

Crowd Counting Consortium

Coding Guidelines 2.0 (2021-2024)

Introduction

The [Crowd Counting Consortium](#) (CCC) encodes [event data](#) on contentious politics in the United States from public information about political gatherings, including but not limited to protests, marches, demonstrations, rallies, sit-ins, die-ins, sickouts, walkouts, caravans, and nonviolent direct actions.¹ For convenience, we often use “protest” as shorthand for the type of events the project covers, but, as the preceding list implies, the scope is broader.

More specifically, CCC follows the [Dynamics of Collective Action](#) project in defining its events of interest as “any type of activity that...is carried out with the explicit purpose of articulating a grievance against a [political] target, or expressing support of a [political] target.” That is a wide net to cast, so we use a few additional criteria to narrow the scope further.

- Following Soule and Davenport (2009: 8), we limit our collection to events that are **open to the public and free of charge**.
- We also limit our work to events that are **nonviolent** in the sense that they are not primarily organized to cause direct harm to any persons. Per Chenoweth (2021: 35–37), in the study of contentious collective action, “nonviolent” does not necessarily mean free of conflict, coercion, or property damage. What we intend to exclude with this criterion are military or paramilitary actions in which attacking or otherwise harming other people is the focal tactic.
- Finally, we only encode data on events that **occur in the United States**, including the U.S. territories of Puerto Rico, Guam, and the U.S. Virgin Islands. The project has occasionally encoded data on events in other countries, but its core mission is limited to making structured data about contentious political events in the United States, and events outside that country are not included in the [compiled and augmented version](#) of the CCC data.

Together, these criteria include all of the aforementioned forms of action while excluding things like paid fundraisers, workshops, trainings, teach-ins, panel discussions, distributions of mutual aid, petitions, commemorative or celebratory parades, concerts, cookouts, court support, block parties, town hall meetings, and signature-collecting or canvassing activities. Claims voiced during comment sessions at meetings of government bodies (e.g., legislative sessions, city

¹ Until April 2023, we also included labor strikes and pickets in our dataset. See the Release Notes section for a brief explanation of why we stopped doing that and how to account for this change in the scope of our collection in gross analyses of the data.

council meetings, school board meetings, court proceedings) are also excluded unless the people making those claims deliberately and successfully disrupted or derailed the larger meeting. Rallies associated with campaigns for public office are included as long as they are open to the public free of charge (i.e., not paid fundraisers). Vigils, memorials, parades, and commemorative gatherings are only included if participants also articulate grievances around broader political issues (e.g., community violence, police brutality, domestic violence, or gun control).

We also choose to define “gathering” broadly and include online events in our collection. Virtual gatherings have become an increasingly common and important form of collective action, and that trend only accelerated in 2020–2021 as a result of the COVID-19 pandemic. In light of this fact, we encode data about virtual versions of events that also meet the criteria enumerated above.

Finally, there is no minimum crowd size for inclusion in the CCC dataset. We record events involving just one or a few participants, as long as they satisfy the criteria enumerated above.

Sources

CCC uses a broad and always-evolving set of sources to find public information about relevant events, including online print and broadcast media sources, social media, and public submissions via an [online form](#). On a day-to-day basis, the research process proceeds along four main channels.

- **Automated web crawl of online print and broadcast media sources.** From 2017 through early 2021, the [Count Love](#) project shared the output of its web crawler with CCC via daily emails. Since early 2021, CCC has run its own instance of that crawler using Count Love’s open-source code. The daily output from the crawler is now stored in an Amazon Web Services bucket and written to a Google Sheet that refreshes daily.
- **Regular review of selected social media accounts.** CCC uses a [Twitter list](#) and Instagram to track accounts of activist organizations and independent journalists who organize or cover relevant events. These feeds are reviewed throughout the day for relevant information. Since late 2021, CCC has also used Telegram to identify and research events, and in November 2022, we began using Mastodon in a similar manner.
- **Google searches on relevant keywords.** CCC coders also use Google searches on relevant keywords to uncover stories the web crawler may have missed. These searches are typically run at least once daily, usually in the morning. Only the News tab is examined, and results are sorted by date and limited to the past 24 hours. The terms “protest” and “rally” are routinely searched, and additional terms such as “vigil,” “march,” and “demonstration” are also searched when time permits.

- **Irregular review of organization websites.** On an ad hoc basis, CCC collects information directly from the websites of organizations hosting large, multi-location protests or maintaining public online calendars or lists of past or upcoming events.
- **Public submissions.** CCC also solicits submissions and corrections from the public via an [online form](#). Responses are reviewed and encoded on an occasional basis. Events identified through this process comprise a tiny fraction of our data set.

Data-Making Process

Coders record information about reported events in [Google Sheets](#) that cover individual calendar months. The sheets are all visible to the public, but CCC only provides edit access to working coders as needed. A compiled and augmented version of the data is also made available to the public via a [GitHub repository](#) maintained by the [Nonviolent Action Lab](#).

Each Google Sheet has two tabs. The “Summary” sheet shows running totals of events and crowd sizes and provides background information on the project. This sheet is not used by coders. The “Tally” sheet is where data is actively entered and edited using the rules described in the Data Fields section below.

