organizations are coded in the `identifier_id` and/or `identifier_text` fields. For many instances, this takes the form of a Wikidata ID with its corresponding text matching the canonical English-language name of the Wikidata entry. Including this information greatly helps researchers who want to (1) track a particular organization or person, (2) need to make fine-grained distinctions that are currently subsumed within a single code (USA MIL = Army + Navy + Air Force + ... and USA GOV = the president + the State Department + the attorney general of Oklahoma + the city of Cambridge + ...) and (3) assign different sector codes using the raw information provided by Wikidata/Wikipedia (e.g., a new sector category for right-wing populist parties and politicians).

4. Depending on PLOVER implementation, the remaining named fields in the `actor/recipient` JSON object are used for additional information beyond what can be coded in the sector secondary modifier. In other words, one item of information—typically it is religion, ethnicity, or position—can be coded in the tertiary code, but only one: this handles virtually all of the current use-cases we know of such as religion, ethnicity, official position, etc.

By effectively crowdsourcing details to Wikipedia/Wikidata, the latter which is generally updated very quickly, often within minutes in the case of a death of a well-known figure, we deal with a major weak point in the CAMEO dictionary approach: the need to maintain coding teams to regularly update dictionaries, and in the absence of this, having information sometimes years out of date. Wikipedia/Wikidata also has a fairly standardized format for biographical information which contains far more detail than a CAMEO entry, but can be parsed with relatively simple programs.

An example `actor` block is below:

```
{
 "raw_text": "Steinmeier",
 "entity_name": "Frank-Walter Steinmeier",
 "wikidata_id": "Q76658",
 "wiki_description": "Minister of Foreign Affairs",
 "country_name": "Germany",
 "sector_code": "GOV",
 "other_sector_codes": ""
 }
```

While the process of generating this block will depend on the specific nature of the coder, in general, the coder will (1) identify an `actor` or `recipient` in the text, in this case, "Steinmeier". (2) Using the context of the article, resolve the mention to its Wikipedia page and ID and report the Wikipedia role or description. (3) Using the information on the page, determine that Frank-Walter Steinmeier was the German Minister of Foreign Affairs and thus GOV.

5.1 Country names

The `actor/recipient` country name fields identify the country or other non-state international actor with which an `actor` or `recipient` is associated. These can be identified with a full name (the default), or optionally abbreviated 3-letter codes, e.g. "Germany" or "DEU." Altogether, the `actor/recipient` country name fields consist of three groups of entities:

1. Independent states according to the Correlates of War (COW) state system membership list.[1] In these instances, reported names and short codes are based on the ISO 3166 English short names and 3-letter codes, not COW names and codes. Aside from the exceptions mentioned next, all dependencies such as the Åland Islands or the Pitcairn Islands are (re)assigned to the country name of their controlling state.

2. A small number of non-independent territories and dependencies, by exception: Antarctica, Hong Kong[2], and Palestine. Note that one US-recognized independent country, Kosovo, also has its own country name entry. It does not have an ISO code; we use "KOS".

3. A small number of international non-state entities, like the UN or NGOs, listed in Table 5.1.

The following subsections provide additional clarifications related to the way different countries and territories are handled.

[1] We include the Holy See in this group, even though it is not in the COW list.

[2] To obtain all events in China, one thus would need to select events both in "China" and "Hong Kong".

5.1.1 Non-state international actors

Table 5.1 presents the seven international non-state entities that are also included in the country names field. With the exception of the UN and EU, the special codes in Table 5.1 are intentionally broad and generic. The application of "international governmental organization" excludes the "United Nations" and "European Union." The latter two cases were separated from "international governmental organization" because they are (a) fairly common, and (b) the largest and arguably among the most institutionalized IGOs.

Table 5.1: International Non-state actor names

Code	Name
IGO	International Governmental Organization
UNO	United Nations
EUR	European Union
NGO	Non-governmental Organization
ISM	International Social Movement
IMG	Transnational Militarized Group
MNC	Multi-national Corporation

5.1.2 Non-independent territories with a separate ISO code

There are 59 dependencies and other non-independent or only partially independent entities that have an ISO 3166 code but are (1) not independent state actors in the COW state system membership list nor (2) appear on the exceptions listed under item 2 above. Examples include the Falkland Islands, Puerto Rico, and French Guiana. Those are folded into their associated controlling state, e.g., `actors/recipients` and events in the Falkland Islands are associated with the United Kingdom, etc.

