

# ***The From Empire to Nation-State dataset***

Andreas Wimmer and Brian Min

UCLA

This document offers a description of a new dataset created to test an institutionalist explanation of war, recently published by the *American Sociological Review* (Wimmer and Min 2006). The *From Empire to Nation-State* dataset is a territory-year dataset spanning 150 territories over the 186 period between 1816 and 2001. Traditional country-year datasets typically include only countries after they have become independent or recognized members of the nation-state system. This project collects data on territories both prior and after they became independent states. The dataset is made up of 28,162 territory-year observations, including a large number of territory-year observations in periods prior to nation-statehood. The unit of observation is territory, a geographic unit distinct from the state or government ruling over that territory at any point in time. We used the division of the world's states in 2001 as a territorial grid, extending these fixed geographic units back to the beginning of the data set in 1816.

The dataset includes three groups of variables: 1) A war dataset that includes all wars fought in the world since 1816, gives their location and classifies them into different types; this war dataset has also been described in a separate document;<sup>1</sup> 2) a group of institutional history variables that track the political institutions under which a territory is governed over time; 3) a group of basic variables such as GDP, population size, political regime type (democracy vs. autocracy), military strength of the center, oil production per capita, ethnic fractionalization, etc.

## **1) War data**

The dataset includes codings on 484 distinct wars fought across 619 territories. These 484 wars include 77 wars of conquest, 111 inter-state wars, 187 non-secessionist civil wars, and 109 secessionist civil wars. Codings of war onset and duration draw upon the work of the Correlates

---

<sup>1</sup> See <http://www.sscnet.ucla.edu/soc/faculty/wimmer/Whitehall.pdf>

of War (COW) project with substantial modifications and numerous additions. We describe these in detail here.

### *Adding wars*

The COW dataset includes wars only for states with diplomatic relationships with Britain or France prior to 1920, or for members of the League of Nations or the United Nations thereafter. To address this Western bias of the COW dataset, we added wars in territories that were not recognized by the Western powers and thus excluded from the COW dataset. This expanded our purview to include most of Latin America in the 19<sup>th</sup> century and most Asian and African territories in years prior to COW recognition as colonies or as independent states. A second bias in COW is produced by the fact that in colonial territories, only wars that involved the colonial power are included. This colonial center bias was overcome by specifically looking for civil wars that happened during periods of colonial rule.

We returned to the original sources of COW and other quantitative studies that have appeared over the past decades. We added 9 wars based on Richardson's "Statistics of Deadly Quarrels" (Richardson 1960). In addition, we drew upon Clodfelter's monumental "Warfare and Armed Conflict" (Clodfelter 2002) and added 24 wars from this source. We also reviewed OnWar.com, an amateur online website that lists a large number of wars with unsystematic but precious information about major lines of conflict. We went through this list, cross-checking new conflicts against other sources, and added 18 new wars from the 19<sup>th</sup> century after verifying date and battle death information from an additional independent source (usually on the internet). Finally, we also updated the list to 2001 adding 7 wars, relying on Gleditsch et al. (2002) and some of the revisions proposed by Gleditsch (2004). A total of 58 new wars were added from these sources.

We used COW's battle death threshold of 1000 per year to decide whether or not to include a war in our database. All newly identified wars for which battle deaths could not be confirmed were ultimately excluded from our list (a considerable number especially in Richardson's list).

We also cross-checked our list against multiple sources including World Military and Social Expenditures (Expenditures [various years] 1987:29-31), Butterworth's (Butterworth 1976) list of wars in the post World War II era, as well as a handful of less extensive lists (Licklider 1995; Miall 1992), none of which provided new information.

While we have made every attempt to develop the most comprehensive list of conflicts possible, there remain potential gaps in our coverage. The most obvious gaps are wars in pre-colonial eras.

Due to a lack of written historical sources and/or historical research for certain parts of the world, particularly regarding battle-deaths, our dataset likely suffers from a certain underestimation of the frequency of warfare in pre-colonial Africa and parts of Asia. We guess that some of the following wars may have reached the 1000 battle death threshold but are excluded from our list because of the lack of reliable estimates: the wars among Yoruba states in pre-colonial Nigeria; the civil wars in Ethiopia and Afghanistan during the middle of the 19<sup>th</sup> century; the wars connected to Buganda expansion in Uganda in the pre-colonial era; wars between the khanates of Central Asia before the Russian conquest.