In the CCC dataset, events, not news reports, are the units of observation. This means that, in the “Tally” sheet, each row represents a single event, defined by the gathering of a particular set of people to articulate a specific set of grievances in a particular place on a certain date. When multiple reports describe the same event, the relevant information is amalgamated into a single record (row). Rows are ordered by date, then alphabetically by locality (city or town) and then state (postal code). When entering data, coders should assume that each record (row) needs to stand on its own and therefore should never reference other rows (e.g., “see above”) in their entries.

Collection is an ongoing endeavor, and coders regularly make additions or revisions to existing records as new information surfaces. This process places twin burdens on coders.

- Coders are asked to take care to avoid duplicating events that are already represented in the data. This means carefully checking existing records for events that occurred in the same locality on the same date before starting a new one.
- When multiple reports exist for the same event, coders have to decide how to synthesize that information. We generally resolve that tension by recording the relevant bits from all of the accounts, using semicolons within fields to make it easy for other researchers to distinguish those bits and make their own choices about how to synthesize them or otherwise handle that uncertainty.
- Coders are asked to capture all relevant links, even if the link provides no additional information. [NB. This clarification was added in January 2023.]

Data Fields

For each event, coders should fill in as many of the following fields as possible, using information from all relevant sources. When in doubt about what to record from a particular source or how to record it, please consult with the project's supervisor(s).

date

Record the date on which the event occurred, in YYYY-MM-DD format.

In the Google Sheets, this column uses data validation to reduce the risk of invalid responses. If you make an invalid entry, the sheet will notify you. When this happens, please take a moment to correct the entry before moving on. NOTE: data validation does not correct for typos in elements of the date (e.g., 2201-07-21 instead of 2021-07-21, or 2020-07-21 for 2021-07-21), so please take care when entering this element.

Some events span more than one day. For events that recur daily or that span multiple days for other reasons (e.g., many strikes and occupations), create a separate record for each day on which the event occurred or continued. The exception to this rule are events that begin one day and continue into the early-morning hours of the next day before wrapping up. In these cases, just create a single record and associate it with the date on which the event began.

locality

Record the name of the city, town, or township in which the event occurred.

- Take time to check for correct spelling. When in doubt, confirm that the spelling you use is recognized by Google Maps.
- Avoid abbreviations. For example, enter "Saint Louis" rather than "St. Louis".
- In cases where a single event began in one locality and traveled to one or more others (e.g., a march or caravan), do not create a separate record for each locality. Instead, choose one locality—usually the start or end point—and enter it here, then enter the names of the other towns in the 'location' field (see below).
- For events occurring in any of the five boroughs of New York City (Manhattan, Brooklyn, the Bronx, Queens, and Staten Island), enter "New York" as the locality, not the borough name. The borough is captured in the 'location' field (see below).
- For named neighborhoods of other cities, use the city name as the locality and, if it is given, capture the neighborhood name in the 'location' field. For example, an event described as occurring in "Roxbury's Nubian Square" would get "Boston" as its locality, and the location might be recorded as "Nubian Square, Roxbury". When in doubt about whether something is a neighborhood or a city, use Google Maps to check.
- In rare cases where the locality is unclear (e.g., "on the Merrimack River" with no other information about event location), leave this field blank.
- For online events, see the guidance specific to that type of event below.

state

Record the two-character postal code for the state or territory in which the locality named above is located.

In the Google Sheets, this column uses data validation to ensure that only valid responses may be entered. If you enter an invalid character pair, the sheet will show you a notification, and nothing will be recorded. Please try again.

If you are not certain of the correct postal code, please look it up. Some of them are easy to confuse (e.g., AR/AK for Arkansas/Alaska, MA/ME for Massachusetts/Maine).

location

Record the location or locations at which the event occurred within the aforementioned locality.

What we're looking for here is information you could use to pinpoint on a map the location(s) of the event within the named locality. Ideally, this is a landmark, address, or intersection you could enter in Google Maps and get a clear result.

- If a specific address is given, record it.
- If a landmark is given instead (e.g., "Black Lives Matter Plaza" or "Cal Anderson Park"), record it, ideally in the format recognized or returned by Google Maps (e.g., "on the capitol steps" in an article about a protest in Boise, Idaho, should be recorded as "Idaho State Capitol").
- If an intersection is given (e.g., "Hollywood and Vine"), record it, ideally in the format recognized or returned by Google Maps (e.g., "Hollywood Blvd and Vine St").
- When none of the above are reported, record whatever other information is given (e.g., "home of Mayor Bill Peduto" or "on the Merrimack River").
- For events occurring in New York City, always include the name of the borough as well as the landmark or address within it (e.g., "633 3rd St, Manhattan" or "Grand Army Plaza, Brooklyn"). When in doubt about the borough, please look it up.
- When events move from location to location, record as many as you can recognize, separated by semicolons. For example, a march that traveled from Foley Square in Manhattan over the Brooklyn Bridge to Barclay's Center in Brooklyn would get "Foley Square, Manhattan; Brooklyn Bridge; Barclay's Center, Brooklyn" as its location. Meanwhile, an event that started in Silver Spring, MD, and traveled to the White House in Washington, DC, might get "Washington" as its locality and then "Silver Spring, MD; The White House" as its location.
- In all cases, you are encouraged to use Google Maps to help clarify or confirm the location and how to record it. As often as possible, we want the entries in this field to be machine readable, meaning they could be fed as is into Google Maps and we could expect to get a specific set of geocoordinates back.

title

Record the name of the event, if one is given.

