# Codebook for

# Rebel Contraband Dataset

Version 1.0

http://www.civilwardynamics.org

When using the data described in this codebook, please cite James Igoe Walsh, Justin M. Conrad, Beth Elise Whitaker, and Katelin Hudak, "Financing rebellion: Introducing the Rebel Contraband Dataset," *Journal of Peace Research* (2018). Report any errors in this codebook, the dataset, or the accompanying set of notes, to jwalsh@uncc.edu. This material is based upon work supported by the U. S. Army Research Office through the Minerva Initiative under grant number W911NF-13-0332.

# 1. Introduction
# 1.1 Motivation

The Rebel Contraband Dataset (RCD) measures if and how non-state armed groups exploit natural resources and engage in crime to fund their violent and other activities.[1] Much research has investigated how natural resource wealth and criminal activity influence the outbreak, conduct, and duration and aftermaths of civil wars. The RCD contributes to this research by developing data on rebel exploitation of natural resources and their use of crime to fund their activities. It addresses two important limitations in existing data sources. First, much of the existing data are collected at the country level of analysis. This is problematic because it is possible for a resource or criminal activity to be located in parts of the country where rebel groups do not operate. In a case such as this, the rebel group might be counted as relying on the resource even when it is not, since both the resource and the group exist in the same country. Second, existing data do not provide information about how, if at all, a rebel group exploits a natural resource or the types of crime in which it engages. Some rebel groups might directly produce and sell resources and use the resulting profits to finance their violence. Others might steal the resource from legitimate producers. And other rebel groups might smuggle the resource from one location to another, but not directly involve themselves in the production and sale of the resource. These different strategies of resource exploitation might be associated with distinct patterns of rebel activity.

# 1.2 Unit of Analysis and Temporal Domain

The unit of analysis is the dyad-year from 1990 to 2015. We identify dyad-years using the DyadID variable from the UCDP Dyadic Dataset version 5.0-2016, available in .csv format at http://ucdp.uu.se/downloads/brd/ucdp-brd-dyadic-50-2016.csv.

Those planning to merge this data with other datasets should note one consequence of this unit of analysis: the same rebel actor may be involved in multiple conflicts simultaneously. For example, Al Qaeda in the Islamic Maghreb was a party to conflicts with both Algeria and Mauritania in 2012. This means that these actors appear multiple times in the dataset in different dyads. Researchers using the actor-year rather than dyad-year as their unit of analysis will want to be aware of this and may want to delete potential duplicate entries.

# 1.3 Workflow and Reliability

Coders for this project located and read sources of information about rebel organizations and, based on this and the variable definitions below, measured if and how a rebel group earned income from natural resources or crime. The basic steps in this process were as follows:

[1] "Contraband" is defined as "forbidden, illegitimate, unauthorized" (see Oxford English Dictionary). This definition is consistent with that used in Fearon (2004), which introduced the term in the context of the contemporary study of civil wars and political violence. See below for specific definitions of distinct forms of contraband.

1. Two coders were assigned to each observation, and initially worked independently to identify sources of information and to code variables from these sources. Coders created a document identifying the sources and justification for each measurement choice.
2. After both coders completed this work, they met with a project manager to compare their coding decisions. Finding accurate information about rebel organizations' use of resources is difficult, and often it is not obvious how a particular variable should be measured based on the available sources. The purpose of this meeting was to determine if the two coders reached different decisions, to resolve these differences through discussion and by collectively consulting the original sources again, and, if necessary, to search for additional information.
3. After this meeting, the manager finalized the measures for each variable and combined and organized the coders' notes into a single document.

Note that this project does not use a formal system of inter-coder reliability. There are two reasons for this. The first is that there is not a single source or type of source that provides relevant information across conflict dyads. For some dyads, most of the available information is found in UN reports; for others, most comes from media sources or NGO reports. This means that coders spent considerable time and effort identifying relevant sources, and that many of these sources required different types of interpretation. A second reason is that rebel groups have incentives to mask or exaggerate the extent to which they rely on different types of natural resources or crime to sustain their activities. This means that many sources have limited information on funding strategies, and the incomplete information found in different sources must be carefully interpreted and combined.

