The Anatomy of Resistance Campaigns (ARC) Dataset: Organization Identification Instructions, Version 2.0

January 15, 2026

Contents

1 Variables in ARC_Events.csv **2**

2 Introduction **2**

3 Key Definitions **3**
3.1 Formal Organizations . . . 3
3.1.1 Organization types . . . 5
3.2 Maximalist demands . . . 7
3.3 Fronts . . . 8

4 Coding Procedure **11**
4.1 GED: Rebel Groups . . . 11
4.2 SCAD events . . . 12
4.2.1 Event types 1-6 . . . 12
4.2.2 SCAD Violent Events: Event Type 8 . . . 16
4.3 Finalizing the codesheet . . . 16

5 Error mitigation strategies **17**

6 Limitations **18**
.

1 Variables in ARC_Events.csv

Variable	Details
country_name	Name of country the event occured in
country_abbr	Country abbreviation
ccode	Correlates of War Country Code
eventid	Event ID, prefixed with scad. for SCAD events and ged. for GED events
data_source	Either SCAD (scad) or GED (ged)
event_startdate	Date the event started (from underlying events source)
event_enddate	Date the event ended (from underlying events source)
org_id	ARC 2.0 organization ID
maximalist	Was the event maximalist (1=Yes)
nonviolent_scad	Was the event a nonviolent SCAD event?

Table 1: Variable descriptions

2 Introduction

This section describes the coding procedure for identifying organizations in the Anatomy of Resistance Camapaigns dataset. This is a condensed version of a longer, more specific and personalised document that researchers on the project used to guide their coding. Please get in contact if you would like this more specific file.

All coding was completed in RMarkdown to enhance transparency and replicability of coding decisions. Researchers worked within customised RMarkdown files based on the country they were assigned. We have kept the underlying events data used in the project (GED and SCAD events) plus the sentences that led to coding decisions in the case of SCAD events. Four variables were coded in this phase of the project:

- Whether there were any formal organizations that participated in events of dissent in the UCDP Georeferenced Event Dataset (Version 19.1, **?** and the Social Conflict Analysis Dataset (Version 3.2, **?**).
- Whether there were any "maximalist" demands made at those events.

- Whether any participating organizations were "fronts" and ties between those fronts and other organizations in ARC.
- The organization type for all identified organizations.

Organizations identified in this first phase were passed to phase two of the project, which coded the features of participating organizations.

This information booklet proceeds by first describing operational definitions for key concepts, followed by the coding procedure for events in the GED and then events in SCAD.

3 Key Definitions

3.1 Formal Organizations

Conceptually, an organizations are structures designed to cohere people and resources – often through collective action – to pursue common goals (**?**). The presence of a formal structure (however thin the hierarchy) intended to aggregate individual efforts towards a defined goal distinguishes organizations from broad social categories such as 'students', 'protesters', or the 'working class'. Empirically, we identify organizations as any group with a proper noun (i.e a capital letter). In many cases this is straightforward, such as "The Darfur Students Association", but in other cases it can be more unclear and we stick to the Proper Noun rule. One example is an event in Burundi that was organized by "the Collective", and we treat this as an organization (i.e "The Collective"). Another example is "The Zapatistas" - we would also treat this as an organization to be coded in the project.

We were inclusive in this first phase of the project and used the second phase to filter out organizations that may have proper nouns but are not formal organizations. This focus on formal organizations also means that we don't include more informal organizational groupings, such as ethnic groups or religious groupings unless they are mentioned in the context of a formal organization (organizations that represent more informal social groups, however, enter the project with the "social base" variable in phase two).

Table 4 shows examples of expressions that count as the positive identification of organizations in protest events. Any combination of these nouns/pronouns identifying the organization and verbs signifying activity were valid for identifying participation. Coding was not limited to these expressions though, if the sentence identified members of an organization at an event or they organized an event, participation was coded even if these specific verbs and nouns were not used. Participation was **always coded for all organizations that were explicitly mentioned in the sentence, we never included fewer organizations than those mentioned in the sentences**. Based on the front rule (see below) we *sometimes* included more, when these organizations had been members of a front for less than 3 years.