On flyers and Facebook event pages, the title is usually easy to distinguish. When it isn't and several phrasings appear, make an educated guess about which part is the title, and err on the side of including too much information rather than too little. With Facebook events and many other calendar listings, the title can often be copied and pasted. To avoid carrying undesired formatting into the Google Sheet, you can copy the title, paste it into a simple text editor (e.g., Notepad or TextEdit), copy it again, and then paste it into the Sheet.

size_text

Record the bit(s) of text describing the number of people who participated in the event.

Note that we do not record “expected” or “predicted” turnout here or in the ‘size_low’ and ‘size_high’ fields that follow. If the event has not yet happened, leave this cell blank. Instead, we wait to see how many people actually show up according to public reports.

- If an exact number is reported (e.g., “37”, “two”), record it.
- When the crowd size is described in vague words—e.g., “a few dozen” or “hundreds”—record those words here. Likewise for combinations of words and numbers (e.g., “more than 20” or “nearly 1,000”). If the source uses words that can be converted to numbers in these phrases, please do the conversion (e.g., “more than a hundred” should be recorded as “more than 100”, but “hundreds” should be recorded as a word).
- If a range is given, record it using “xx-xx” format (e.g., “30-40” and “30 to 40” should both be recorded here as 30-40).
- When different sources provide different estimates, include them all, separating them with semicolons (e.g., “dozens; a few dozen; more than 50”). The order does not matter.
- When a single source gives multiple estimates (e.g., “a few dozen to start...crowd grew to more than 100”), use only the largest one.
- When no sources describe the crowd size but there are photos or videos available, use that imagery to try to estimate the crowd size. If you do this, record either “count pic” or “count video” in this field, depending on which one you (primarily) used. Do not try to estimate crowd size from a video or a picture of a very large event (roughly thousands or higher).
- When no sources describe the crowd size and pictures and video are unavailable or uninformative but the event has a Facebook Event page, check that page after the event for a count of people who “went” to it. If there is one, record it here as FB: Number. For example, if the Facebook event page says 37 people went, put FB: 37 in this column.
- If a coder witnesses an event or receives information about event size from a credible source and no other credible information about crowd size is available, record “eyewitness” in this field.

size_low

Record the lowest reported estimate of the crowd's size, using the following rules to convert words into numbers. If only a single, specific number is reported, record it here and in 'size_high'. Do not include commas in the numbers (so, 1000 rather than 1,000).

- With fuzzy plural terms, assume the lowest possible multiple. So, for example, “dozens” becomes 24, “hundreds” becomes 200, and “thousands” becomes 2000.
- Interpret “several” and “a few” as 3. So, for example, “a few hundred” and “several hundred” would both become 300.
- If a range is reported, record the low end of it here.
- If you must estimate the crowd size from photos or video, you are encouraged to recognize that your estimate will be imprecise and guess a range rather than a single number. If you do this, record the low end of that range here.
- When an exact count is given or a Facebook Event page count is used, record it here.

size_high

Record the highest reported estimate of the crowd's size, using the following rules to convert words into numbers. If only a single, specific number is reported, record it here and in 'size_low'. Do not include commas in the numbers (so, 1000 rather than 1,000).

- With fuzzy plural terms, assume the lowest possible multiple. So, for example, “dozens” becomes 24, “hundreds” becomes 200, and “thousands” becomes 2000.
- Interpret “several” and “a few” as 3. So, for example, “a few hundred” and “several hundred” would both become 300.
- If a range is reported, record the high end of it here.
- If you must estimate the crowd size from photos or video, you are encouraged to recognize that your estimate will be imprecise and guess a range rather than a single number. If you do this, record the high end of that range here.
- When an exact count is given or a Facebook Event page count is used, record it here.

organizations

Record the names of all organizations mentioned in your sources as having participated in the event in any way, separated by semicolons.

Organizational participation can take many forms, from organizing and leading the event to sponsoring or co-sponsoring it to providing one or more speakers for it to just showing up to the event as a recognizable presence (e.g., with a banner or in a uniform). Err on the side of inclusion. If a reporter quotes a protester and associates them with a named organization, include that organization's name. The goal is to facilitate research into connections and interactions between organizations, so we want to capture as much of this information as possible.

When no organizations are mentioned as organizing or participating in the event, leave this field blank.

participants

Record words or phrases describing the participants in the event, separated by semicolons (e.g., “employees” or “parents; teachers; students”).

As with the ‘organizations’ field, err on the side of inclusion. The goal is to capture as much information as we can about the kinds of people who participated, as distinct from any organizations they represent or to which they belong.

When no such descriptions are given or when they are uninformative (e.g., “protesters”), leave this field blank.

claims

Summarize the political grievances or demands of event organizers and participants. Use commas to separate claims.

This field is very important, so take care when recording it. Think of this field as a space to describe what the protest is about, or *why* participants are protesting. Try to paint as vivid and complete a picture as you can in as efficient a manner as possible. This helps casual users of the data understand what the event was about, and it gives more advanced users of the data richer material to draw connections or contrasts across events. Recognizing that this is a bit of an art form, here are some general pointers.

