Guide to 2023 ACLED Column Changes

On 20 March 2023, ACLED made changes to the structure of the dataset by removing three existing columns and adding three new columns. These changes aimed to eliminate redundancies and reduce potential sources of confusion while also introducing more useful variables for analysis. In addition, some existing columns were repositioned within the dataset to improve the readability of the event data.

Three columns – *event_id_no_cnty*, *data_id*, *and iso3* – have been removed from the dataset. The decision to remove these columns came after careful consideration and internal reviews of their usefulness. Alongside these column removals, ACLED introduced three new variables to facilitate analysis and provide users with additional information about events. The new columns are *disorder_type*, *civilian_targeting*, and *tags*, all of which are described in detail below.

This guide outlines the aforementioned changes to the ACLED dataset and explains how users can adjust to them. Users should review the guide carefully to ensure the updates do not interrupt their workflow.

Column Removals

event_id_no_cnty

ACLED originally created the *event_id_no_cnty* column so users could sort event IDs within a single country. However, other columns (e.g. *event_date* or *timestamp*) can be used to sort events chronologically and do so more accurately. Because of its redundancy and potential inaccuracy when used for chronological sorting, the *event_id_no_cnty* column was removed from the ACLED dataset. If a user still requires this column, it can easily be reproduced using the *event_id_cnty* column (see below for details).

data_id

The *data_id* column contained auto-generated IDs representing each event's row within the

ACLED dataset. Importantly, the values in the data_id column were not static, but changed each time the data were updated. The dynamic nature of the *data_id* column created confusion, as some users would mistake the values for static, unique IDs. ACLED removed the column from the dataset to reduce this confusion and prevent the use of an incorrect ID variable in the future. To uniquely identify and track events, users should always use the *event_id_cnty* column, which contains unique IDs that remain static even as the dataset is updated.

iso3

The ACLED dataset contained several columns indicating the country in which an event occurred. Two columns in particular – *iso* and *iso3* – allowed users to easily join ACLED data with external datasets. The *iso* and *iso3* columns, respectively, provided users with a country's unique numeric code and three-letter code from the International Organization for Standardization (ISO). These columns provided different, but entirely interchangeable, country identifiers. ACLED therefore removed the *iso3* column from the dataset in order to eliminate this redundancy and create space for more useful columns.

Column Additions

disorder_type

The new *disorder_type* column provides users with a broader classification of event types. This new classification system will allow users to more easily identify and filter relevant event categories, particularly those that are often used in ACLED methodology documentation and analysis. Each event will be assigned a disorder type based on the *event_type* and *sub_event_type* columns:

disorder_type	***event_type/sub_event_type***
Political violence	• *Battles* • *Explosions/remote violence* • *Violence against civilians* • *Mob violence* • *Excessive force against protesters*
Demonstrations	• *Protests (all sub-event types, including excessive force against protesters)* • *Violent demonstration*
Strategic developments	• *Strategic developments*

Note that the *disorder_type* categories are not mutually exclusive, as the *excessive force against protesters* sub-event type (a subset of the *protests* event type) is classified under both *political violence* and *demonstrations*.

civilian_targeting

The new *civilian_targeting* column denotes that violence in an event mainly or solely targeted civilians. Without this column, users can only identify civilian targeting events by applying a combination of filters across *event_type*, *sub_event_type*, and various actor columns. The *civilian_targeting* column eliminates the need for such complex filtering, as it will contain one of two values: "Civilian targeting," which indicates that civilians were targeted during the event, or blank (null), which indicates that ACLED found no reports that civilians were the main or sole target in the event. The lack of a civilian targeting designation does not rule out the possibility that civilians were affected by violence in the event, however (e.g. as 'collateral damage' in the context of a *battle or explosions/remote violence event*).

tags

ACLED uses a variety of tags to provide standardized information about events. For example, tags may denote the size of a demonstration, whether women were specifically targeted in a violent incident, or connections to a particular political movement (e.g. "stop the steal" in the United States). Tags were previously included in the *notes* column within square brackets, which could make it difficult to filter events by tag or to extract tag information (e.g. size of a demonstration). The addition of a standalone *tags* column facilitates the extraction and analysis of tagged events. All tags that were in the *notes* column were shifted to the new *tags* column.