Table 2: Expressions identifying participation

Noun/Pronoun	Verb
Organization X	organized the event
Activists from organization X	called for the event
Supporters of organization X	organized the event
The leader/leaders/leadership of organization X	participated in the event
Representatives of organization X	were present at the event
	urged/encouraged members/supporters/activists to join
	joined the event

Only organizations that participated in dissent on the territory of a given country are included in that country's register of active organizations. A Senegalese-based political party that engaged in a protest for democratization in Senegal would be included in the ARC data, while an activist group in London, or South Africa protesting in London or South Africa for democracy in Senegal would not. For example, in relation to the complicated case of the Islamic State (IS), only affiliates of IS that were locally based are counted as actors in a given country. For example, *Jund al-Khilafah* (Soldiers of the Caliphate) is an IS affiliate based in Algeria and for Algeria we would include this organization as an actor in ARC, however, we would not also include *Ansar Bayt al-Maqdis* which was the Egyptian affiliate of IS as an actor in Algeria. This is even though both organizations are – in ARC – considered to be a part of the IS 'front' organization. More details about front organizations can be found below.

3.1.1 Organization types

Organization types are based on the general goals of the organizations, not the types of events that they participated in (i.e maximalist or non-maximalist). Organization types were coded according to the following typology:

Our definitions of economic sectors come from the World Bank:

Industrial Sector: "Industry corresponds to ISIC divisions 10-45 and includes manufacturing (ISIC divisions 15-37). It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. Data are in current U.S. dollars."

Table 3: Organization Types

Code	Type	Org. Goals	Example
1	*Revolutionary*	Maximalist goals of regime change (not separatism)	AFDL: Alliance of Democratic Forces for the Liberation of the Congo
2	*Political Party*	Win elections	UFDG: Union of Democratic Forces of Guinea
3	*Industrial Organization*	Worker's rights/interests in the industrial sector	CTNG: The National Confereration of Guinean Workers
4	*Service Organization*	Professionals's rights/interests in the service sector	ONAT: Tunisian Bar Association
5	*Women's*	Women's rights/interests	AFTD: Tunisian Association of Democratic Women
6	*Ethnic/Indigenous*	Ethnic rights/interests	NCBIP: National Coordinating Body for Indigenous Peoples
7	*Religious*	Coordinate religious activities	FNAI: Tunisian Front of Islamic Associations
8	*Student or Youth*	Youth/student interests	GUTS: General Union of Tunisian Students
9	*Other CSOs*	Broad, universal non-maximalist goals (i.e human rights) or CSOs that do not fit other categories	CNOSCG: National Council of Civil Society Organizations
10	*Criminal*	Illegal economic activity	M18: Marra 18
11	*Peasant/Agrarian Organization*	Interests of workers in agriculture	FETRAGOM: Federation of Agricultural Workers of Gros-Morne
12	*Separatist*	Territorial Independence or Autonomy	MAK: Movement for the Autonomy of Kabylie (Algeria)

Service Sector: "Services correspond to ISIC divisions 50-99 and they include value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services. Also included are imputed bank service charges, import duties, and any statistical discrepancies noted by national compilers as well as discrepancies arising from rescaling. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The industrial origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3 or 4."

Agricultural Sector: "Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3 or 4."

3.2 Maximalist demands

Whether there were maximalist demands made at an event is a binary variable (1 = yes, 0 = no). We define *maximalist* demands as **calls for changes in the political structure which, if implemented, would significantly alter the executive's immediate access to state power, the rules with which executives are selected, or the policy or geographic areas for which the executive has the right to make laws.**

Examples of maximalism include demands that a head of state resign via a non-institutional method, demands for democratization in autocratic settings, demands to enfranchise an ex-

cluded social group, and demands for regional or ethnic autonomy or independence. Maximalist demands exclude calls for political change that fall short of these fundamental aspects of executive power, such as improved human rights protections, or changes in public spending. A series of borderline demands and their treatment can be found in the FAQ section of the coding instructions.