- **Start by summarizing the participants’ stated or implied purpose of the event in one or a few short phrases** that begin with “for” or “against” or, where relevant, “in remembrance of”, “in celebration of”, “in commemoration of”, or “in solidarity with”. For example, instead of writing “Trump rally”, write something like “for Donald Trump”. Or, instead of writing “stop Line 3”, write something like “against Line 3 oil pipeline”.
- Once you have provided that high-level summary, **capture verbatim as many claims made by organizers and participants as you can**, including phrases from signs, banners, and chants evident in photos or videos.
 - Because claims are comma-separated, omit commas within verbatim claims.
 - Where verbatim claims include line breaks, use “ / ” to represent the break.
 - When recording chants, end the claim with an exclamation point (e.g., “stop fossil fuels!”). To avoid confusing slogans from placards and banners with chants, do not include exclamation points at the end of verbatim captures of those (e.g., a sign reading “climate action now!” should be recorded as “climate action now”).

- If participants use a common slogan that links the event to a wider movement (e.g., “Black Lives Matter” or “stop the steal”), be sure to include that slogan in the summary, too.
- In cases where images are an important part of a sign or banner, put a word or phrase describing the image in brackets (e.g., “never again [coat hanger]” or “we need to talk about the elephant in the womb [GOP]” at a rally for abortion access).
- If pictures or videos show signs or other iconography that convey or imply specific claims or grievances, incorporate those into your summary as well, even if those issues were not raised verbally by event organizers or leaders. For example, if numerous participants in a 2021 protest against a county coronavirus mask mandate are carrying American flags and flags that read “Trump 2020,” you might include “for Donald Trump, for patriotism” in your summary of the claims along with something like “against county coronavirus mask mandate”.
- When protesters carry flags, list them in the summary as well, using the structure “_____ flag”, where the blank is filled with the name or descriptor of the flag, e.g., “American flag”, “Gadsden flag”, “LGBTQ+ Pride flag”, “Pine Tree flag”, “Mexican flag”. [NB. This clarification was added in January 2023.]
- **When relevant, make connections to broader issues explicit in your summary.** For example, protests against the Line 3 project often involve concerns about water pollution and the rights of indigenous peoples whose traditional lands the project crosses. So, instead of just writing “against Line 3 oil pipeline” for events where that’s the case, write something like “against Line 3 oil pipeline, for environmental protection, for Indigenous peoples’ rights” if those other issues are evident in protesters’ claims as well. Ditto for protests in response to police shootings of Black people, where, instead of just writing “against police shooting of [name]”, you might write something like “for justice for [name], against police brutality, against racism” (and include “Black Lives Matter” as well if protesters specifically used the phrase).
- **When crafting this summary, assume that none of the other fields are visible to the person reading it.** So, for example, if the event involves teachers picketing for higher pay at Long Beach School District, don’t write something like “for higher pay” and assume that data users could infer the “teachers” and “Long Beach High School” parts from the ‘participants’ and ‘location’ fields, respectively. Instead, write something like “for higher pay for teachers in Long Beach School District”.
- CCC applies regular expressions to the claims summaries you write to try to attach tags representing one or more of a few dozen major political issues (e.g., racism, environment, education, LGBTQ+ rights) with each event. Those regular expressions rely on keywords or phrases. **If there are specific issues you think an event should be associated with, be sure to include at least one keyword or key phrase from the relevant regular expression(s) in your summary.** For example, if you’re describing a protest against mask mandates in schools, you would want to write something like “against coronavirus mask mandates in schools” instead of the shorter “against mask mandates” so the keyword “coronavirus” would trigger the ‘covid’ tag and the term “schools” would trigger the ‘education’ tag. For a list of issue tags and associated

keywords and key phrases, see [this document](#). For the actual regular expressions, see [this R script](#).

valence

Use this field to capture the left-right political orientation of the participants' claims as follows.

2 = right wing
1 = left wing
0 = other/neither

This is a judgment call, and we recognize that it will often be unclear. Here are some general pointers.

- Events organized by or in support of Republican Party organizations or candidates generally warrant a 2, while events organized by or in support of Democratic Party candidates or organizations generally warrant a 1.
- Events expressing claims traditionally associated with one or the other major U.S. political party should be coded accordingly. For example, the GOP has long advocated for restricting abortion rights, so “pro-life” events warrant a 2. Ditto for events advocating for Second Amendment rights. By contrast, events advocating for abortion rights or for gun control would generally get a 1 here, as those positions are traditionally associated with the Democratic Party.
- Events focused on hyper-local issues (e.g., opposition to a new residential development in the neighborhood) or essentially non-partisan issues (e.g., domestic violence, suicide awareness, drug addiction, animal cruelty) should generally be coded as 0.
- Events focused on foreign affairs generally warrant a 0 as well. The rare exceptions are cases where claims about foreign affairs are mixed with explicitly partisan claims about domestic issues (e.g., “stand with Israel, vote for Trump” or “solidarity with Palestine, against racism, defund police”).
- For claims that are not associated with either major political party but are broad and politicized, you will need to use your judgment. For example, while the core claims of many actions associated with the 2020 George Floyd uprising were not endorsed by leaders of the Democratic Party, we chose to give them a valence of 1 because they were clearly left wing and were firmly rejected by President Trump and other Republican leaders.
- When using your judgment about fuzzier cases, aim to be consistent with past practice on similar issues. When in doubt, consult project leaders.

event_type

List the forms the event took, as described by participants or observers, separated by semicolons. To make this field more useful to researchers, try to limit your entries to the following list. If you come across something truly exceptional, though, go ahead and innovate.