Preparing for Column Updates

Anyone using ACLED data downloaded prior to 20 March 2023 will be using an outdated column structure. Those users will need to take certain steps to ensure that the column changes do not interrupt their workflow, regardless of whether they actively use the specific columns that were removed. Even users who do not currently use the removed columns may be affected by changes in column positions. **All users should follow the steps below to review and update any scripts and/or Excel files used to interact with ACLED data.** The following examples focus specifically on Excel, R, and Python, but the underlying logic can easily be applied to other platforms and programming languages.

Step 1: Ensure that any references to the column index number (i.e. the column's position number) are updated.

The position of nearly all ACLED columns changed as a result of the column updates. Therefore, references to column numbers or letters must be updated in all scripts and Excel formulas. These are the new column positions:

Column Name	Column Letter (Excel)	Column Number	Positional Change
event_id_cnty	A	1	2 to the left
event_date	B	2	3 to the left
year	C	3	3 to the left
time_precision	D	4	3 to the left
disorder_type	E	5	New column
event_type	F	6	2 to the left
sub_event_type	G	7	2 to the left
actor1	H	8	2 to the left
assoc_actor_1	I	9	2 to the left
inter1	J	10	2 to the left
actor2	K	11	2 to the left
assoc_actor_2	L	12	2 to the left
inter2	M	13	2 to the left
interaction	N	14	2 to the left
civilian_targeting	O	15	New column
iso	P	16	14 to the right
region	Q	17	Same position
country	R	18	Same position
admin1	S	19	Same position
admin2	T	20	Same position
admin3	U	21	Same position
location	V	22	Same position

latitude	W	23	Same position
longitude	X	24	Same position
geo_precision	Y	25	Same position
source	Z	26	Same position
source_scale	AA	27	Same position
notes	AB	28	Same position
fatalities	AC	29	Same position
tags	AD	30	New Column
timestamp	AE	31	1 to the right

Examples:

a. Excel

Before the column change, this VLOOKUP formula would have accessed the *event_date* column by referencing column index **3** (i.e. the third column in the selected data range).

	A	B	C	D	E	F	G	H	I
1	*data_id*	*iso*	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*event_type*	*sub_event_type*
2	7348008	780	TTQ971	971	2018-12-29	2018	1	Violence again	Attack
3	8888831	780	TTQ970	970	2018-12-28	2018	1	Battles	Armed clash
4	7348033	780	TTQ969	969	2018-12-27	2018	1	Violence again	Attack
5	7347992	780	TTQ968	968	2018-12-26	2018	1	Battles	Armed clash
6	7347931	780	TTQ967	967	2018-12-23	2018	1	Violence again	Attack
7	7348041	780	TTQ966	966	2018-12-21	2018	1	Violence again	Attack
8	7347938	780	TTQ965	965	2018-12-18	2018	1	Violence again	Attack
9	7884729	780	TTQ964	964	2018-12-17	2018	1	Protests	Peaceful protest
10	7347986	780	TTQ961	961	2018-12-16	2018	1	Violence again	Abduction/forced
11	8566225	780	TTQ962	962	2018-12-16	2018	1	Violence again	Attack
12									

	Formula	=VLOOKUP("TTQ971",C2:E11,3,FALSE)
	Result	2018-12-29

After the column change, the same VLOOKUP formula now references column index **2** because *event_id_no_cnty* is no longer present, shifting *event_date* to the second column in the selected range.

	A	B	C	D	E	F	G
1	*event_id_cnty*	*event_date*	*year*	*time_precision*	*disorder_type*	*event_type*	*sub_event_type*
2	TTQ971	2018-12-29	2018	1	Political violence	Violence against c	Attack
3	TTQ970	2018-12-28	2018	1	Political violence	Battles	Armed clash
4	TTQ969	2018-12-27	2018	1	Political violence	Violence against c	Attack
5	TTQ968	2018-12-26	2018	1	Political violence	Battles	Armed clash
6	TTQ967	2018-12-23	2018	1	Political violence	Violence against c	Attack
7	TTQ966	2018-12-21	2018	1	Political violence	Violence against c	Attack
8	TTQ965	2018-12-18	2018	1	Political violence	Violence against c	Attack
9	TTQ964	2018-12-17	2018	1	Demostrations	Protests	Peaceful protest
10	TTQ961	2018-12-16	2018	1	Political violence	Violence against c	Abduction/forced
11	TTQ962	2018-12-16	2018	1	Political violence	Violence against c	Attack
12							