Some wars of conquest may also have been missed because accurate death toll counts of the local populations are rare. In the first versions of COW, only deaths of imperial forces were taken into consideration. Later versions included the death toll of local fighters but it is unclear the extent to which previously excluded wars were added retrospectively (cf. Sambanis 2004). Our impression is that this was not done systematically, and we have added a handful of such wars where we came across well documented cases.

#### *New coding for the location of wars*

The COW dataset of wars provides information on the state participants involved in a war but offers no direct information on the location of these wars. In order to address this actor bias and the lack of information on where a war actually took place, we added locational codes for all wars, taking the current division of the globe into states as the grid on which to map each war. The OnWar.com database and Clodfelter's list were our main sources for determining battlefield locations; where necessary, we did Internet searches to find information on the geographical locations of major battles. If battlefields were located on more than one territory (such as during the Russian revolution), we coded multiple locations. Following the coding rules that COW used for determining who counts as a participant in a war, a territory was coded as a war location where at least 100 died in battle or 1000 troops were actively engaged.

There were a few cases where we diverted from these coding rules: in some civil wars, the forces that aim at overthrowing the government may set up a base of operation outside the territory of the country. These bases may come under attack by across-the-border operations of government forces. We decided that such cross-border pursuits did not justify adding a second location to the war (this was relevant for the civil wars in Nicaragua, Angola, Zimbabwe and Turkey).

The locational coding produces some oddities, mainly in cases where expanding empires meet outside of their core territories, vying for control over a region that has no local force strong enough to participate in battle. According to the territorial logic, this war is then related not to the two empires, but to the territory on which the battles took place. The cases are the Russo-Japanese war of 1904 which is attributed to China (and not to Russia or Japan), the Russo-Persian war of 1826 which is coded as relating to Armenia, Afghanistan, and Turkey (but neither Russia nor Persia), the Italo-Ethiopian war of 1887 which is related to Eritrea (not to Ethiopia), Russia vs. Central Asian Rebels of 1931 which is coded as a war in China. We used these locational classifications for the sake of consistency.

### *Reclassification*

Most datasets of war, including the COW dataset, classify wars based upon the type of actors involved without regard to the political goals of those actors. Thus an inter-state war is one involving two sovereign state actors while a civil war involves a sovereign state actor against a domestic non-state actor. For certain theoretical purposes, one might be less interested in the status of participants in the Western-centered international system and more in the political goals they seek to achieve: building a nation-state, enlarging the domain of an empire, gaining power in an existing state, etc.

In our dataset, inter-polity wars can either be wars of conquest or interstate wars. Civil wars can be either be non-secessionist or secessionist civil wars, with secessionist wars subdivided into non-nationalist and nationalist wars of secession. In addition, we also divided civil wars into ethnic and non-ethnic wars following the same criterion used by Fearon and Laitin (2003) in their post-1945 civil war dataset.

Our typology of wars relates in the following way to the old COW war types:

Main types	Inter-polity wars		Civil wars			
Sub-types	Wars of conquest	Inter-state wars	Secessionist civil wars		Non-secessionist civil wars	
Definition of sub-types	Expansion of state territory, permanent incorporation of new territories and populations; resistance against such expansion.	Fight between states over borders and territory, regional hegemony (but without aim of permanent incorporation as in wars of conquest)	Fight against the political center with the aim to establish an independent state		Fight between groups, at least one of which represents the central government, over domestic power relations, degree of autonomy of provinces or ethnic groups, tax burden, dynastic succession etc.	
Sub-sub-types			Non-nationalist wars of secession	Nationalist wars of secession	Ethnic civil wars	Non-ethnic civil wars
Definition of sub-sub-types			Fight for a separate, non-modern state (an independent khanat, sultanat, kingdom, tribal confederacy)	Fight for a separate, modern nation-state	Lines of conflict defined in ethnic terms and/or significant recruitment on the basis of ethnic networks	Lines of conflict not defined in ethnic terms and no recruitment on the basis of ethnic networks
COW category that corresponds most closely	Imperial wars	Inter-state wars, but with some reclassifications into wars of imperial conquest if war goal is permanent absorption of enemy territory into empire	Colonial wars, if aim is founding of a pre-modern state; some wars from civil war category added	Colonial wars, if aim is founding of an independent national state as opposed to less taxes, changes in administrative structures, re-installation of privileges etc. Some wars added from civil war category if goal is independent national state	Civil wars but with some wars added from colonial war category, if war goal is reduction of taxes, changes in administrative principles, re-installation of privileges etc.	