- protest
- counter-protest
- demonstration
- rally
- march
- strike
- picket
- walkout
- sickout
- caravan
- boat parade
- motorcycle ride
- bicycle ride
- run
- walk
- banner drop
- direct action

N.B. Starting in 2022, we began to put most events into one of the following categories according to the following definitions.

- **Protest.** A crowd gathering to express disagreement with, or disapproval or anger or frustration toward, a specific individual or organization that is at or near the crowd's gathering point (e.g., a politician giving a speech, a corporate headquarters, a bank branch, a construction site, a city hall), or in negative reaction to a recent or current event (e.g., the killing of George Floyd, the reversal of Roe v. Wade).
- **Rally.** A crowd gathering to demonstrate and reinforce their mutual support for a set of political demands or claims, typically by making speeches directed at each other. These include but are not limited to campaign rallies, and these events sometimes include or support press conferences.
- **Demonstration.** A crowd gathering to demonstrate their support for a set of political demands or claims to an external audience, typically with things like signs and banners and flags displayed to passers-by in a public space. Pickets may also be considered a form of demonstration.
- **March.** A crowd moving on foot from one location to another as a means to express their support for a set of political demands or claims.
- **Caravan.** Same as a march, but in motorized vehicles (e.g., cars, trucks, motorcycles).
- **Bicycle ride.** Same as a march or caravan, but mostly or entirely with bicycles.

- **Direct action.** An individual or crowd using civil disobedience or other physical actions that do not directly threaten to harm other humans as a way to express or advance their claims. Common forms involve occupations of space that impede others from doing things to which the protesters object (e.g., occupying a target lawmaker's office, forming a line to block the entrance to a target corporation's building, or chaining people to construction equipment on a target corporation's work site) and damaging property materially or symbolically relevant to protesters' claims (e.g., anti-capitalist protesters breaking windows in a Starbucks or anti-police protesters tagging a police station with graffiti). In the context of unrest in response to political events, looting may also be construed as a form of direct action and should be labeled as such.
- **Counter-protest.** See the [dedicated section below](#) for more on what these are.

police_measures

In words or phrases, briefly describe notable measures or actions taken by police or other law enforcement on the scene. If no information is provided about police presence or actions, leave this field blank.

In many cases, this field will be blank or something as simple as "on scene" or "followed marchers". In cases where police took more aggressive action, you might see a summary like "riot gear; formed skirmish line; declared unlawful assembly". In cases where police used additional crowd-control measures, try to be as specific as possible (e.g., "pepper spray; flash-bangs; less-lethal munitions"). If information is available about the size or scope of the law enforcement presence, it is helpful to capture that as well (e.g., "dozens of officers" or "National Guard and state troopers deployed").

participant_measures

In words or phrases, briefly describe notable measures or actions taken by participants in the event being recorded. If no notable actions were taken, leave this field blank.

As with the summary of police measures, this field will often be left blank because participants didn't do anything unusual or otherwise notable, or because no information was available about protesters' conduct.

This field can be especially useful for describing the nature of civil disobedience or direct actions in cases where those were involved (e.g., "roadblock", "sleeping dragon", "protesters locked themselves to construction equipment", "vandalized building facade"), or for describing interactions between protesters and police (e.g., "heckled police; threw water bottles") or protesters and counter-protesters (e.g., "verbal confrontation and scuffles with counter-protesters", "physical fighting with protesters").

When relevant, this field should also be used to capture information about targets of protesters' actions and how they were targeted. For example, if protesters shouted and held banners during a speech by a public figure or at a city council meeting, please note that here and include the name of the target (e.g., "disrupted speech by President Biden with shouting", "disrupted city council meeting with banners and chants", "bird-dogged Mayor Adams as he left press conference"). Likewise for actions at the homes or offices of individuals or organizations (e.g., "banged pots and pans outside home of U.S. Rep. Lou Correa", "heckled arriving guests at private fundraiser featuring Vice President Harris", "chanted at Northrup Grumman offices"). These names and verbs can be used by future researchers to identify events that featured these kinds of disruptions, or that targeted specific types of public figures or organizations, or both.

police_injuries

Record the number of police or law enforcement officers on the scene who were reportedly injured during the action.

- If a single number is reported, enter the number (e.g., 3).
- If a phrase is used to describe an ambiguous number of injuries (e.g., "more than 3"), record the phrase.
- If officers were reportedly injured and neither of the above applies, enter "unspecified".
- If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., the latest report from a law enforcement agency). If none appears authoritative, record "unspecified".
- If no officers were reportedly injured, leave blank.

participant_injuries

Record the number of participants in the action who were reportedly injured during the action.

- If a single number is reported, enter the number (e.g., 3).
- If a phrase is used to describe an ambiguous number of injuries (e.g., "more than 3"), record the phrase.
- If protesters appear to have been injured but neither of the above applies, enter "unspecified".
- If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., an organizer's after-action report). If none appears authoritative or all accounts are ambiguous, record "unspecified".
- If no participants were reportedly injured, leave blank.

Note that injuries of counter-protesters should not be counted here. Instead, they should be recorded in the row representing the counter-protest.

arrests

Record the number of participants in the action who were reportedly arrested or detained during the action.