	Formula	=VLOOKUP("TTQ971",B2:F11,2,FALSE)
	Result	2018-12-29

b. R

Before column change	After column change

`# Select event_id_cnty (column 3) and notes (column 28)`
`event_notes <- acled_df %>%`
`  select(3, 28)`
	`# Select event_id_cnty (column 1) and notes (column 28)`
	`event_notes <- acled_df %>%`
	`  select(1, 28)`

c. Python

Before column change	After column change
`#Subsetting to event_id_cnty and notes` `event_notes = acled_df.iloc[0:9,[2,27]]`	`#Subsetting to event_id_cnty and notes` `event_notes = acled_df.iloc[0:9,[0,27]]`

Step 2: Ensure that any references to column letters are updated. (Note: This step is specific to Excel.)

As indicated in the previous step, column positions changed, which therefore impacted column letters in Excel (as outlined in the chart in step 1). Users should adjust to these changes by updating any references to column letters in Excel.

Examples:

a. Excel

Before the column change, this COUNTIF formula accessed the *sub_event_type* column (column letter I) by referencing cells I2:I11.

	A	B	C	D	E	F	G	H	I	J	K	L
1	data_id	iso	event_id_cnty	event_id_no_cnty	event_date	year	time_precision	event_type	sub_event_type	actor1	assoc_actor_1	inter1
2	7348008	780	TTQ971	971	2018-12-29	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
3	8888831	780	TTQ970	970	2018-12-28	2018	1	Battles	Armed clash	Unidentified Armed Group (Tr		3
4	7348033	780	TTQ969	969	2018-12-27	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
5	7347992	780	TTQ968	968	2018-12-26	2018	1	Battles	Armed clash	Unidentified Armed Group (Tr		3
6	7347931	780	TTQ967	967	2018-12-23	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
7	7348041	780	TTQ966	966	2018-12-21	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
8	7347938	780	TTQ965	965	2018-12-18	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
9	7884729	780	TTQ964	964	2018-12-17	2018	1	Protests	Peaceful protest	Protesters OWTU: Oilfield Wo		6
10	7347986	780	TTQ961	961	2018-12-16	2018	1	Violence against civilian	Abduction/forced d	Unidentified Armed Group (Tr		3
11	8566225	780	TTQ962	962	2018-12-16	2018	1	Violence against civilian	Attack	Unidentified Armed Group (Tr		3
12												

Formula	=COUNTIF(I2:I11, "Attack")	
Result	6	

After the column change, the same COUNTIF formula accesses the sub_event_type column (column letter G) by referencing cells G2:G11.

	A	B	C	D	E	F	G	H	I	J
1	event_id_cnty	event_date	year	time_precision	disorder_type	event_type	sub_event_type	actor1	assoc_actor_1	inter1
2	TTQ971	2018-12-29	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
3	TTQ970	2018-12-28	2018	1	Political violence	Battles	Armed clash	Unidentified Armed Group (		3
4	TTQ969	2018-12-27	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
5	TTQ968	2018-12-26	2018	1	Political violence	Battles	Armed clash	Unidentified Armed Group (		3
6	TTQ967	2018-12-23	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
7	TTQ966	2018-12-21	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
8	TTQ965	2018-12-18	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
9	TTQ964	2018-12-17	2018	1	Demostrations	Protests	Peaceful protest	Protesters OWTU: Oilfield W		6
10	TTQ961	2018-12-16	2018	1	Political violence	Violence against civilian	Abduction/forced	Unidentified Armed Group (		3
11	TTQ962	2018-12-16	2018	1	Political violence	Violence against civilian	Attack	Unidentified Armed Group (		3
12										

Formula	=COUNTIF(G2:G11, "Attack")
Result	6

Step 3: Eliminate all named references to the removed columns.

References to the *data_id, event_id_no_cnty*, or *iso3* column names will produce errors and/or cause formulas and scripts to fail. Users should either delete all references to the removed columns or manually recreate the desired columns (see steps 4-6 for more details on replacing and/or reproducing the removed variables).