3.3 Fronts

A front is an organization formed to coordinate the political activities of constituent organizations. We include all of the member organizations as participants in a event (specifically, a sentence) where they have been members of the front for less than three years. Most organizations join a front when it forms so the ties from member organizations to the front and the age of the front are usually the same. But this is not always the case as organizations leave and join fronts over time.

Another way to think about how to coding organizations and fronts in the sentences and their ties, is to think about how old each organization-front tie is. Where ties between a member organization and a front are less than three years old, we code that organization as participating wherever the front participates and then add the front ties at the end. Where those ties are more than three years old we don't include the member organization as a participant when the front participates (we code just the front) and the front tie does not need to be included at the end. For most cases member organizations all join fronts at the same time, so each organization-front tie is about the same age as the front itself. But this is not always the case.

We never go more than one-level down a front for all fronts that are explicitly mentioned in a sentence. Sometimes members of a front are also fronts themselves. But we do not code the members of this "front within a front" unless that organization is explicitly identified as a participant in a sentence separately. Then we would also include all of the first-level participants of this front.

One example is the Sudanese Revolutionary Front, which was an alliance of four rebel

groups formed in November 2011. Since the years we were coding for SRF were 2016 and 2017, this front is more than 3 years old, so we don't include it's constituent organizations as fronts. However, the UCDP website says that "The United People's Front for Liberation and Justice (UPFLJ), consisting of 17 small groups from eastern Sudan yet based mainly outside of the country, also joined later in October 2013." Since 2016 is less than 3 years from October 2013 (i.e this front tie would only reach the 3 year threshold in October 2016), then we include the UPFLJ as a constituent organization. It's probable that the UPFLJ was formed in May 2013, meaning that we should also include the constituent organizations of this front, but there is very little to no information on this, so we will include only the UPFLJ. The constituent organization is absorbed into the front (not treated as a separate organization) if it remains a part of the front for three or more years. So in this case, in 2017 (more than 3 years after UPFLJ joined the SRF) we treat the SRF as a single organization.

The only exception to this rule is where there was explicit evidence that constituent organizations fully merged and ceased to exist independently. For example, the Trade Unions Congress of Swaziland (TUCOSWA) is a merger of the Swaziland Federation of Trade Unions (SFTU) and the Swaziland Federation of Labour (SFL). The SFTU and SFL deregistered as trade unions after the merger and ceased to exist.

While any such threshold is arbitrary, we argue this strikes a balance between theoretical and practical concerns. The UCDP amalgamation rule might be practical for some rebel groups, but obscures too much when examining the broad fronts that often make up nonviolent movements. However, to disaggregate all fronts would be impractical and would overlook fronts' typical move towards unity over time. In an ideal world we could precisely identify when fronts ultimately merge into a unitary organization, but in our experience information is simply too sparse to achieve this. As such, we rely on an arbitrary cut-off, but one that we believe balances theoretical and resource demands.

Figure 1 is a flowchart used to help making decisions about fronts.

We used the following set of standardised sources to guide researchers when doing additional research, beyond the newsire text, for fronts:

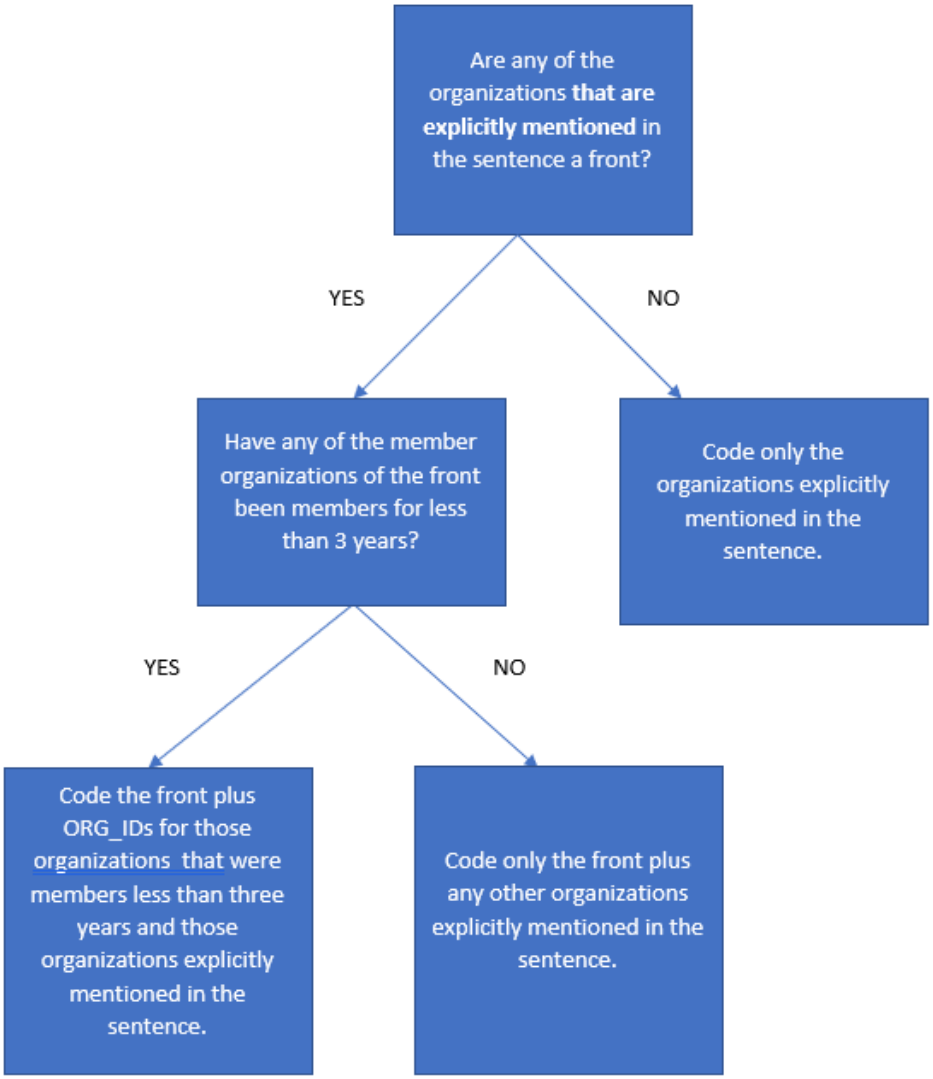

Figure 1: What to do with fronts

Here is a standardised list of sources that you should consult when working out whether an organization is a front or not:

1. Web searches (international organizations, known international news organizations, academic sources, even Wikipedia can give you an idea of whether an organization is a front). Try searching for the name in the local language as well.
2. Targeted FACTIVA Searches for the name of the front.
3. Political Handbook of the World (PHoW) if it's a front of political parties, Uppsala Conflict Data Program Conflict Encyclopaedia (`https://ucdp.uu.se/`, Trade Unions of the World (TUoW)
4. The Historical Dictionary Series (if available - many can be accessed through the library).
5. Refworld `https://www.refworld.org/`
6. Web searches for the organization's website, facebook page, twitter page

4 Coding Procedure

Researchers used an RMarkdown file for each country-case, which guided coders through the procedure of coding organizational participation in: (1) GED events, (2) SCAD event types 1-6, (3) Violent SCAD events, and (4) Codesheet finalization including organization names, types, and front ties. Each of these steps is discussed in more detail below. If users would like the very specific coding instructions given to researchers, including the FAQ, please get in contact with the research team.

4.1 GED: Rebel Groups

The first section of the ID file contained code to identify rebel groups in the Georeferenced Event Dataset (GED) (**?**).