- If a single number is reported, enter the number (e.g., 3).
- If a phrase is used to describe an ambiguous number of arrests (e.g., “more than 3”), record the phrase.
- If protesters appear to have been arrested but neither of the above applies, enter “unspecified”.
- If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., a law enforcement agency’s after-action report). If none appears authoritative or all accounts are ambiguous, record “unclear”.
- If no participants were reportedly arrested, leave blank.

Note that a citation alone is not enough to count toward this category. To count here, participants must be arrested or detained.

Note also that arrests of counter-protesters should not be counted here. Instead, they should be recorded in the row representing the counter-protest. If it is unclear from reports whether arrested individuals were protesters or counter-protesters or both, put “unspecified” in the ‘arrests’ field for both events.

property_damage

In words or short phrases, summarize any property damage caused by participants in the action (e.g., “graffiti; broken windows; dumpster fire”). If no property damage is reported, leave blank.

police_deaths

Record the number of police or law enforcement officers who reportedly died during the action.

- If a single number is reported, enter the number (e.g., 3).
- If a phrase is used to describe an ambiguous number of deaths (e.g., “more than 3”), record the phrase.
- If officers reportedly died and neither of the above applies, enter “unspecified”.
- If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., the latest report from a law enforcement agency). If none appears authoritative, record “unspecified”.
- If no officers reportedly died, leave blank.

participant_deaths

Record the number of participants in the action who reportedly died.

- If a single number is reported, enter the number (e.g., 3).
- If a phrase is used to describe an ambiguous number of deaths (e.g., “more than 3”), record the phrase.
- If protesters reportedly died but neither of the above applies, enter “unspecified”.
- If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., an organizer’s after-action report). If none appears authoritative or all accounts are ambiguous, record “unspecified”.
- If no participants reportedly died, leave blank.

Note also that deaths of counter-protesters should not be counted here. Instead, they should be recorded in the row representing the counter-protest.

macroevent

When relevant, this field is used to record a unique ID that links a counter-protest (or counter-protests) to the event it (or they) target. (For more on what constitutes a counter-protest, see [Counter-protests](#) below.) The goal is to facilitate analysis of protester/counter-protester interactions by making it easy to filter and group events where these interactions occurred.

Macroevent IDs are composed of three parts: a date, a location, and some brief descriptor of the central event’s content or theme. For example a Black Lives Matter protest and counter-protest in Boston on June 10, 2021, might be given matching macroevent IDs that read “20210610-boston-blm”.

The length of the ID doesn’t matter much, and the exact construction of the IDs (including typos) is not really important, either. The main concerns are the following.

- IDs must be unique to a given protest/counter-protest pair or set (hence the use of the date-location-descriptor construction, which virtually ensures this)
- IDs must match exactly within protest/counter-protest pairs or sets (best practice is to use copy/paste to ensure this)
- IDs should not contain any spaces or special punctuation marks, just numbers and letters and hyphens (e.g., “20211001-nyc-abortion”, not “20211001 New York abortion”)

[NB. From January 2021 until December 2022, we also used this field to capture other types of linkages between events, including explicit coordination across events in numerous places on the same date and continuity of an event in a single location across time. In December 2022, we changed the coding guidelines and revised the historical data to restrict this field to linking protest/counter-protest sets only. This was the use case that drove us to create this field in the first place, but we found it very difficult to reliably filter to those sets when the field also included IDs for other types of connections. So, we simplified.]

notes

Use this field to record any information or observations about the event that you think are interesting or important but are not already captured in the preceding fields. For example, were there notable speakers at the event? Do you want to clarify something about how or why the event was coded? This is your space to use or not, but aim to keep your observations civil and non-judgmental.

coder

Coders should enter their first and last initials (e.g., “JU”). If you add information to an event already coded, add your initials after a semicolon and a space (e.g., “JU; JP”).

source[n]

Paste the URL of a relevant source. Or, for sources that do not have URLs, enter a short description (e.g., “public submission” or “eyewitness”).

Special Cases

A few types of events commonly arise that require some additional explanation on how to encode them.

Future Events

If you come across information about an event that is planned but has not yet happened, add the event to the spreadsheet but leave blank the columns relating to things like attendance, arrests, property damage, and injuries. Sometimes you will later come across an after-event report that will allow you to fill in the missing details.

If you later learn that the event was canceled, delete the row. If you learn that the event was rescheduled, edit the date and move it to the correct place in the sheet.

Counter-protests

A counter-protest is an action that is organized or occurs in direct response to another action and usually (but not always) engages directly with that original action. We treat actions as counter-protests whenever a) an event appears to have been organized primarily for purposes of confronting or challenging another event or b) when organizers describe their action as a counter-protest to another CCC-coded event and hold it on the same date, even if the counter does not take place in the same location as the action it is targeting.

This category includes planned protests, rallies, demonstrations, and the like that intend to confront other actions (e.g., demonstrators outside a venue in which a political candidate is holding a rally, or antifascist activists who mobilize to confront a right-wing march). CCC also uses this category to capture more spontaneous confrontational actions with apparent political intent, including individuals or small groups who engage in counter-messaging protesters in a sustained way (e.g., following and heckling them) or who violently attack protesters (e.g., ram protesters with their vehicle). Note, however, that an action described as a response or counter to another action is not coded as a counter-protest if it occurs on a different date.

A protest and a counter-protest each have their own row in the sheet. Whenever you record a counter-protest, be sure to do the following three things.