The workflow outlined above was designed to maximize reliability given these constraints. We used two coders who initially worked independently of each other to code each dyad to ensure that the initial search for information and its interpretation were not influenced by social dynamics. These coders then met with a project manager to discuss the different sources they identified and how their coding decisions were similar or different. This meeting identified gaps in each coders' work to this point. After this meeting, the coders were typically instructed to revisit existing sources and to look for additional sources to clarify the coding for each variable. We continued this process until the team reached a consensus about how to code each dyad-year, or to conclude that there is not enough information to determine if a group engaged in a particular strategy.

These decisions were documented in detail in the set of actor notes that accompanies this codebook. The actor notes include explicit descriptions of what information led to each coding decision and citations that support these decisions. The actor notes highlight observations where we were particularly uncertain about our coding decisions. As discussed below, the dataset also includes explicit measures of the degree and type of uncertainty the coders faced in making decisions. In some cases, we shared these conclusions with experts to solicit their opinions about the selection of sources and coding decisions. When these experts identified points that we may

have overlooked or mis-interpreted, we repeated the coding process using the suggestions that they made.

# 2. Variables

## 2.1 Identification Variables

These variables provide basic information about the dyad. The variables are pulled directly from the UCDP Dyadic Dataset; consult the codebook for that project for specific definitions:

SideA
SideB
Year
DyadID
SideBID
Incomp
Terr
Year
Intensity (the UCDP name for this variable is "Int")
Type
GWNoA
GWNoLoc
Region

We add the following variable:

Intervening: In some cases, there are gaps of time when a dyad is not included in the original UCDP dataset. For example, a dyad might appear in 2000, disappear for 2001 and 2002, and then reappear in 2003. We coded data for these years that appear between two years in the dataset. This variable takes a value of 1 when the dyad-year row is one of these intervening years, and a value of zero otherwise.

Note that these intervening years often do not appear in other datasets that use the UCDP format. Users who are merging with such datasets may wish to drop these observations.

## 2.2 Natural Resource Variables

These variables record if and how a rebel group profits from natural resources in a given year.

The variables are composed of two elements: the "funding strategy" and the resource. The funding strategy is how the rebel group profits from resources. There are four funding strategies described in greater detail below: extortion, theft, booty futures, and smuggling.

The resource is simply the natural resource that the rebel group is exploiting. The natural resources included in this project include:

animal: high-value animal products as ivory
bauxite

cannabis: includes marijuana or hemp
cassiterite
charcoal
coal
cobalt
coca
cocoa
coffee
coltan
copper
diamonds (alluvial)
diamonds (primary)
gold
iron
mercury
oil
opium
rubber
tea
timber
tin
titanium
wolframite
zinc
agriculture: any high-value agricultural product not otherwise listed above
drugs: used when specific drug is unknown
gems: used when the specific gemstone is unknown
minerals unknown: used for any mineral not listed above
other: used for any natural resource not otherwise listed here

Since there are 4 funding strategies and 31 resource categories (26 specific resources and 5 residual categories), there are a total of 124 variables. The variable names are recorded in the format "strategy_resource", for example, rubber_extortion or timber_smuggling. In the variable names, booty futures is abbreviated as bf.

These are dichotomous measures, taking a value of 1 if the rebel group used a funding strategy in the year, and a value of 0 if it did not. If, after extensive searches of relevant sources, there is no evidence that the group used this strategy, the dataset records a value of 0.

Note that these variables are not mutually exclusive. This means that a rebel group may pursue more than one of these funding strategies in a given year.

If a group pursues a particular strategy, then there are additional variables to complete as well. For example, if a group is identified as engaging in theft, we also identified how certain we are

that the group engaged in this strategy, the value of the items they stole, where the theft occurred, and so on. See below for more details.