Events arising in specific dependencies can still be recovered through the use of PLOVER's geolocation fields when those events contain a sufficiently precise geolocation.

5.1.3 Partially recognized (disputed) entities

There are several disputed territories which are not recognized by the US that receive special handling, like the Transnistria region in Moldovia. They are treated as follows:

- Instead of a separate country name and code, they are associated with the nominally sovereign country that claims control over that territory. As such, for example, Transnistria would be treated as part of Moldovia, not a separate country.
- However, they receive a special "PRE: partially recognized entity" *sector* tag, in addition to whatever other sector tags they would receive. Thus an event involving the Transnistrian breakaway government would receive as country name "Moldovia", but the sector tags would include in "PRE" in addition to (in this instance) the "GOV" code.

Presently the list of such cases (with the country they are associated with) is: Abkhazia (Georgia), Luhansk People's Republic (Ukraine), Donetsk People's Republic (Ukraine), Nagorno-Karabakh (Azerbaijan), Northern Cyprus (Cyprus), Transnistria (Moldovia), Somaliland (Somalia), South Ossetia (Georgia), and Western Sahara (Morocco).

5.2 Sector Codes

Sector codes are 3-letter codes that identify the broad societal/organizational sector that an `actor` or `recipient` is associated with. They provide a more fine-grained categorization of `actors/recipients` than just their country, and should thus be especially useful for analyzing patterns in domestic, not international, events.

Table 5.2: Sector Codes

Code	Frequently used codes
COP	Police forces, officers, criminal investigative units, protective agencies
GOV	Government: the executive, governing parties, coalitions partners, executive divisions
JUD	Judiciary: judges, courts
LEG	Legislature: parliaments, assemblies, lawmakers
MIL	Military: troops, soldiers, all state-military personnel/equipment
PRM	Paramilitary organizations not in opposition to government
PTY	Political parties (see Note 1)
REB	Rebels: armed opposition groups or individuals (see Note 3)
SPY	State intelligence services
UAF	Unidentified armed forces ("unknown gunmen")
UNK	Generic unidentified actors (e.g. actor for "two civilians were killed by a bomb")
	Less frequently used codes
AGR	formally or informally organized agricultural labor; peasants
BUS	business: individuals companies, and enterprises, not including MNCs
CRM	individual criminals and criminal gangs
CVL	civilians, protesters, activists, and other unarmed domestic entities that lack affiliation with the organized groups that appear within the sectors presented above and below
EDU	educators, schools, students, or organizations dealing with education
ELI	former government officials who currently do not hold positions in the government or armed forces, including exiled/ousted officials
JRN	journalists, newspapers, radio, television, web sites (see Note 4)
LAB	formally or informally organized labor in services or manufacturing
MED	individuals and organizations dealing with health (see Note 4)
REF	refugees and internally displaced persons
OPP	political opposition: opposition parties, individuals, anti-government activists; typically assigned as a secondary code
PRE	partially recognized entities, such as breakaway regions, partially recognized states, or disputed territories (see subsection 5.1.3 and Note 2)
REL	religious organizations and institutions
SOC	any organization or movement that is considered part of "civil society" not otherwise covered above

5.2.1 Sector code notes

1. For PTY, the use of additional sector labels such as GOV or OPP help to distinguish political parties in government or in opposition from more general references to political parties.

2. PRE is used to identify actors associated with partially recognized territories like Transnistria. It is added to whatever other sector codes and actor would receive, e.g. a mention of the Transnistrian breakaway government would receive the "GOV" and "PRE" codes, while proper Moldovian government would only receive the "GOV" code. See subsection 5.1.3.

3. For militarized groups, we drop the INS (insurgent) and SEP (separatist) distinctions incorporated into CAMEO during the research phase of ICEWS: these can be resolved on the basis of the group identity and group objectives are frequently ambiguous in any case.

4. Two modifications of CAMEO sector codes: in CAMEO 'MED' was "media" and 'HLH' was "medical" but no one could remember those.

Chapter 6

Data Fields

Note: This chapter is still under development and is not likely to be of interest to most readers. Not all items discussed within this chapter have been implemented.