Examples:

a. Excel

Before the column change, this FILTER formula directly referenced *event_id_no_cnty* as the first column in the selected range of columns. (Note: The filter formula does not automatically provide column headers as shown in the result; these were added manually.)

	A	B	C	D	E	F	G	H	I	J	K
1	*data_id*	*iso*	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*event_type*	*sub_event_type*	*actor1*	*assoc_actor_1*
2	7348008	780	TTQ971	971	2018-12-29	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
3	8888831	780	TTQ970	970	2018-12-28	2018	1	Battles	Armed clash	Unidentified Armed Group (Trinidad an	
4	7348033	780	TTQ969	969	2018-12-27	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
5	7347992	780	TTQ968	968	2018-12-26	2018	1	Battles	Armed clash	Unidentified Armed Group (Trinidad an	
6	7347931	780	TTQ967	967	2018-12-23	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
7	7348041	780	TTQ966	966	2018-12-21	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
8	7347938	780	TTQ965	965	2018-12-18	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
9	7884729	780	TTQ964	964	2018-12-17	2018	1	Protests	Peaceful protest	Protesters (Trinidad an	OWTU: Oilfield V
10	7347986	780	TTQ961	961	2018-12-16	2018	1	Violence against civiliar	Abduction/forced d	Unidentified Armed Group (Trinidad an	
11	8566225	780	TTQ962	962	2018-12-16	2018	1	Violence against civiliar	Attack	Unidentified Armed Group (Trinidad an	
12											

Formula	=FILTER(step3_before[[event_id_no_cnty]:[sub_event_type]],step3_before[fatalities]>0)						
Result	event_id_no_cnty	event_date	year	time_prec	event_type	sub_event	
	970	2018-12-28	2018	1	Battles	Armed clash	
	969	2018-12-27	2018	1	Violence against	Attack	
	966	2018-12-21	2018	1	Violence against	Attack	
	962	2018-12-16	2018	1	Violence against	Attack	

However, after the column changes, this formula returns an error (#REF) because *event_id_no_cnty* no longer exists. To avoid this error, *event_id_no_cnty* is replaced with *event_id_cnty* in the formula.

	A	B	C	D	E	F	G	H	I
1	*event_id_cnty*	*event_date*	*year*	*time_precision*	*disorder_type*	*event_type*	*sub_event_type*	*actor1*	*assoc_actor_1*
2	TTQ971	2018-12-29	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	
3	TTQ970	2018-12-28	2018	1	Political violence	Battles	Armed clash	Unidentified Armed Group	
4	TTQ969	2018-12-27	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	
5	TTQ968	2018-12-26	2018	1	Political violence	Battles	Armed clash	Unidentified Armed Group	
6	TTQ967	2018-12-23	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	
7	TTQ966	2018-12-21	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	
8	TTQ965	2018-12-18	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	
9	TTQ964	2018-12-17	2018	1	Demostrations	Protests	Peaceful protest	Protester:	OWTU: Oilfield W
10	TTQ961	2018-12-16	2018	1	Political violence	Violence against civiliar	Abduction/forced c	Unidentified Armed Group	
11	TTQ962	2018-12-16	2018	1	Political violence	Violence against civiliar	Attack	Unidentified Armed Group	

Formula	=FILTER(step3_after[[event_id_cnty]:[sub_event_type]],step3_after[fatalities]>0)						
Result	event_id_cnty	event_date	year	time_precision	disorder_type	event_type	sub_event
	TTQ970	2018-12-28	2018	1	Political violence	Battles	Armed clash
	TTQ969	2018-12-27	2018	1	Political violence	Violence against civilians	Attack
	TTQ966	2018-12-21	2018	1	Political violence	Violence against civilians	Attack
	TTQ962	2018-12-16	2018	1	Political violence	Violence against civilians	Attack

b. R

Before column change	After column change

`# Select a subset of columns`
`new_df <- acled_df %>%`
`  select(data_id, event_id_cnty,`
`  event_date, country, latitude,`
`  longitude, notes, iso3)`
	`# Select a subset of columns`
	`new_df <- acled_df %>%`
	`  select(event_id_cnty,`
	`  event_date, country, latitude,`
	`  longitude, notes)`

c. Python

Before column change	After column change
`#Subsetting some columns` `new_df=acled_df[[“data_id”, “event_id_city”,“event_date”,“country”,“latitude”,“longitude”,“notes”,“iso3”]]`	`#Subsetting some columns` `new_df=acled_df[[“event_id_city”,“event_date”,“country”,“latitude”,“longitude”,“notes”]]`

Step 4: Begin using the *event_id_cnty* column in place of *event_id_no_cnty*, or regenerate the *event_id_no_cnty* column.