## 2.2.1 Extortion

Variable names are listed in bold below. In the dataset, these names are slightly different; all letters are in lower case, and spaces are replaced with underscores (for example, extortion_location).

**Extortion**: This variable can take a value of zero (no extortion) or 1 (extortion). In this strategy, the group uses violence or the threat of violence to earn money directly or indirectly from the production of natural resources. Examples include:

- The armed group coerces workers to extract resources;
- The armed group protects illegal miners or farmers of illicit goods from the police or military, in exchange for a share of the proceeds;
- Rebels sell resources to outsiders on behalf of producers;
- The armed group hires workers, groups, or firms to engage in production;
- Rebels bomb a pipeline unless the operator of pipeline pays them protection money;
- Rebels kidnap workers at a mine and release them in exchange for ransom payment.

Extortion is an ongoing activity. It is not episodic. This is important because sometimes it is difficult to distinguish extortion from theft, defined below. Sometimes an armed group will loot a producer (such as a mine owner or farmer cultivating illegal drugs) that it wishes to extort, in order to demonstrate the consequences of not complying with its "offer" of protection. But this theft should subside once the producer acquiesces.

Armed groups also sometimes earn funds from cooperating with those who smuggle natural resources. This cooperation might resemble the common-sense understanding of the term extortion. For example, the armed group might threaten to attack or expose the smugglers unless they receive payment. For this project, the term extortion is only used when the armed group earns income directly or indirectly from the production of the resource. Since smugglers do not actually produce the resource, we do not include threats against smugglers by armed groups as extortion. Instead, we code this using the smuggling category below.

**Extortion Location**: The region or regions of the country where the group engages in extortion. This string variable records the first order administrative region of the country. This variable is left blank if the value for Extortion is equal to zero or if the information is not available.

**Extortion Geo KW:** Any additional geographic keywords or information. This variable is left blank if the value for Extortion is equal to zero or if the information is not available.

**Extortion Keywords:** Any keywords that further describe the activity of the rebel group. For example, if the group hires and pays workers, relevant text could include "hire workers". If no such terms are useful or relevant, this variable is blank.

**Extortion Funds:** The amount that the rebel group earns from extortion. It can take one of the following three values:

1 if the group earns less than $5 million from this activity;

2 if the group earns more than $5 million from this activity;

-99 if it is not possible to determine how much the group earns from this activity.

**Extortion Amount**: Textual information about the amount of funds the group earns from this activity. If no such information is available, this is left blank. This variable is left blank if the value for Extortion is equal to zero.

**Extortion Certainty:** See below for the definition of this variable.

## 2.2.2 Theft

**Theft:** This variable can take a value of zero (no theft) or 1 (theft). The rebel group engages in periodic theft of natural resources. In cases of theft, the rebel group does not have an ongoing relationship or contact with the producers. Examples of theft might include:

- Stealing oil from pipelines with the objective of selling the oil;
- Robbing warehouses or other storage facilities containing natural resources.

The key difference is that extortion requires that the armed group expect a payment or reward for stopping its actions or threats. Theft occurs when the group's only expectation is to obtain the natural resource for resale.

Sources of information occasionally identify the occurrence of theft, but not a clear attribution to a rebel group in our dataset. For example, the activity might be attributed to "rebels" or "bandits", but not to a particular rebel group. If this is all of the information that is available, the dataset does not attribute the theft to the rebel group.

**Theft Location**: This is identical to the corresponding variable for Extortion defined above.

**Theft Geo KW:** This is identical to the corresponding variable for Extortion defined above.

**Theft Keywords:** This is identical to the corresponding variable for Extortion defined above. For example, if the group steals resources from a convoy, the variable entry would be "convoy theft".