Researchers coded all organizations that participated in state-based violence in the GED as participants in ARC. Therefore all rebel groups that are in the GED in our sample period are in ARC. We also treat all rebel groups as making maximalist demands of the state because this is an inclusion criteria in the GED to be a participant in state-based conflict.

Working with the GED therefore primarily involved attributing organization identifiers (ORG-IDs) to rebel groups and deciding whether rebel groups are *fronts* and conducting additional research to find which members of the front met the three year rule (above) and then adding these organizations as participants.

4.2 SCAD events

The real value-added of ARC comes in the coding of organizations in SCAD events. For each SCAD event included in the ARC dataset, researchers used searches of newswire texts to identify whether organizations participated in the event and whether maximalist demands of the state were made.

4.2.1 Event types 1-6

First, we only included events that were targeted at the central government and were of issue types 1 through to 6 (demonstrations, riots, and strikes). We coded "anti-government violence" events separately below. For this subset of events, researchers followed the procedure below.

Researchers first obtained details of the start and end dates of the event and the event description from the SCAD notes, such as in the example below:

```
>
> case_output <- scad_sud %>% filter(eventid==i) %>%
+   dplyr::select(issuenote, notes, actor1, actor2, actor3, startdate, enddate, eventid)
> case_output
                                                
```

Press, BBC, Al Jazeera, Xinhua, Trend News Agency and AllAfrica (for African Coutries) and EFE (for Central American cases).

FACTIVA SEARCH STRING: hlp=COUNTRYNAME* AND hlp=(protest* OR strike* OR riot* OR demonstrat* OR marche* OR rally* OR rallie*) NOT "striker" NOT "strike a deal" AND (SN="Associated Press" OR SN="BBC" OR SN="Agence France" OR SN="Reuters" OR SN="AllAfrica" OR SN="EFE" OR SN="jazeera" OR SN="Xinhua" OR SN="Trend News Agency")

The search string was identical for every event that we researched. Researchers then downloaded the newswires found by the search string with a unique file ID. These newswires were then loaded back into RStudio through the RMarkdown file and broken into individual sentences. Sentences were preprocessed to show the top 50 sentences that were most likely to have information relevant to coding maximalist demands and organizational participation, based on a prediction model. This model was optimzed to show the number of sentences (50) that researchers read vis a vis the detection of maximalist claims and organizational participation. Coders were also shown the sentence preceding and following these top 50 sentences. Researchers then read each of the sentences, looking for evidence of organizational participation or maximalist demands.

Recall that the definition of a maximalist demand is **calls for changes in the political structure which, if implemented, would significantly alter the executive's immediate access to state power, the rules with which executives are selected, or the policy or geographic areas for which the executive has the right to make laws.**

Organizations were identified where the newswire text indicated that a formal organization (using the proper noun rule) organized or was present at an event. The following table is a set of typical circumstances under which positive identification of organizational participation might be made.

If evidence was found for either maximalist demands or organizational participation, this was recorded at the *sentence* level, such that we can retrieve the specific sentences upon

Table 4: Expressions identifying participation

Noun/Pronoun	Verb
Organization X	organized the event
Activists from organization X	called for the event
Supporters of organization X	organized the event
The leader/leaders/leadership of organization X	participated in the event
Representatives of organization X	were present at the event
	urged/encouraged members/supporters/activists to join
	joined the event

which each coding decision was made. Each sentence that contained either evidence of organizational participation or maximalist demands was recorded, even if there were multiple sentences that indicated the same or similar information. That is, even though we aggregate the information to the event level in the final dataset, there can be multiple sentences per event that led to a coding decision. For example, it could be that sentence `3` leads to the identification of three organizations (`ORG-001, ORG-002, and ORG-003`), and sentence `10` leads to the identification of two of those organizations (`ORG-002 and ORG-003`) then both sentences were coded with the organizations identified specific to those sentences. We have retained the sentences that led to the coding of either organizational participation or maximalism for each event. Cases can have no maximalist claims and no organizations, maximalist claims and no organizations, maximalist claims and organizations and non-maximalist claims and organizations. So we are identifying organizations that participate in non-maximalist events.