References to *event_id_no_cnty* should be replaced by *event_id_cnty*, which is simply *event_id_no_cnty* with a country abbreviation prefix added. Note that *event_id_no_cnty* is not a unique identifier of an event. Nonetheless, if users still require this column, it can be regenerated by removing the country abbreviation from values in the *event_id_cnty* column.

Examples:

a. Excel

Option 1 - Replace all references to *event_id_no_cnty* with *event_id_cnty*

Before the column change, the COUNTA and UNIQUE formulas shown here referenced the *event_id_no_cnty* column to count the number of unique events in the data. (Note that this approach was prone to errors and, as in this case, did not produce the desired result because *event_id_no_cnty* was not a unique event identifier. This example highlights one of several justifications for removing this column from the ACLED dataset.)

	A	B	C	D	E	F	G	H
1	*data_id*	*iso*	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*event_type*
2	7348008	780	TTQ971	971	2018-12-29	2018	1	Violence against civiliar
3	8888831	780	TTQ970	970	2018-12-28	2018	1	Battles
4	7348033	780	TTQ969	969	2018-12-27	2018	1	Violence against civiliar
5	7347992	780	TTQ968	968	2018-12-26	2018	1	Battles
6	7347931	780	TTQ967	967	2018-12-23	2018	1	Violence against civiliar
7	7348041	780	TTQ966	966	2018-12-21	2018	1	Violence against civiliar
8	7347938	780	TTQ965	965	2018-12-18	2018	1	Violence against civiliar
9	7884729	780	TTQ964	964	2018-12-17	2018	1	Protests
10	7347986	780	TTQ961	961	2018-12-16	2018	1	Violence against civiliar
11	8566225	780	TTQ962	962	2018-12-16	2018	1	Violence against civiliar
12	8577777	780	SOM962	962	2018-12-16	2018	1	Violence against civiliar
13								

	Formula	=COUNTA(UNIQUE(Table14[event_id_no_cnty]))
	Result	10

However, after the column change, the same formula returns an error, as *event_id_no_cnty* does not exist. To address the error and correctly count unique events, the reference to *event_id_no_cnty* is replaced by *event_id_cnty*.

	A	B	C	D	E	F	G
1	*event_id_cnty*	*event_date*	*year*	*time_precision*	*disorder_type*	*event_type*	*sub_event_type*
2	TTQ971	2018-12-29	2018	1	Political violence	Violence against c	Attack
3	TTQ970	2018-12-28	2018	1	Political violence	Battles	Armed clash
4	TTQ969	2018-12-27	2018	1	Political violence	Violence against c	Attack
5	TTQ968	2018-12-26	2018	1	Political violence	Battles	Armed clash
6	TTQ967	2018-12-23	2018	1	Political violence	Violence against c	Attack
7	TTQ966	2018-12-21	2018	1	Political violence	Violence against c	Attack
8	TTQ965	2018-12-18	2018	1	Political violence	Violence against c	Attack
9	TTQ964	2018-12-17	2018	1	Demostrations	Protests	Peaceful protest
10	TTQ961	2018-12-16	2018	1	Political violence	Violence against c	Abduction/forced disap
11	TTQ962	2018-12-16	2018	1	Political violence	Violence against c	Attack
12	SOM962	2018-12-16	2018	1	Political violence	Violence against c	Attack
13							