The re-classifications that were necessary to arrive at this new classification were numerous and are discussed briefly here. First, we treated non-colonial empires (the Ottoman; Habsburg; China; Romanov; Abyssinian empires) and Communist empires (the Soviet Union), in the same way as colonial empires (French, Portuguese, British, Dutch). Therefore, rebellions against Ottoman rule e.g. in the Balkans (the Greek, Serbian etc. “wars of liberation”) were classified in the same category as the anti-colonial wars in Algeria or Angola. By contrast, COW classifies the Cretan uprisings against the Ottomans as civil wars and the Algerian wars of national liberation as extra-state wars.

Secondly, we departed from COW’s distinction of two sub-types of extra-state wars, again basing our distinction on the political projects pursued by actors. “Nationalist wars of secession” were defined as rebellions against the political center (an imperial center or an already established nation-state) with the expressed aim of establishing a separate state representing a nationally defined people, thus conforming to the modern ideal of the nation-state. If the break-away

movement was not motivated by nationalist ideology, but rather pre-modern principles of political legitimacy (such as a khanate in Western China or the adjacent Russian territories; an independent emirate or sultanate in the Middle East etc.), we counted this as a non-nationalist war of secession. If, however, a rebellion against a political center was basically directed against certain specific aspects, such as laws that infringed on traditional rights, or new taxes, or direct administration by the center's officers and administrators, without challenging the borders of the existing state, we defined this as a non-secessionist civil war.

There is obviously a fine line between secessionist and non-secessionist civil wars, since many tax rebellions turned into nationalist wars of liberation, and many anti-imperial movements were composed of groups with different motives. The Druze rebellions against the French in Lebanon from 1925 to 1927, which were initiated by Druze mountain tribes resisting direct administration and later joined by Arab Syrian nationalists, is a case in point. Another problem is that it may be difficult to distinguish whether the demand for a new state is driven by nationalist or non-nationalist motives. For example, was the semi-independent Bosnian province under a Bosnian Vizier demanded by Christian rebels from the Ottoman Sultan in 1836 a modern nation-state? Did the independent Khanate that Muslim rebels fought for in 1863 in China or the re-installation of the Mogul during the so-called Indian mutiny in the middle of 19<sup>th</sup> century represent nationalist ideals? We decided on the basis of contextual and historical judgment whether or not the nationalist elements were dominant in these and some other borderline cases. We also had to distinguish between cases where the demand for independence was of a tactical nature (the Karen's threat to establish an independent state) or represented a long-term strategic objective (such as when a coalition of leaders from various ethnic groups headed by Uygurs established the short lived "Islamic Republic of Eastern Turkistan" in what is today western China).

The second category of extra-state wars in the COW dataset is wars of imperial conquest. These are wars fought by an imperial power seeking to enlarge the territory under its control by incorporating a territory as a dependent entity into its imperial domain. Attempts at "pacifying" the hinterland (such as the wars in Libya against the Sanusi tribes in the 1920s) are also coded as wars of imperial conquest. In sum, 54 wars that COW defines as "state vs. independent non-state", a sub-category of extra-state wars, were included in this new category of wars of imperial conquest. We also added many wars that COW had categorized as inter-state wars, since many such wars were fought against imperial encroachments and ended, in all but a handful of cases such as in Afghanistan, with the defeat of independent kingdoms or tribal confederacies and their

incorporation into an imperial domain. We reclassified 9 wars from “inter-state” to “wars of imperial conquest”. Not that since our units of observation are territories, the two World Wars are treated as a series of different war episodes, and we determined for each of them the most appropriate classification. The wars connected to Hitler’s occupation of Eastern Europe, to give an example, was coded as wars of conquest, while the battles in England were coded as resulting from an inter-state war, since there was no plan in the German Generalstab to conquer and permanently incorporate the British state into the domains of the new Reich. All of the above reclassifications and additions resulted in 484 total wars including 77 wars of conquest, 111 inter-state wars, 187 non-secessionist civil wars, and 109 secessionist civil wars.

#### *War rate calculation*

We assigned a 1 in the year of war onset and a 0 in all other years. In addition, we coded an ongoing war variable to equal 1 in all years in which a war was fought and a 0 for all years of peace. Most of the analysis in the ASR paper focuses on the war onset variable rather than the ongoing war variable.