**Theft Funds:** This is identical to the corresponding variable for Extortion defined above.

**Theft Amount**: This is identical to the corresponding variable for Extortion defined above.

**Theft Certainty:** See below for the definition of this variable.

## 2.2.3 Booty Futures

**Booty Futures:** This variable can take a value of zero (no booty futures) or 1 (booty futures). The rebel group earns income by promising another actor exploitation rights over the natural resource in the event that the rebel group gains formal control over the territory (Ross, 2012). These other actors might be another country, a foreign firm (such as a resource firm or a firm of mercenaries that are promised a concession to exploit the resource if the rebels gain control of it), etc. The non-state actor may or may not be involved in the direct production of the resource. The key element of booty futures is that the rebel group promises and receives payments **today** in exchange for a promise to allow another actor to profit from a natural resource **in the future**.

**Booty Futures Location**: This is identical to the corresponding variable for Extortion defined above.

**BF Geo KW:** This is identical to the corresponding variable for Extortion defined above.

**Booty Futures Type**: This is identical to the corresponding variable for Extortion defined above.

**Booty Futures Keywords:** This is identical to the corresponding variable for Extortion defined above.

**Booty Futures Funds:** This is identical to the corresponding variable for Extortion defined above.

**Booty Futures Amount**: This is identical to the corresponding variable for Extortion defined above.

**Booty Futures Certainty:** This is identical to the corresponding variable for Extortion defined above.

## 2.2.4 Smuggling

**Smuggling:** This variable can take a value of zero (no smuggling) or 1 (smuggling). The rebel group earns income by directly engaging in or protecting those who smuggle natural resources illegally. As explained above, extortion of people who smuggle natural resources is coded as smuggling. Most smuggling involves transporting the goods across international borders.

**Smuggling Location**: This is identical to the corresponding variable for Extortion defined above.

**Smuggling Geo KW:** This is identical to the corresponding variable for Extortion defined above.

**Smuggling Keywords:** This is identical to the corresponding variable for Extortion defined above.

**Smuggling Funds:** This is identical to the corresponding variable for Extortion defined above.
**Smuggling Amount**: This is identical to the corresponding variable for Extortion defined above.

**Smuggling Certainty:** This is identical to the corresponding variable for Extortion defined above.

## 2.3 Crime Variables

The dataset measures the degree to which an armed group engages in criminal fundraising activity that does not involve natural resources. The criminal activity must be highly profitable and sustained to be coded as such. Many rebel groups engage in small-scale criminal activity. Such activity is not always reported reliably in the sources used in this project. Measuring small-scale crime would likely be very difficult and would produce a biased estimate of the true rate of crime. For this reason, the dataset includes only criminal activities that are mentioned in multiple sources, and/or which the sources make clear are an important source of finance for the rebel group.

The dataset includes the following crime variables. Each takes a value of zero (when the armed group does not engage in the criminal activity) or one (when the group does engage in this criminal activity) in each year of the dataset:

**Theft**: This includes only large-scale theft, such as multiple bank robberies, stealing large numbers of vehicles or other valuable goods, and so on. It might also be described in sources as "robbery" or "hijacking". Theft includes credit card fraud.

**Extortion**: The group "taxes" much of the economic activity that occurs in a particular area. Examples can include roadblocks, providing "protection" to marketplaces and villages, and so on.

**Smuggling human**: The group engages in or sponsors criminal smuggling of people, often across international borders. This includes smuggling of refugees and others who might be willing participants in the activity.

**Smuggling other:** The group engages in or sponsors criminal smuggling of things other than natural resources or humans, often across international borders.

**Humanitarian Aid:** The group requires aid organizations to pay it in order to deliver aid or to continue operations in a particular area, or the group steals humanitarian aid before it can be delivered to its intended recipients.

**Piracy:** The group engages in or sponsors piracy (meaning the theft, ransoming, etc. of sea vessels, not the violation of intellectual property rights).

**Kidnapping International:** The group kidnaps people in exchange for payments; at least one of the victims is from a country outside the area of the conflict.