Formula	=COUNTA(UNIQUE(Table25[event_id_cnty]))	
Result	11	

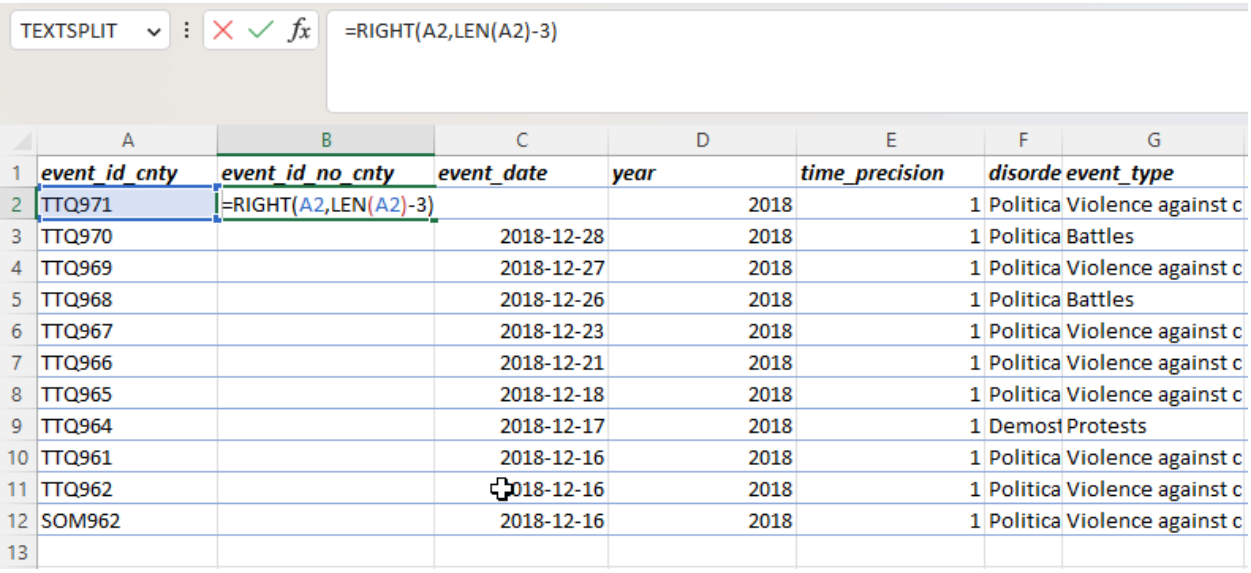
Option 2 - Regenerate *event_id_no_cnty*

1. Create a new column named "*event_id_no_cnty*."

	A	B	C	D	E	F	G
1	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*disorde*	*event_type*
2	TTQ971		2018-12-29	2018	1	Politica	Violence against c
3	TTQ970		2018-12-28	2018	1	Politica	Battles
4	TTQ969		2018-12-27	2018	1	Politica	Violence against c
5	TTQ968		2018-12-26	2018	1	Politica	Battles
6	TTQ967		2018-12-23	2018	1	Politica	Violence against c
7	TTQ966		2018-12-21	2018	1	Politica	Violence against c
8	TTQ965		2018-12-18	2018	1	Politica	Violence against c
9	TTQ964		2018-12-17	2018	1	Demost	Protests
10	TTQ961		2018-12-16	2018	1	Politica	Violence against c
11	TTQ962		2018-12-16	2018	1	Politica	Violence against c
12	SOM962		2018-12-16	2018	1	Politica	Violence against c
13							

2. Insert the following formula in the first row of the new column:

=RIGHT({event_id_cnty cell reference},LEN({event_id_cnty cell reference})-3)



	A	B	C	D	E	F	G
1	event_id_cnty	event_id_no_cnty	event_date	year	time_precision	disorde	event_type
2	TTQ971	=RIGHT(A2,LEN(A2)-3)		2018	1	Politica	Violence against c
3	TTQ970		2018-12-28	2018	1	Politica	Battles
4	TTQ969		2018-12-27	2018	1	Politica	Violence against c
5	TTQ968		2018-12-26	2018	1	Politica	Battles
6	TTQ967		2018-12-23	2018	1	Politica	Violence against c
7	TTQ966		2018-12-21	2018	1	Politica	Violence against c
8	TTQ965		2018-12-18	2018	1	Politica	Violence against c
9	TTQ964		2018-12-17	2018	1	Demost	Protests
10	TTQ961		2018-12-16	2018	1	Politica	Violence against c
11	TTQ962		2018-12-16	2018	1	Politica	Violence against c
12	SOM962		2018-12-16	2018	1	Politica	Violence against c
13							

3. Fill the column by clicking the autofill icon (the small square in the bottom left corner of the cell) and dragging it down.