- Label it a “counter-protest” in the ‘type’ field.
- Assign a common macroevent ID to it and the event it is countering, using copy/paste to ensure that the IDs match.
- Use the ‘participant_measures’ fields in the two rows to capture available information about the nature of interactions between protesters and counter-protesters (e.g., “verbally confronted and scuffled with protesters”).

Everything else should be handled just as you would any other event.

NOTE: If an event that is initially planned as a counter-protest occurs but the event it is meant to counter does not occur (i.e., is canceled), the event that does occur should be labeled something else in the ‘type’ field (e.g., “rally”, “demonstration”), and no ‘macroevent’ ID should be assigned.

Online Events

As noted in the Introduction, CCC covers online events that otherwise meet its criteria. This means that we capture information about things like virtual rallies, protests, vigils, and walkouts (e.g., student or teacher Zoom-outs during the COVID-19 pandemic) but not online petitions, paid fundraisers, celebrations, and so on.

Online events are treated just like other events with three exceptions.

- The locality in which an online event occurs is not always specified or clear. Here is how we handle that ambiguity.
 - If an online event is associated with a specific town, record it as you usually would.
 - If the event is associated with a specific state but not a town, enter the name of the state capital in the ‘locality’ field. For example, an online get-out-the-vote rally for Wisconsin Democrats would get “Madison” as its locality and “WI” as its state.

- If the event is intentionally national in scope, use “Washington” and “DC” as the locality and state, respectively.
- If the event is intentionally international in scope, do not record it.
- In the ‘location’ field, enter “online”.
- In the ‘type’ field, add “virtual” to the type: “virtual rally”, “virtual vigil”, and so on.

Riots

CCC does not apply the label “riot” to any events. As Chenoweth (2021: 55) notes, that term has become pejorative and is often used by authorities or opponents to delegitimize protesters’ actions. In the light of the political weight of this term, we choose to apply one or more of the terms listed above—including “direct action” in cases where protesters deliberately destroy or seize property—instead of labeling them as riots.

Diffuse Coordinated Actions

Some events involve coordinated nonviolent action in service of advancing political claims, but participants in the event do not gather or otherwise act collectively in a specific location as part of the protest. Sickouts are a common example, but other forms of protest (e.g., [wearing certain colors to school](#)) or coordinated noncompliance may also occur and are germane to CCC’s collection.

CCC handles these events similar to the way we handle online ones.

- If the action is associated with a specific location (e.g., a single school or a workplace), record the ‘locality’, ‘state’, and ‘location’ as you usually would.
- If the action occurs across multiple locations in a single city or town, record the relevant ‘locality’ and ‘state’, then enter “citywide” in the ‘location’ field.
- If the action occurs across multiple locations in a single state, enter the name of the state capital in the ‘locality’ field, then put “statewide” in the ‘location’ field.
- If the action is national in scope, use “Washington” and “DC” as the locality and state, respectively, and put “nationwide” in the ‘location’ field.
- For ‘type’, use specific terms where relevant (e.g., “sickout”, “walkout”). If none of those apply, enter “coordinated action” as the event type, and then use the ‘participant_measures’ field to briefly describe what that action was (e.g., “wore black or red to school”).

FAQ

Do I enter an event even if sources don't provide any crowd counts?

Yes, just leave empty the columns pertaining to crowd size (size_text, size_low, size_high).

Do I enter an event that has not happened yet?

Yes, but leave blank the columns relating to attendance, arrests, damage, injuries, and the like. Sometimes you will later come across an after-event report that will allow you to fill in those missing details. If you later learn that the event was canceled, delete the row. If you learn that the event was rescheduled, edit the date and move it to the correct place in the sheet.

For location, do I need to include every reported detail?

No. Focus instead on trying to capture major addresses or landmarks that Google Maps recognizes (and you are encouraged to use Google Maps to sort out what that means). In cases where an event moves from location to location, record as many of those major points as possible, separating them with semicolons.

What do I enter for locality if the event spanned more than one town or traveled between towns?

Pick one notable locality (often the start or end point) and enter it in the 'locality' field, and then list the other towns in the 'location' field. For example, if protesters caravanned from Palo Alto to Sacramento and then rallied at the state capitol there, you might enter "Sacramento" as the locality, "CA" as the state, and then record "Palo Alto, CA; California State Capitol" in the 'location' field.

What if protesters from two locations or organizations join together? Are they listed as a single event or two separate ones?

If participants organize separately and begin their events separately, then their actions are listed as two separate events, even if they eventually merge their actions ([example](#)). However, if the source or sources list a meeting point so people can travel together to an event, the meeting point—e.g., a train station or parking lot—is not listed as its own event. Also, if multiple organizations each advertise on their own that they are bringing a group to the same action, that action is still listed as one event, not one event per organization.

Should I list an event that only appears on a Facebook event page or on an organizational calendar or website?

Yes. Keep the event listed even if you cannot find a news report. We assume it is more likely that an event happened but did not get media coverage than that the listed event didn't happen.

Do we count online protests?

Yes, we also track crowds that gather online. See the Online Events section under the Special Cases header above for guidance on how to encode them.

How do we handle events that have both an online and in-person component?