In addition to providing a coding ontology, PLOVER is also intended to provide a standardized data exchange format using *named* data fields instead of the current system where the content of data fields is usually determined by *location* in some delimited format such as `.csv`. Standardizing these field names will simplify the merging and reuse of datasets, and such data are far easier for a human to read.

Despite the apparent complexity of the formats discussed here, note that the only required field we have added to "event data classic" is the `id` identifier, so the simplest form of an event record would look like

```
{
"id" : "PHOXv1-20160724-0042",
"date" : 2016-07-24,
"actor" : [{"code":"USA"}],
"recipient" : [{"code":"CAN"}],
"event" : ["CONSULT"]
}
```

For ease of parsing and use, we suggest formatting PLOVER in newline-delimited JSON (JSONL) format, with each event formatted as one valid JSON entry, each on a separate row.

Except in the small number of cases where a standard format is specified, the content of the field is left open, and in particular "number" should be interpreted as "number or code": for example instead of providing the number of individuals killed, a dataset might use a set of categories giving ranges. Similarly, fields such as `context` can take multiple values: typically these would be formatted using a JSON "array" structure—which is to say, a list—but responsibility for handling these details is left to the data provider and users. Providers should feel free to include named fields beyond those provided here but if a data set codes or extracts information corresponding to one of the existing fields, please use that name.

Table 6.1: PLOVER JSON

Name	Content	Note	Required?
id	unique identifier	1	Y
has_event	event has been coded (True/False)	5	N
date	date in YYYY-MM-DD format		Y
time	ISO 8601-formatted time	2	N
enddate	date in YYYY-MM-DD format		N
endtime	ISO 8601-formatted time	2	N
actor	list of entity objects		Y
recipient	list of entity objects		N
event	list of event categories		Y
event_loc	location object for event		N
event_text	list of texts of event		N
quad_code	1, 2, 3 or 4		N
event_scale	floating point scale value		N
mode	list of modes	3	N
context	list of contexts	3	N
link	link identifier	4	N
text	text from which the record was coded	6	N
text_info	textInfo object for text		N
cite_info	citeInfo object for text		N
coder	coder identification		N
coded_date	date of coding		N
coded_time	time of coding in ISO 8601-formatted time	2	N
comment	any text		N

Notes:

1. The identifier should be unique within the data set; it is the responsibility of the user to reconcile identifiers across data sets

2. ISO 8601 allows a number of different formats for times depending on the level of detail. Formatting should be such that a string of the form `date + 'T' + time` should yield an ISO-8601 datetime.

3. `event`, `mode` and `context` fields can have multiple entries; they do not need to resolve to a single value, and in fact this is likely to occur fairly frequently in classifier-based systems which work with the general sense of a sentence, in contrast to dictionary-based systems which look for specific sets of words. Multiple event categories would be used in a single record if the source and recipient actors are the same; they would resolve to multiple records if the source and recipient actors are different, as might occur in a compound sentence.

4. This can be used to create a common reference across multiple related events, such as demonstrations in multiple locations organized by the same group.

5. This is typically set to False when the record is part of a pre-processing pipeline

6. This slot will only be filled when the creator of the record has appropriate intellectual property rights for the text: this tends to be the exception rather than the rule

Table 6.2: Information object for entities

Name	Content
code	3-char top-level entity code (e.g., country)
sector	3- or 6-char sector (GOV, MIL, etc)
entity_text	extracted text for source
identifier_id	unique identifier ID for source [see Note 1]
identifier_text	unique identifier name for source [see Note 1]
religion	religion (code or text)
ethnicity	ethnicity (code or text)
office	office or official position (code or text)
gender	gender (code or text)
age	integer

Notes:

1. These fields would be used to resolve the name of an entity that occurs in multiple forms—for example "Islamic State", "IS", "ISIS", "Daesh"—into a single form or code. This should be the Wikidata ID for the entity. For example, the Islamic State's is https://www.wikidata.org/wiki/Q2429253 and its canonical Wikidata name is "Islamic State of Iraq and the Levant".

Table 6.3: Information object for text

Name	Content
sequence	sequence number of sentence
start	character offset for start of text
end	character offset for end of text
text_story	list of sentences from full story text

Table 6.4: Information object for size

Name	Content
dead	number killed
injured	number injured
arrested	number arrested

Notes:

1. These fields are included as standard names because they are most likely to be used in event systems, but users should feel free to add additional fields for numbers that are not related to location.