**Kidnapping:** The group kidnaps people in exchange for payments; all of the victims are from the country or area where the conflict occurs.

**Other:** Types of crime not mentioned above.

Most criminal activity involving drugs will be captured by the natural resources dataset. Activity involving chemically manufactured drugs, such as crystal meth, is counted in the crime dataset.

Each type of crime has additional variables associated with it:

**Crime Location**: This is identical to the corresponding variable for Extortion defined above.

**Crime Geo KW:** This is identical to the corresponding variable for Extortion defined above.

**Crime Keywords:** This is identical to the corresponding variable for Extortion defined above.
**Crime Funds:** This is identical to the corresponding variable for Extortion defined above.

**Crime Amount**: The amount of funds the group earns from this activity. If no such amount is available, record a value of -99.

**Crime Certainty:** This is identical to the corresponding variable for Extortion defined above.

# 3. Certainty

This variable measures how certain we are about each of the measures recorded above. Certainty variables are included for any instance in which a value of 1 was recorded for one of the substantive variables described above.

**Certainty**: This measures the degree of confidence in determining that an armed group engaged in a particular activity. This is a nominal variable. It can take six values:

Higher Certainty (1): The activity is documented by two or more credible and independent sources (such as academic articles, NGO reports, media stories) or in a single detailed United Nations report that is itself based on the collection and analysis of a large number of sources.

Lower Certainty/One Source (2): Activity is documented by only one credible and independent source.

Lower Certainty/Bias (3): Activity is documented only by a source or a range of sources, such as a national government or a rebel group, that may have an interest in influencing perception that the rebel group in question engages in the activity.

Lower Certainty/Identity (4): Activity is documented in at least one credible source, but insufficient details are provided to code each group year with higher certainty. In countries where more than one group is operating, the source might use terms such as “rebels” or “insurgents”, but not identify the specific group by name. In cases such as this, the coding team seeks to infer the identity of the group from the area in which the strategy occurs and other information. If this is not possible, the relevant strategy variables are coded as zero, indicating insufficient evidence to conclude that the strategy is being used.

Lower Certainty/Timing (5): Here the source or sources might not indicate the years in which the group begins or ends the use of a particular activity. The coding team again seeks to infer the

start and end dates for each strategy from other available information. If this is not possible, the relevant strategy variables are coded as zero, indicating insufficient evidence to conclude that the strategy is being used. Inferences should be clearly documented and explained in the relevant notes document.

High Certainty/Timing (6): For group years in which we record an exploitation strategy being used, this activity is documented by two or more credible and independent sources (such as academic articles, NGO reports, media stories). However, the sources are not clear on the specific start and end dates for the use of the strategy.

# 4. General Notes

## 4.1 Initiation and Termination of Activities

When a specific activity is mentioned in the source material, the rebel group is coded as engaging in that activity for the given year. In subsequent years, the rebel group is only coded as engaging in that same strategy if it is explicitly mentioned again. In other words, if extortion is used by a group in 1998, but then is not mentioned again until 2001, the intervening years are coded as no extortion occurring.

Note that variables are coded for the entire year. If a particular activity is used at any point during the year, the dataset records the appropriate value for the entire year.

## 4.2 Splinter Groups

In situations where an organization experiences a schism or breaks into smaller groups, the new factions are still considered part of the original organization, unless explicitly identified by UCDP as a new non-state actor. For example, the Rassemblement Congolais pour la Démocratie (RCD) in the Democratic Republic of the Congo split into two factions, RCD-Kinsangani and RCD-Goma, in late 1999. UCDP, however, continues to treat the organization as a unitary actor during the ensuing years. The coding for RCD thus includes activities undertaken by either or both factions as long as they are listed as a single group in the UCDP dataset.

# 5. References

Fearon, James D. 2004. Why do some civil wars last so much longer than others? *Journal of Peace Research* 41(3): 275-301.

Ross, Michael. 2012. *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations*. Princeton: Princeton University Press.