In order to calculate war onset rates as in Figures 4 and 5 of the ASR paper, all war onset years needed to be related to a unique episode of imperial incorporation or nation-state creation. For example, Algeria is incorporated into the French Empire in 1848 and becomes an independent nation-state in 1963. The Franco-Algerian war of 1954 was thus identified to occur 9 years before nation-state creation and 106 years after imperial incorporation. In some cases, wars occur in territories with more than one instance of imperial incorporation or nation-state creation. For example, the Iraqi-Kurdish war of 1961 occurs both after Iraq’s absorption into the British Empire from 1914-1932 and after its incorporation into the Ottoman Empire from 1531-1913. In this case and as a general rule, the war is related only to the most recent episode of institutional transformation, and is thus coded as occurring 47 years after imperial incorporation. Exceptions were made only if a war was causally responsible for bringing about a subsequent transition, in which case we assigned the war to the latter. This was the case for wars of conquest, which led to imperial incorporation, and a number of nationalist wars of liberation, which led to the establishment of nation-states.

## **Institutional history variables**

### *Date of imperial incorporation and nation-state creation*

We first coded the years of the incorporation into empire and of nation-state creation (as sometimes different from the juridical date of independence, which is coded in the COW dataset). Of the 156 territories in our dataset, 140 territories were incorporated into an empire (and 92 during the temporal range of this dataset, 1816-2001), and 150 experienced nation-state creation. In order to determine the year in which a territory was considered to be part of a larger political entity (usually an empire), we searched for evidence of one of the following and coded the year of incorporation to whichever came first:

- The territory is effectively administered by an occupying force
- A garrison is established that aims at expanding military control over the territory
- The territory becomes a protectorate or colony.

The establishment of military posts that serve only to provide military protection to foreign traders, however, was not treated as a case of imperial incorporation. Temporary military occupation that lasted three or fewer years and that was not intended to permanently “absorb” the occupied territory into the state was not coded as imperial incorporation and were coded as periods of military occupation. Some territories have been conquered by multiple empires; some were governed by several empires contemporaneously. These complexities had to be recorded in our imperial history data file.

Nation-state creation is coded as the year in which a state begins to be self-governed in the name of a nationally defined people and no longer according to dynastic or religious principles. Two closely intertwined elements of the national principle are crucial: the state rules in the name of a nation and it is able to govern effectively without foreign intervention. More precisely, 1) a nation-state has a written constitution that a) defines a core national group, b) introduces equality before the law for all members of the nation, and c) provides for “popular rule” by some form of elected body; and 2) The state had to have *de facto* control over its foreign policy. Both criteria had to be fulfilled in order to define a polity as a modern nation-state. The U.S., Liberia, and South Africa were treated as nation-states as soon as equality before the law was finally introduced, i.e. after abolition of slavery, the introduction of voting rights for the indigenous population or the black majority respectively. Some states still are not modern nation-states according to this definition (Saudi Arabia, Bhutan, Brunei).

The “control of foreign policy” criterion proved to be the most problematic. It is a matter of historical judgment and definitional precision to decide how much control a state must have over its foreign policy to be classified as sovereign, given various forms of shared sovereignty across history, such as Canada’s dominion status in the British empire, the quasi-dominion status of Zimbabwe, the quasi-independent foreign policy of Egypt when it was still under the strong influence of Britain though still part of the Ottoman empire. The shared sovereignty during the transition period in many decolonizing states also produces ambiguity: was Cambodia’s “fifty percent independence” that the French granted from 1950 onwards enough to classify it as a sovereign state or is it rather in 1953 when legal independence was reached? We decided to regard dominion status (or quasi-dominion status) as providing “enough” control over foreign policy, but think that full independence in situations of decolonization is necessary to consider a state as sufficiently sovereign.

Some states have experienced several episodes of nation-state formation, interrupted by new episodes of colonization (e.g. Haiti, the Dominican Republic, and the Baltic and Caucasian states). If a modern nation-state split into two or more separate nation-states (Czechoslovakia, Pakistan, the Central American Republic), we coded a new episode of nation-state formation for both territories if the split-away territory comprised at least one third of the entire population. A reunification with a change in population of at least one third also was considered as the formation of a new nation-state (Yemen, Germany, and Vietnam). A nation-state had to exist at least three years to enter our dataset (we thus excluded the Kurdish Republic of Mahabad and the Republic of Eastern Turkistan and a couple of other states).

#### *Types of institutional rule over a territory*

The coding of imperial incorporation and nation-state creation together with the coding of some additional types of polities allowed us to construct a full institutional history for each territory from 1816 to 2001. Figure 1 that shows the portion of the world’s surface governed by different institutional principles over time. The file uses the following classificatory grid. It results from cross-tabulating types of institutional rule over a territory (nation-state principles, imperial principles, military occupation, “other”) with the political status of a territory, i.e. whether it is governed autonomously or whether it is ruled as a dependency. As the following table shows, only seven out of the sixteen possible categories were used for codings since there were no (or only very few) empirical constellations that would have fit into the other types.