Table 6.5: Information object for citations

Name	Content
corpus	name or other identifying information
citation	bibliographic citation or database identifier for text
url	URL for text
title	title for text
language	language of text (ISO 639-1 two-letter codes)
publication	name of text publisher
license	license covering text
copyright	copyright covering text
coder	identifying information for any event extraction system used
codebook	reference for the codebook used to code the text, e.g. `plover-base-1.3.1` or `plover-protest-0.3`
version	version of data set

Table 6.6: Location object: location information returned by Mordecai3, drawing from the Geonames database.

Name	Content
extracted_place	original place name extracted from the text
resolved_place	name of geographical location the event was resolved to (unicode). The rest of the fields provide more information on the returned location
geonameid	integer id of record in geonames database
lat	latitude in decimal degrees
lon	longitude in decimal degrees
city	name of the city
district	name of the admin 2 (district/county) level
province	name of the admin 1 (province/governorate/state) level
country	name of the country
country_code	country code (3 character ISO code, to match the actor/recipient countries
feature_class	high-level code, such as A for area or P for populated place. see http://www.geonames.org/export/codes.html
feature_code	detailed feature type, such as PPLX for neighborhood, ADM2 for district, etc. See http://www.geonames.org/export/codes.html
resolution	code indicating the level of resolution. 5=sub-city, 4=city, 3=district, 2=province, 1=country

6.1 Adding to PLOVER: protest example

OEDA was founded on the principle that there should not be "one data set to rule them all": different implementations will have different strengths. As an example, a protest-specific coder could add more fields to the event record for things like the participant size (a numeric amount or size category), the number of people who were injured, the number of people arrested, etc. This section briefly outlines how PLOVER could be extended to code specific event types in greater detail. A protest-optimized coder could also include protest-specific contexts like the ones in Table 6.7.

Table 6.7: PROTEST contexts

Name	Content
election	elections
political	political and constitutional reforms
economic	economy, jobs
food	food, water, subsistence
env-disaster	environmental issues, disasters incl. earthquakes, floods, fires
discrimination	ethnic discrimination, ethnic issues
religion	religious discrimination, religious issues
education	education
foreign	foreign affairs/relations
war	domestic war, violence, terrorism
rights	human rights, democracy
pro-govt	pro-government
independence	independence or separatist movements

Adapted from Salehyan and Hendix, *Social Conflict Analysis Database* (SCAD) Version 3.2: `https://www.strausscenter.org/images/codebooks/SCAD_32_Codebook.pdf`

Chapter 7

CAMEO vs. PLOVER

As noted in the introduction, CAMEO was originally developed for academic research under U.S. National Science Foundation funding in the early 2000s, and was based on the WEIS system. The canonical citation for CAMEO is Schrodt et al. (2009), and the detailed manual, ca. 2012, is found at `http://eventdata.parusanalytics.com/data.dir/cameo.html`. The CAMEO event framework was very much the work of Deborah Gerner and Ömür Yilmaz, with contributions by various coders in the Kansas Event Data System project; the enitity framework was strongly influenced by the VRA "IDEA" coding system developed in the late 1990s (Bond et al., 2003). The CAMEO manual contains an extended discussion of the issues considered in transitioning from WEIS to CAMEO. Additional details on the development of the automated coding underlying CAMEO can be found in Schrodt (2006) or `http://eventdata.parusanalytics.com/utilities.dir/KEDS.History.0611.pdf`.

Considerable additional work on CAMEO was done in the early 2010s first in the context of the DARPA ICEWS research program, then later in the operational deployment of ICEWS by teams at BBN and Lockheed which was eventually incorporated into the Dataverse public data: details of this work on found in the internal documentation of that data.