	A	B	C	D	E	F
1	event_id_cnty	event_id_no_cnty	event_date	year	time_precision	disorde
2	TTQ971	971	2018-12-29	2018	1	Politica
3	TTQ970		2018-12-28	2018	1	Politica
4	TTQ969		2018-12-27	2018	1	Politica
5	TTQ968		2018-12-26	2018	1	Politica
6	TTQ967		2018-12-23	2018	1	Politica
7	TTQ966		2018-12-21	2018	1	Politica
8	TTQ965		2018-12-18	2018	1	Politica
9	TTQ964		2018-12-17	2018	1	Demost
10	TTQ961		2018-12-16	2018	1	Politica
11	TTQ962		2018-12-16	2018	1	Politica
12	SOM962		2018-12-16	2018	1	Politica
13						

	A	B	C	D	E	F	G
1	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*disorde*	*event_type*
2	TTQ971	971	2018-12-29	2018	1	Politica	Violence against c
3	TTQ970	970	2018-12-28	2018	1	Politica	Battles
4	TTQ969	969	2018-12-27	2018	1	Politica	Violence against c
5	TTQ968	968	2018-12-26	2018	1	Politica	Battles
6	TTQ967	967	2018-12-23	2018	1	Politica	Violence against c
7	TTQ966	966	2018-12-21	2018	1	Politica	Violence against c
8	TTQ965	965	2018-12-18	2018	1	Politica	Violence against c
9	TTQ964	964	2018-12-17	2018	1	Demost	Protests
10	TTQ961	961	2018-12-16	2018	1	Politica	Violence against c
11	TTQ962	962	2018-12-16	2018	1	Politica	Violence against c
12	SOM962	962	2018-12-16	2018	1	Politica	Violence against c
13							

b. R

Before column change	After column change
Option 1: Replace all the references to *event_id_no_cnty* with *event_id_cnty*	

`# Sort by event_id_no_cnty`
`acled_df <- acled_df %>%`
`  arrange(event_id_no_cnty)`
	`# Sort by event_id_cnty`
	`acled_df <- acled_df %>%`
	`  arrange(event_id_cnty)`

Option 2: Regenerate *event_id_no_cnty*	

`# Sort by event_id_no_cnty`
`acled_df <- acled_df %>%`
`  arrange(event_id_no_cnty)`
	`# Regenerate event_id_no_cnty and sort using str_extract`
	`acled_df <- acled_df %>%`
	`  mutate(event_id_no_cnty = str_extract(event_id_cnty, '\\d+')) %>%`
	`  arrange(event_id_no_cnty)`

`# Sort by event_id_no_cnty`
`acled_df <- acled_df %>%`
`  arrange(event_id_no_cnty)`
	`# Regenerate event_id_no_cnty and sort using gsub`
	`acled_df <- acled_df %>%`
	`  mutate(event_id_no_cnty = gsub("\\D", "", event_id_cnty)) %>%`
	`  arrange(event_id_no_cnty)`

b. Python

Before column change	After column change
Option 1: Replace all references to *event_id_no_cnty* with *event_id_cnty*	

`# Sort by event_id_no_cnty`
`acled_df =`
`acled_df.sort_values("event_id_no_cnty")`
	`# Sort by event_id_cnty`
	`acled_df=`
	`acled_df.sort_values("event_id_cnty")`

Option 2: Regenerate *event_id_no_cnty*	

`# Sort by event_id_no_cnty`
`acled_df=`
`acled_df.sort_values("event_id_no_cnty")`
	`# Regenerate event_id_no_cnty`
	`acled_df["event_id_no_cnty"] = acled_df["event_id_cnty"].str.replace(r'\D+', '', regex=True)`
	
	`# Sort by the regenerated event_id_no_cnty`
	`acled_df = acled_df.sort_values("event_id_no_cnty")`

Step 5: Begin using the *event_id_cnty* column in place of the *data_id* column.