Political Status \ Institutional principles		Nation-state	Empire	Other	Military occupation
Autonomous		<i>Autonomous nation-states</i> (e.g. France)	<i>Imperial centres</i> (e.g. Turkey under the Ottomans)	<i>Other autonomous states</i> (Buthan, Saudi Arabia, Somalia)	--
Dependent	... on a nation-state center	<i>“Internal colonies”</i> (e.g. Georgia under Soviet rule)	<i>Colonial dependency</i> (e.g. Algeria under the French)	-- <sup>2</sup>	<i>Militarily occupied territories</i>
	... on an imperial center	-- <sup>3</sup>	<i>Imperial dependency</i> (e.g. Bosnia under Ottoman rule)		
	... on a center governed by “other principles”	-- <sup>4</sup>			

The distinction between imperial and nation-state principles of governing a territory is consistent with our definition of the turning points. As soon as a territory experiences “imperial incorporation”, it is supposed to be governed by the institutional principles of empire. Based on Eisenstadt (1963:10-24) and Howe (2002:13-20), empire is defined by the following features: centralized bureaucratic forms of government, the domination of a core region over peripheries, an ethnically or culturally defined hierarchy between rulers and ruled, and claims to universal legitimacy—whether referring to a revolutionary ideology (as of the Soviet Union), a *mission civilisatrice* (as of colonial empires) or religious conversion (such as of the Spanish empire). As soon as a territory experiences nation-state creation, it is by definition ruled according to nation-state principles. These include centralized bureaucratic forms of government, uniform rule (without an institutionalized differentiation between core and periphery on the territory in question), equality of citizens (replacing hierarchy in empires) and government in the name of a bounded national community (rather than some universal aspiration). Note that the nation may be

<sup>2</sup> In this category we would subsume traditional client states of nation-states, empires, or other traditional polities. We avoided using this category since it is difficult to determine at which point a political alliance between a stronger and a weaker state makes such states dependent territories (e.g. Korea and China).

<sup>3</sup> Would be a dependent part of an empire that is governed like a nation-state, perhaps Hong Kong under China, if we would code China as an empire.

<sup>4</sup> Would be a dependent part of an “other” type of center which rules the territory according to nation-state principles, imperial principles, or “other” principles (Central Asia under the Mongols; Crete under Venetians).

imagined as multi-ethnic and multi-religious, as in Switzerland or India, or as mono-ethnic and – religious, as in France and Japan.

All territories governed by other institutional principles were assigned to the category of “other.” Absolutist kingdoms know centralized bureaucracies, as do nation-states and empires, but lack the center-periphery structures and the universalist forms of legitimacy of empires. In contrast to Eisenstadt (1963, chap. 1) and in line with Stephen Howe (2002), we exclude the absolutist kingdoms and principalities of Western Europe from our definition of empire. We did not want to assign Wurttemberg before Bismarck or the Papal State before Garibaldi to the same category as imperial China or the Spanish empire. In contrast to nation-states, absolutist states are not based on the equality of all citizens—which makes a difference even if a nation-state is ruled by a dictator with the powers of an absolutist king.<sup>5</sup> Feudal states, tribal confederacies (such as the Sanusi of Libya), city states (Switzerland before 1848), and patrimonial empires (e.g. the Tukolor or Mongol empires) all lack centralized bureaucracies.

Note that we exclude “informal empires” (Mann 2006) such as of the contemporary United States or the dispersed hegemonic “empire” of Hardt and Negri (2000) from our definition of empire, since these are not politically coherent entities. Note also that following the territorial logic on which our project is based, we coded the political institutions governing a particular *territory*, not those of entire states. Thus, the territory of Great Britain is classified as a nation-state, even while it was the core of a large colonial empire. The territory of the contemporary United States is a nation-state, even while Guam is governed according to imperial principles.

## **Other basic variables**

### *Types of political regimes (democracy versus autocracy)*

Our dataset contains four regime type dummy variables for democracy, autocracy, anocracy, and anarchy with only one of these dummies equal to 1 in any given territory-year. We relied primarily upon the 20-point scores from the Polity IV project, using the standard +6 and -6

---

<sup>5</sup> Our definition of the nation-state is based on the broad typologies developed in political sociology, rather than on the regime types of political science (democracy, autocracy, etc.). We thus assume that the difference between nation-states and absolutist states asserts itself even within the same political regime type: modern dictators such as Idi Amin cannot rule in the same way as Louis XIV; they cannot rely on dynastic legitimacy, but instead have to show that their government benefits “the people” of Uganda (e.g. by expelling Indian traders as “parasites” from the national home).

cutoffs to distinguish between democracies and autocracies. In order to identify the regime type of pre-independent territories, we diverted from Fearon and Laitin's procedure (Fearon and Laitin 2003) and coded all colonial dependencies as autocracies upon discovering that by following the the Polity IV coding rules, we would never arrive at an anocracy or democracy value for these regions.