7.1 Summary of changes

- A set of standardized names ("fields") for JSON (`http://www.json.org/`) records are specified for both the core event data fields and for extended information such as geolocation and extracted texts; most of these fields are optional and where available we use existing specifications, for example the `http://geonames.org` geographical location field names, ISO-3166 country identifiers and ISO-8601 date and time formats.
- Only the 2-digit event "cue categories" have been retained from CAMEO.
- The details in the 3- and 4-digit categories are now delegated to the optional `mode` and `context` fields: see Section 2.1 for further discussion of this.
- A set of scaled "PLOVER scores" has been systematically derived from the "Goldstein scores" found in the ICEWS data set.
- The CAMEO 01 and 02 categories dealing with comments have been eliminated.[1]

[1] Ironically, this reverses a decision McClelland belatedly made—and later regretted—in the WEIS specification in the 1960s.

- The CAMEO 08 "YIELD" category has been split into verbal (CONCEDE) and material (RETREAT) components.
- The "target actor" event component was renamed `recipient` for clarity and to better match the terminology used in the NLP literature on event extraction (Halterman, 2020). "Source actor" was renamed `actor` to reduce confusion with the textual source of the event.
- `Actors/Recipient` entries are now standardized using references to Wikipedia.
- The complexity of substate codes has been limited, and the allowable substate modifiers have been substantially simplified.
- Standard optional fields have been defined for some categories, and `recipient` is optional in some categories.

7.2 CAMEO to PLOVER translation

Table 7.1: PLOVER equivalents to CAMEO cue categories

CAMEO code	CAMEO text	PLOVER category
01	MAKE PUBLIC STATEMENT	dropped
02	APPEAL	dropped
03	EXPRESS INTENT TO COOPERATE	AGREE
04	CONSULT	CONSULT
05	ENGAGE IN DIPLOMATIC COOPERATION	SUPPORT
06	ENGAGE IN MATERIAL COOPERATION	COOPERATE
07	PROVIDE AID	AID
08	YIELD (081 to 083)	CONCEDE
08	YIELD (084 to 087)	RETREAT
09	INVESTIGATE	ACCUSE
10	DEMAND	REQUEST
11	DISAPPROVE	ACCUSE
12	REJECT	REJECT
13	THREATEN	THREATEN
14	PROTEST	PROTEST
15	EXHIBIT FORCE POSTURE	MOBILIZE
16	REDUCE RELATIONS	SANCTION
17	COERCE	COERCE
18	ASSAULT	ASSAULT
19	FIGHT	ASSAULT
20	USE UNCONVENTIONAL MASS VIOLENCE	FIGHT (see Note 1)

Notes:

1. For unconventional weapons, the `mode` in the FIGHT record would be set to `unconventional`.
2. Generally, everything at the 3- and 4-digit level should simply be reduced to the 2-digit cue category and converted accordingly. Depending on your specific application, you might

want to make some exceptions to this—for example a CAMEO "015: Acknowledge or claim responsibility" might be considered AGREE and a CAMEO "016: Deny responsibility" might be considered REJECT—but we are not making general recommendations on this. Except to suggest that for the benefit of those trying to replicate your work, you carefully document any such decisions.

7.3 Summary of event categories

Event Type	Quad	Modes	Requires Recipient	Other Fields
AGREE	V-Coop.	Yes		
CONSULT				
SUPPORT				
CONCEDE				
COOPERATE	M-Coop.		Yes	
AID		Yes		
RETREAT		Yes		
REQUEST	V-Conf.	Yes		
ACCUSE		Yes		
REJECT		Yes		
THREATEN		Yes		
PROTEST	M-Conf.	Yes	Yes	event_loc
SANCTION		Yes	Yes	
MOBILIZE		Yes		
COERCE		Yes		
ASSAULT		Yes	Only if one-sided or SideAB, otherwise all participants in source	event_loc, dead, injured, size

Table 7.2: Quad categories in PLOVER

Quad category	PLOVER categories	Numeric
Verbal cooperation	AGREE, CONSULT, SUPPORT, CONCEDE	1
Material cooperation	COOPERATE, AID, RETREAT	2
Verbal conflict	REQUEST, ACCUSE, REJECT, SANCTION, THREATEN	3
Material conflict	PROTEST, MOBILIZE, COERCE, ASSAULT	4

7.4 A note on CRIME

A separate event category for crime has been added and removed several times as PLOVER was being drafted. While criminal activity is important to capture in event data (Osorio, 2015; Osorio and Reyes, 2017), we have decided to not include a separate event category for it for several reasons:

1. Whether activity is criminal or not often depends on on the identity of the actor: actions undertaken by rebel groups may not fit within a definition of CRIME, while the same action taken by a drug cartel might. We have generally tried to avoid relying on the identities of actors in order to define events, due in part to the implementation of past coders, which did not use actor information to code event types.