As discussed in the section on 3.3, where a participating organization was identified as a *front* we coded the participation of the front plus any member organizations whose ties to the

front were less than three years old. Please refer to the flow chart in the 3.3 section for how these decisions were made. The front rule means that we *never include fewer organizations than those explicitly mentioned in the sentence text, but we sometimes include more based on the front rule.*

We also coded events as having organizational participation or maximalist demands if the SCAD event notes or actor codes indicated that this was the case.

For very long events that returned too many newswires (specifically longer than 21 days and returning more than 200 newswires) we included only newswires from the first 10 days of the event and the last 10 days of the event. This circumstance was very rare.

Some events overlapped in time and the newswire texts also overlap across events. For example, there may be a protest on Friday by youth over student fees and a strike on Saturday over wage increases. The newswires for the Saturday event will sometimes include reports of the Friday event. Where we were clearly able to separate the events by the issues or demands, only organizations and demands related to that specific event have been included. However, with many events it was difficult to tell whether other events that are mentioned in the newswires are related or not to the SCAD event that is being coded. Where researchers could confidently say that an event mentioned in the newswires is different to the one being coding (perhaps because the issues are clearly different or because the text indicated it occurred on a different day) then those sentences would not be included as evidence as maximalism or organizational participation. We did our best to work out whether an protest or event in a newswire is related to the one we were coding, but **erred on the side of caution, where it's better to code the sentences if we were not sure than not code them.** In other words, *we were inclusive rather than exclusive.* For example, pro-democracy protests can also include protests by many other groups demanding many different things. It's difficult to say in these cases whether a protest is related or not, so treat them as related and code the sentences as you would normally. We leant on the side of inclusion rather than exclusion in these situations.

We repeated this procedure for each relevant event in the SCAD dataset.

4.2.2 SCAD Violent Events: Event Type 8

We then coded organizations associated with anti-government violence in SCAD. These were treated in a similar way to GED events, where we coded organizations based on the actor names and the SCAD event text only. We coded organizational participation in these events only where there was a proper noun indicated in the texts. For example, we would code organizational participation where (for example) SPLA-IO particpated but not where "bandits" were indicated as participating. As with the GED (and earlier SCAD events) we also checked whether organizations were fronts.

4.3 Finalizing the codesheet

The final part of the coding procedures involved adding information on organization names, types and information about front ties.

First, each unique organization ID was assigned a name (and multiple names if needed). We followed a standard procedure of using an acronym first, followed by a colon, then the fuller name of the organization. For example: `FAO: The Fabulous Amazing Organization`.

Second, each organization was attributed a primary organization type based upon it's main goals (please refer to the 3.1.1 section for a list of organization types). In general this was inferred from case knowledge developed from coding the case, but some additional research was sometimes needed. Only in a very small number of cases were we not able to attribute a primary organization type. Each organization can have only one type per country-year. In the vast majority of cases this is one type for the entire coding period. Where it was clear that organizations changed type over the coding period (say from a "revolutionary" rebel group to a political party), this was coded by attributing types to specific year blocs.

Finally, researchers then coded front ties for organizations that were identified as fronts. Please refer to section 3.3 for a more detailed definition and coding procedure for fronts. Fronts involved some additional research to identify front members using sources mentioned in the 3.2 section. Ties were coded from the front to the identifiers for each organization

that was a member of that front. These ties can change over time (as ties become older than 3 years, organizations leave, or new members join) and where this was the case we coded specific combinations of front ties for the relevant year blocs.

At this point the coding for the country was completed. The RMarkdown file then automatically aggregates the coding into event-level and organization-country-year data, preserving the sentences used to positively identify maximalist demands and organization types. The above coding procedure was replicated for each country in our sample.