For non-colonial imperial territories (governed by the land-based classical empires of Austria-Hungary, the Ottomans etc.), the dependent territories received the same regime classification as the imperial center. We followed this procedure also for the settler colonies of Canada, New Zealand and Australia as soon as these territories became part of the British Empire.

All territories that were neither empires nor independent nation-states were given an individual code, choosing between democracy, anarchy (no central government or no statehood at all, i.e. the -77 polity code), autocracy (traditional states such as khanates, emirates etc.), or anocracy (e.g. the Swiss republics before 1848). For pre-colonial African territories, we relied on Müller's (1999) atlas of pre-colonial cultures, which aggregates ethnographic data on political systems to decide between anarchy (absent of very weak statehood; Polity code -77) and autocracy (presence of a governing central state). All "simple states without social classes", "feudal states", "complex states" were classified as autocracies, while territories with only low or no political centralization were defined as anarchies.

We defined years of military occupation (Polity code -66) as autocracy, consistent with our way of coding imperial territories. For years of transition (Polity code -88), we interpolated Polity scores and identified the most appropriate regime category by year.

### *Change in military personnel*

To estimate changes in the military strength of the political center, we created a Military Personnel Change variable that calculated the percent deviation of the current year's military personnel level from the average level over the prior five-year period. We relied on COW's National Material Capabilities dataset to develop an estimate of the number of military personnel in each territory.

Unfortunately, COW's data do not include numbers for separate colonial armies such as the British Indian army and count only those under the direct command of the motherland's government. Still, it is the best available measurement of the capacity of an imperial center to suppress rebellions in their dependent territories.

For independent nation-states, this variable provided an estimate of the change in the state's domestic military strength. For colonies and imperial dependencies where no COW data were available, we calculated the change in military strength of that territory's imperial ruler, thus providing a rough estimate of the ability of the colonial power to suppress rebellion.

#### *GDP and population size*

By far the most extensive and reliable historical data on income and population come from Maddison (Maddison 2003). Our dataset includes a GDP/capita variable and a population size variable. For the Soviet and Yugoslav successor states, the Czech Republic and Slovakia, Pakistan and Bangladesh we took the earliest available data for the territories that were to become separate states, calculated their shares of the undivided countries, and then extrapolated back into the past until the earliest data available for the undivided country. We excluded GDP data for Turkey up to 1923 because they seemed to relate to the entire empire. Germany's data reflect various borders over time; no data series for Germany in its post-1990 borders are available yet.

#### *Political Discrimination*

The Minorities at Risk (MAR) dataset collected by Ted Gurr and collaborators is the best source available for information on the political status of ethnic groups since World War II. However, it has the disadvantage, as noted by Fearon (Fearon 2003:196), of including only groups that have either shown some minimal degree of political mobilization and/or that have been substantially discriminated against. For our purposes, we assume that the MAR group list provides a reliable census of groups that have been discriminated against and that groups that do not appear in their list have not suffered from discrimination nor have they been the basis for political mobilization. For each of 284 "at risk" minority groups, the MAR dataset provides an annual political discrimination score running from 1 (underrepresentation addressed by affirmative action) to 4 (an Apartheid type situation). Groups not represented in the MAR dataset are assumed to not be the victims of description and imply a 0 score. Using these group-level discrimination scores and group-level population data, we created a population weighted-average Discrimination Score for all territories and scaled the score to a 100-point scale. Territories with no groups reported in the MAR dataset received a 0.

### *Oil production per capita*

We generate an oil production per capita variable based upon historical data (Mitchell Various years) that includes annual oil production estimates for a comprehensive cross-section of oil-producing states and colonies. Many studies use either a dummy variable for oil exporter or calculate the share of oil exports to GDP. However, a per capita figure represents an improved operationalization since it is not dependent on the strength of other economic sectors, as are the percent of GDP figures, and the risk of collinearity is reduced considerably compared to dummies (cf. Humphreys 2005).