2. Crime overlaps with other event categories, especially ASSAULT, which would make it difficult to train a CRIME classifier that did not pick up events that better belong in other categories.

That said, PLOVER still includes mechanisms for crime-type events to be coded. Researchers who are interested in criminal behavior have two primary options for locating it in PLOVER events:

1. Identify events taken by criminal actors. As coders move away from hand-constructed dictionaries to resolve actors, many more groups will be coded. Researchers will be able to subset events to those undertaken by specific criminal groups (e.g. the Sinaloa Cartel) or by using the CRM actor code.

2. Use the **crime** and **illegal_drugs** contexts: see Table 2.1.

7.5 Some residual issues

In the discussions leading to the development of PLOVER, several additional open issues were raised that we have decided to remain agnostic on are listed below. These issues have not yet been fully implemented in most coding projects employing PLOVER to date.

Temporal markup: This is emerging as a major issue in event extraction, particular among users who are interested in the long-standing objective of automated chronology generators. While there are some significant efforts on this in the NLP community—`http://www.timeml.org/`—we don't feel we currently have the experience required to make recommendations.

De-duplication: There is no consensus on this beyond noting that the widely-used "one-a-day filtering" is controversial, and it is a topic where there is currently active research and experimentation, so we're leaving it alone.

Required entities: PLOVER, in contrast to CAMEO, makes the `recipient` optional for some event types. One outstanding question is whether the `actor` should also be optional for some event types. Some event types often leave the `actor` implicit, for example, "4 people were arrested/killed in a suicide bombing etc." These have no explicit, named `actor` so they will not be coded by PLOVER. Similarly, many natural disasters do not fit neatly into an entity-centric approach to coding ("mudslides destroyed dozens of houses"). We could consider relaxing this requirement to increase our recall, but at the potential cost of more false positives and greater conceptual complexity.

Bibliography

Azar, E. E.
1980. The conflict and peace data bank (COPDAB) project. *Journal of Conflict Resolution*, 24:143–152.

Azar, E. E.
1982. *The Codebook of the Conflict and Peace Data Bank (COPDAB)*. College Park, MD: Center for International Development, University of Maryland.

Azar, E. E. and T. Sloan
1975. *Dimensions of Interaction*. Pittsburgh: University Center for International Studies, University of Pittsburgh.

Bond, D., J. Bond, C. Oh, J. C. Jenkins, and C. L. Taylor
2003. Integrated data for events analysis (IDEA): An event typology for automated events data development. *Journal of Peace Research*, 40(6):733–745.

Goldstein, J. S.
1992. A conflict-cooperation scale for WEIS events data. *Journal of Conflict Resolution*, 36:369–385.

Halterman, A.
2020. *Extracting Political Events from Text Using Syntax and Semantics*. PhD thesis, MIT Political Science.

Leng, R. J.
1987. *Behavioral Correlates of War, 1816-1975. (ICPSR 8606)*. Ann Arbor: Inter-University Consortium for Political and Social Research.

McClelland, C. A.
1967. World-event-interaction-survey: A research project on the theory and measurement of international interaction and transaction. University of Southern California.

McClelland, C. A.
1976. *World Event/Interaction Survey Codebook (ICPSR 5211)*. Ann Arbor: Inter-University Consortium for Political and Social Research.

O'Brien, S. P.
2010. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104.

Osorio, J.
2015. The contagion of drug violence: spatiotemporal dynamics of the mexican war on drugs. *Journal of Conflict Resolution*, 59(8):1403–1432.

Osorio, J. and A. Reyes
2017. Supervised event coding from text written in spanish: Introducing Eventus ID. *Social Science Computer Review*, 35(3):406–416.

Schrodt, P. A.
2006. Twenty years of the Kansas event data system project. *The Political Methodologist*, 14(1):2–8.

Schrodt, P. A., D. J. Gerner, and Ö. Yilmaz
2009. Conflict and mediation event observations (CAMEO): An event data framework for a post Cold War world. In *International Conflict Mediation: New Approaches and Findings*, J. Bercovitch and S. Gartner, eds. New York: Routledge.