Some events combine in-person and online elements, often but not always in the form of a livestream of the in-person action. In these cases, create separate records (rows) for the in-person and online parts, and use a common macroevent ID to link the two. This approach is similar to how we handle coordinated actions across multiple in-person locations. It also gives maximum flexibility to researchers who may want to limit analysis to in-person actions (so they can filter out the online gathering without losing the in-person action), and it allows us to record distinct information for the two (or more) parts when relevant (e.g., about the number of participants). (NOTE: This does not extend to participants or observers who independently livestream an event. We are talking specifically about cases where organizers of an in-person action also organize a parallel online component or gathering.)

What if an event recurs daily, weekly, or monthly?

Each day or iteration counts as a unique event and gets its own row.

What if a single event lasts for multiple days in a row?

Same story: each day or iteration counts as a unique event and gets its own row. The partial exceptions are events that start one day and last into the early-morning hours of the next; in these situations, only record a single event on the date the action started.

In the arrest column, do we count protesters who are issued citations?

No. To count as arrests, participants must be arrested or detained. A citation alone is not enough to be listed.

Release Notes

This section provides information on significant changes to the collection or coding process, to make other researchers aware of them and provide guidance on how to account for them in any uses of the data set.

April 2023: Ended Coverage of Labor Strikes

As of 1 April 2023, CCC stopped collecting data on labor strikes.

Since late 2020, the [ILR Labor Action Tracker](#) project at Cornell University has been doing an excellent job making structured data on labor strikes in the U.S. So, in early 2023, we decided to stop duplicating that project's work. Labor strikes have always been a bit of a stretch of CCC's idea of political crowds, and we have limited resources, so it seemed like a sensible change to us. We continue to collect on rallies, protests, and similar actions in support of striking workers, but not on the strikes and workers' pickets themselves.

Time and resources permitting, we may try to extract and archive all records of labor strikes from the historical data (January 2017–March 2023). In the meantime, researchers doing longitudinal comparisons of gross event counts and the like are advised to filter out records with either “strike”, “strike; picket”, or “walkout; strike; picket” (in no particular order) as the complete text of the ‘type’ field in the compiled and augmented version of the dataset. Note, though, that some waves of student-led demonstrations in the late 2010s calling for action on climate change were assigned “strike” as the event type at the time. So, if you are attempting gross comparisons that include that period, you'll probably want to include a carveout in your filtering rules that excludes events with a valence of 1 and “environment” as an issue tag.

Works Cited

Chenoweth, Erica. 2021. *Civil Resistance: What Everyone Needs to Know*. New York: Oxford University Press.

Soule, Sarah and Christian Davenport. 2009. "Velvet Glove, Iron Fist, or Even Hand? Protest Policing in the United States, 1960-1990," *Mobilization: An International Quarterly* (Vol. 14, Issue 1): 1-22.

Appendix: Modifications to Consider

A running list of potential changes to the coding procedures, in no particular order.

Ordered indicator of protester violence

One of the most-cited ACLED reports from early 2021 compares levels of violence across police and protesters across protester types. Right now, CCC's summary measures of police and protester interactions are too crude to do this well (injuries, deaths, arrests, and property damage). One option would be to create an ordered indicator of the intensity of violence we use to record the highest observed level of violence by each group at the event, something like:

- 0 None reported or observed
- 1 Verbal confrontation or threats
- 2 Minor physical confrontations (e.g., shoving, scuffles) with no reported injuries
- 3 Physical confrontation with injuries
- 4 Menacing with lethal weapons (e.g., pointed gun, waved knife, aimed car)
- 5 Injurious physical attack (e.g., sustained brawl, stabbing, shooting)

One tricky thing about this would be distinguishing between actions directed at police and actions directed at counter-protesters (where relevant). To address this, we could create separate fields for those two.

Ordered indicator of police violence

Same idea, but for the police. Something like:

- 0 Not on scene or on scene but observed only
- 1 Modest constraints, no direct action (e.g., tape, barricades, warnings, riot gear)
- 2 Substantial constraints, no direct action (e.g., formed skirmish line, blocked road)
- 3 Moderate direct action (e.g., pepper balls, flash-bangs, beanbags, kettling)
- 4 Heavy direct action (e.g., tear gas, batons, gunshots)

Protest target

Devin Judge-Lord, a postdoc at HKS in 2021-22, suggested this one. He studies the effects of activism on policy and thought it would be useful to have a field that, when relevant, identifies the government agency or elected official or other actor whom protesters are targeting with their demands (e.g., president, Congress, school board, EPA, etc.). I think our work is too broad to reduce this to a short list we could treat as a categorical indicator, but we could just make it a text field to start. We could also pair that field with a categorical indicator for the level of government targeted (federal, state, county, local).

Clearer and stricter list of event types

This comes from a conversation among Erica, Jay, and Victoria. Right now, the set of event types tracked in the 'event_type' field is not clearly defined, nor are the event types themselves.

If we want this category to be analytically useful, we probably need to tighten it up with a list of the set of event types covered (probably including an “other” for ambiguous or edge cases) and definitions or descriptions of those types. Some distinctions probably don’t matter for most use cases (e.g., protest vs. demonstration), but others do (e.g., strikes and direct actions). If we did this, coders could continue to use the ‘participant_measures’ field to provide more nuance or detail (e.g., a sit-in might get “demonstration” in ‘event_type’ but then “sit-in” as one bit of info in ‘participant_measures’). Also, Erica pointed out that “non-cooperation” might be an important and interesting thing to track as an event type for things like boycotts and sickouts and such.