## Bibliography

- Butterworth, Robert Lyle. 1976. *Managing Interstate Conflict, 1945-1974*. Pittsburg: University Center for International Studies.
- Clodfelter, Michael. 2002. *Warfare and Armed Conflicts. A Statistical Reference to Casualty and Other Figures, 1500-2000*. Jefferson: McFarland.
- Eisenstadt, Samuel N. 1963. *The Political Systems of Empires*. New York: Free Press.
- Expenditures, World Military and Social. 1987. *World Military and Social Expenditures, vol. 11*. Leesburg: WMSE Publications.
- Fearon, James D. 2003. "Ethnic and cultural diversity by country", in *Journal of Economic Growth* 8 (2):195-222.
- Fearon, James D. and David D. Laitin. 2003. "Ethnicity, insurgency, and civil war", in *American Political Science Review* 97 (1):1-16.
- Gleditsch, Kristian. 2004. "A Revised List of wars between and within Independent States, 1816–2002." *International Interactions* 30:231–62.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Erikson, Margareta Sollenberg, and Håvard Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39:615–37.
- Hardt, Michael and Antonio Negri. 2000. *Empire*. Cambridge: Harvard University Press.
- Howe, Stephen. 2002. *Empire. A Very Short Introduction*. Oxford: Oxford University Press.
- Humphreys, Macartan. 2005. "Natural resources, conflict, and conflict resolution. Uncovering the mechanisms." *Journal of Conflict Resolution* 49:508-537.
- Licklider, Roy. 1995. "The consequences of negotiated settlements in civil wars, 1945-1993", in *American Political Science Review* 89 (3):681-690.
- Mann, Michael. 2006. "American Empire: Past and Present". Paper given at the conference Comparative Social Analysis Workshop, Department of Sociology, UCLA.
- Maddison, Angus. 2003. *The World Economy: Historical Statistics*. Paris: OECD.
- Miall, Hugh. 1992. *The Peacemakers. Peaceful Settlement of Disputes since 1945*. New York: St. Martin's Press.
- Müller, Hans-Peter. 1999. *Atlas vorkolonialer Gesellschaften. Sozialstrukturen und kulturelles Erbe der Staaten Afrikas, Asiens und Melanesiens. Ein ethnologisches Kartenwerk für 95 Länder mit Texten, Datenbanken und Dokumentationen auf CD-ROM*. Berlin: Reimer.
- Richardson, Lewis Frey. 1960. *Statistics of Deadly Quarrels. Edited by Quincy Wright and C.C. Lienau*. Pittsburgh: Boxwood Press.
- Sambanis, Nicholas. 2004. "What is civil war? Conceptual and empirical complexities of an operational definition", in *Journal of Conflict Resolution* 48 (6):814-858.
- Wimmer, Andreas and Brian Min. 2006. "From empire to nation-state. Explaining wars in the modern world, 1816-2001", in *American Sociological Review* 71 (6)

## Waves of War Dataset 1.0

### Variable List

Variable	Description	Source	Number of observations	Mean	Minimum value	Maximum value
yearc	Year-territory		28162			
year	Year		28162		1816	2001
country	Territory name		28162			
onset	War onset on territory		28162	0.0215894	0	1
war	War on territory		28162	0.076486	0	1
waname	Name of war		2151			
warno	War number	COW and author codings	2151			
wartype	Type of war		2151			
yrbeg	Beginning year of war		2151		1816	1998
yrend	Ending year of war		2151		1817	2001
anarc	Anarchy dummy	Author calculations from Polity IV	28162	0.0148072	0	1
anarcl	Lagged anarchy indicator	Author calculations from Polity IV	28012	0.0148151	0	1
anoc	Anocracy dummy	Author calculations from Polity IV	28162	0.2750515	0	1
anocl	Lagged anocracy indicator	Author calculations from Polity IV	28012	0.2751321	0	1
area2001	Size of territory in square kilometers in 2001	World Bank Development Indicators	27976	885011.8	9250	1.71E+07
asia	Asia region dummy	Fearon & Laitin (2003)	27790	0.1437208	0	1
autoc	Autocracy dummy	Author calculations from Polity IV	28162	0.586393	0	1
autocl	Lagged autocracy indicator	Author calculations from Polity IV	28012	0.5886406	0	1
cowcode	COW code for territory	COW Project	28162	451.854	2	920
democ	Democracy dummy	Author calculations from Polity IV	28162	0.1234642	0	1
democl	Lagged democracy indicator	Author calculations from Polity IV	28012	0.1211624	0	1
eeurop	Eastern Europe region dummy	Fearon & Laitin (2003)	27790	0.1815761	0	1
ethfrac	Ethnic fractionalization	Fearon & Laitin (2003)	27232	0.4175961	0.001	0.92
gdp	GDP (million 1990 International Geary-Khamis dollars)	Maddison, World Economy: Historical Statistics	16972	69914.24	28	7965795
gdppc	GDP per capita	Maddison, World Economy: Historical Statistics	17075	2741.089	201.83	43806.46
gdppcl	Lagged per capita gdp	Maddison, World Economy: Historical Statistics	16932	2712.469	217.83	43806.46
implag	Years to/from imperial incorporation	Author calculations	16045	83.8205	-129	753
imppower	Name of imperial power	Author codings	15286			
instab	Change in regime type in prior 2 years	Author calculations	28162	0.0606846	0	1
instabl	Lagged instability indicator	Author calculations	28012	0.0605812	0	1
lamerica	Latin America region dummy	Fearon & Laitin (2003)	27790	0.1544081	0	1
lmtnest	ln(mountainous terrain)	Fearon & Laitin (2003)	27232	2.217459	0	4.55
lnpop	ln(population)	Maddison, World Economy: Historical Statistics	21348	8.452219	1.504077	14.05876
lnpopl	Lagged ln(population)	Maddison, World Economy: Historical Statistics	21205	8.445625	1.504077	14.04987
milperc	Deviation from 5-year average of governing military's personnel	Author calculations from COW National Material Capabilities 3.0	22742	0.0200949	-1	4