As noted previously, the *data_id* column did not provide any analytical information for users and was not a static ID, meaning that it could not be used to uniquely and consistently identify a particular event. **The *event_id_cnty* column is the only column that serves as a static, unique event identifier.** Any references to the *data_id* column intended to be used as a unique event ID should immediately be replaced with *event_id_cnty*.

Examples:

a. Excel

Before the column change, the COUNTA and UNIQUE formulas shown here referenced the *data_id* column to count the number of unique events in the data. This approach was incorrect, as *data_id* was not a static identifier. If users combined ACLED events retrieved at different times into a single dataset, this may have introduced duplicated *data_id*'s, resulting in an incorrect count of the number of unique events. This error is highlighted in the following screenshot, in which the calculation does not produce the correct count of unique events.

	A	B	C	D	E	F	G	H
1	*data_id*	*iso*	*event_id_cnty*	*event_id_no_cnty*	*event_date*	*year*	*time_precision*	*event_type*
2	7348008	780	TTQ971	971	2018-12-29	2018	1	Violence against civilian
3	8888831	780	TTQ970	970	2018-12-28	2018	1	Battles
4	7348033	780	TTQ969	969	2018-12-27	2018	1	Violence against civilian
5	7347992	780	TTQ968	968	2018-12-26	2018	1	Battles
6	7347931	780	TTQ967	967	2018-12-23	2018	1	Violence against civilian
7	7348041	780	TTQ966	966	2018-12-21	2018	1	Violence against civilian
8	7347938	780	TTQ965	965	2018-12-18	2018	1	Violence against civilian
9	7884729	780	TTQ964	964	2018-12-17	2018	1	Protests
10	7347986	780	TTQ961	961	2018-12-16	2018	1	Violence against civilian
11	8566225	780	TTQ962	962	2018-12-16	2018	1	Violence against civilian
12	8566225	780	SOM962	962	2018-12-16	2018	1	Violence against civilian
13								

	Formula	=COUNTA(UNIQUE(Table14[data_id]))
	Result	10

After the column change, the absence of the *data_id* column causes the formula to produce an error (#REF). To avoid the error, the reference to *data_id* in the formula is replaced with *event_id_cnty*. In addition, because the *event_id_cnty* is a unique event identifier, the formula returns the correct result.

	A	B	C	D	E	F
1	*event_id_cnty*	*event_date*	*year*	*time_precision*	*disorder_type*	*event_type*
2	TTQ971	2018-12-29	2018	1	Political violence	Violence against c
3	TTQ970	2018-12-28	2018	1	Political violence	Battles
4	TTQ969	2018-12-27	2018	1	Political violence	Violence against c
5	TTQ968	2018-12-26	2018	1	Political violence	Battles
6	TTQ967	2018-12-23	2018	1	Political violence	Violence against c
7	TTQ966	2018-12-21	2018	1	Political violence	Violence against c
8	TTQ965	2018-12-18	2018	1	Political violence	Violence against c
9	TTQ964	2018-12-17	2018	1	Demostrations	Protests
10	TTQ961	2018-12-16	2018	1	Political violence	Violence against c
11	TTQ962	2018-12-16	2018	1	Political violence	Violence against c
12	SOM962	2018-12-16	2018	1	Political violence	Violence against c
13						

Formula	=COUNTA(UNIQUE(Table25[event_id_cnty]))
Result	11

b. R

Before column change	After column change

`# Count the number of sources per event`
`source_count <- acled_df %>%`
`  separate_rows(source, '; ') %>%`
`  count(data_id)`
	`# Count the number of sources per event`
	`source_count <- acled_df %>%`
	`  separate_rows(source, '; ') %>%`
	`  count(event_id_cnty)`

c. Python

Before column change	After column change
`# Setting the data frame index to data_id` `old_columns_dataset = old_dataset.set_index("data_id")`	`#Setting the data frame index to event_id_cnty` `new_columns_dataset= new_dataset.set_index("event_id_cnty")`

Step 6: Begin using the iso column instead of the *iso3* column.

Users who used the *iso3* column, including users who relied on this column to merge ACLED data with external datasets, should replace all references to *iso3* in scripts or spreadsheets with references to the *iso* column. Users should review the examples in step 5 for guidance on how to update column references.