5 Error mitigation strategies

We implemented several error mitigation strategies. Several are discussed above, including using a consistent search string and concepts and a standardized (although non-exclusive) set of verb-noun combinations to identify organizational participation.

Most important among our error mitigation strategies were weekly team meetings. In these meetings, maintained over the entire coding period from 2021-2022, the research team discussed unclear coding decisions and built shared understandings of the main concepts, and how our coding rules were to be applied in unclear circumstances. These meetings also helped us to build consistent responses to borderline cases. Although the impact of our team meetings is difficult to quantify, we believe the many hours invested in this process is one of our most important bias mitigation strategies. These meetings also resulted in an FAQ section, which can be found in the coding instructions.

In addition, we conducted a round of tests in 2020 to identify areas where researchers disagreed across key concepts. We had 5 researchers code 10 randomly selected SCAD events and 5 randomly selected GED events, collecting information on whether the event had maximalist demands and the organizations that participated. These tests were **not** designed as intercoder reliability tests with the intention of producing agreement statistics, rather as training exercise to help us identify areas where we could improve the coding process for increased reliability. Many of the coding procedures above were *developed* as a response to these tests, not as a test of their reliability. It is also important to note

that these tests were undertaken in a context where researchers were recommended not to communicate with each other about the cases (i.e. without the weekly team meetings where we would discuss and resolve cases).

For coding maximalism, across the 10 SCAD events, all five coders perfectly agreed in 9/10 events and, in the one exception, 4/5 coders agreed. Out of 50 "chances" to get the coding wrong, researchers got this coding "correct" 49 times. Across the 10 events, the average agreement per event was 98%.

It's more difficult to assess agreement on organization identification because there is no set range on the number of organizations that could be identified. One approach is to say that every organization identified across the 5 coders was an "opportunity" to agree. By this approach an event that had a maximum of 4 organizations identified and 5 coders would have 20 chances at agreement. Using this method, there were 80 opportunities at agreement across the 10 events and 5 coders, upon which coders agreed 64 times, or 80%.

For those cases where researchers disagreed, our tests revealed that this was often because researchers saw different sentences because they used slightly different search strings, not because they interpreted the information especially differently. A major change following these tests was to fully standardize the search string, such that all coders saw the same information. This was retrofit to existing cases following these tests.

Overall, we consider the agreement on the ID side of the data to be high and are confident in the reliability of the coding of maximalism and organization identification. The main concern to our mind is that there continue to be organizations that are not identified in the data, either because they were not mentioned in search strings, or because coders omitted them based on their reading of the text. While the risk of the former is important, we think the risk of the latter should lead only to small differences.

6 Limitations

Despite employing the strategies above, there remain limitations to the data of which users should be aware. Biases inherent in newswire texts also apply to our data. Biases related

to urban versus rural locations, violent versus nonviolent, and large versus small events are also very likely present in our data. Users should be aware that the baseline reports that we use to identify organizations are likely biased towards urban, large, events that potentially involve the use of violence. While our inclusion of regional news digests (i.e. All Africa and EFE) might reduce some of these biases, they will not be removed entirely (**?**).

In addition to this, our method for identifying organizations is likely biased toward larger organizations and those with experience interacting with the media or with dedicated media wings. Overall, we think this probably translates into a bias toward political parties and large trade unions, at the expense of smaller CSOs that may be crucial to mobilization processes but less visible to the media. Organizations that make more extreme demands or use more extreme, attention-grabbing tactics might also be overrepresented in the data. Although the addition of research into front members might reduce some of this bias, we think that it remains present in the data. These biases also likely apply to larger events, where newswires have only a limited number of words with which to report on events and despite a large number of organizations in participation, may only report (using proper nouns) the largest, most media-savvy, or most attention-grabbing organizations. Another way to think about this bias is that we suspect our data are biased towards a combination of the "main" organizations that were involved in an event and those with features that make them more attention-grabbing or available to the media.