milpercl	Lagged deviation from 5-year MA of governing military personnel	Author calculations from COW National Material Capabilities 3.0	22600	0.0203457	-1	4
nafrme	North Africa-Middle East region dummy	Fearon & Laitin (2003)	27790	0.0900324	0	1
nbcivil	Number of neighboring civil wars	Author calculations	28162	0.134685	0	6
nbconq	Number of neighboring wars of conquest	Author calculations	28162	0.0574178	0	6
nbinter	Number of neighboring interstate wars	Author calculations	28162	0.094205	0	10
nbnatind	Number of neighboring civil wars of nationalist independence	Author calculations	28162	0.072225	0	15
nbnonind	Number of neighboring civil wars of non-nationalist independence	Author calculations	28162	0.0252823	0	3
nsflag	Years to/from nation-state creation	Author calculations	28162	-16.74554	-182	352
nsfyear	Year of nation-state creation	Author codings	28162	1925.324	1649	1998
ocivil	Onset of non-secessionist civil war	Author calculations	28162	0.0067112	0	1
oconq	Onset of war of conquest	Author calculations	28162	0.0031603	0	1
oil	Oil production (thousand metric tons/yr)	Mitchell, International Historical Statistics	28162	3405.529	-827.5	557320
oilpc	Oil production per capita	Author calculations	28133	0.347498	-0.01	202.76
oilpcl	Lagged oil production per capita	Author calculations	27985	0.3427407	-0.01	202.76
ointer	Onset of inter-state war	Author calculations	28162	0.0072438	0	1
ointrap	Onset of intra-polity war (civil + natind + nonind)	Author calculations	28162	0.0111853	0	1
onatind	Onset of nationalist war of independence	Author calculations	28162	0.0034799	0	1
ononind	Onset of non-nationalist war of independence	Author calculations	28162	0.0009942	0	1
pdemnb	Percent of neighbors democratic	Author calculations	28162	0.0543914	0	1
pocivil	Prior ongoing non-secessionist civil war	Author calculations	28162	0.0192103	0	1
poconq	Prior ongoing war of conquest	Author calculations	28162	0.0067467	0	1
pointer	Prior ongoing inter-state war	Author calculations	28162	0.0127477	0	1
pointrap	Prior ongoing intra-polity war (civil + natind + nonind)	Author calculations	28162	0.0352958	0	1
poldisc	Political discrimination (population-weighted score)	Minorities at Risk project	5252	16.28891	0	89
poldiscl	Lagged political discrimination index score	Author calculations	5139	16.38289	0	89
ponatind	Prior ongoing nationalist war of independence	Author calculations	28162	0.0138129	0	1
pononind	Prior ongoing non-nationalist war of independence	Author calculations	28162	0.0022726	0	1
ponset	Prior ongoing war	Author calculations	28162	0.0548967	0	1
pop	Population (thousands)	Maddison, World Economy: Historical Statistics	21349	20262.41	0	1275392
relfrac	Religious fractionalization	Fearon & Laitin (2003)	27232	0.3757753	0	0.78
ssafrica	Sub-Saharan Africa region dummy	Fearon & Laitin (2003)	27790	0.2893127	0	1
western	Western region dummy	Fearon & Laitin (2003)	27790	0.14095	0	1