**Oona A. Hathaway & Scott J. Shapiro**

**Conquest and State Size Database**

**Summary of Data Sources and Technical Documentation**

The goal of the project is to create a dataset to help us understand the changing nature of territorial conquest from 1816 to 2014.

Please cite the data as follows: Oona A. Hathaway & Scott J. Shapiro, Conquest and State Size Database (2017), http://www.theinternationalistsbook.com/data.html. This data is analyzed and described at length in Oona A. Hathaway & Scott J. Shapiro, *The Internationalists* (Simon & Schuster 2017), Chapters 13 and 14.

# **Correlates of War’s Territorial Change (TC) data**

The construction of our analysis dataset begins with the Correlates of War’s (COW) Territorial Change (TC) data. The dataset is an attempt to “identify and code all territorial changes involving at least one nation-state . . . for the period 1816-2014.” The data consist of 838 observations and include detailed information about the conflict and political entities involved. In particular, the data are categorized by process of transfer, including conquests, annexations, cessions, secessions, unifications, and mandated territory. Transfers are also coded according to whether they constitute an entry to/exit from the state system or independence, as well as whether there was a military conflict associated with the transfer. Finally, data on conflicts include information on the size of the territory and an estimate of the population occupying the territory at the time of the transfer.[[1]](#footnote-1)

**Corrections to the TC data**

While working with the data, we discovered several irregularities, which appear to be mistakes in the underlying TC data. We perform a number of corrections to account for errors in estimated territory size and recording of the gaining/losing entities. For example, our version of the territorial change data provides a more complete picture of the break-up of Germany following World War II and its subsequent reunification.

**Identifying Conquest in the TC data**

In order to properly analyze military conquest, we narrow our analysis to the subset of the TC data that are (1) associated with a military conflict or (2) were otherwise coded as a conquest by COW.

# **Coded Data**

In order to properly understand the changing nature of territorial conquest, we constructed seven legal categories in the TC data (**Q1-Q7**), described below. In addition to **Q1-Q7**, the data include the following: **o\_n** is an internal tracking number, **tc\_number** corresponds to the Correlates of War Territorial Change data ID, **o\_coder1** and **o\_coder2** identify the two coders, each of whom coded the transfer independently. Where the two coders disagreed, the disagreement was resolved in conversation with the authors and the team of coders.

**Q1. Sovereignty in Dispute.** Was the losing entity’s sovereignty over the transferred territory in dispute at the time of transfer?That is, did other states not recognize the sovereignty of the losing entity over the transferred territory? The available answers are (0) no; (1) yes; and (9) unclear.

**Example 1:** Suppose every state in the international system recognizes State A’s control over a parcel P of land. A few years later, State B invades A and takes control of P. The answer would be “no” (0).

**Example 2:** Suppose few or no states in the international system recognize State A’s control over a parcel P of land (perhaps because it was taken in a previous illegal conquest). A few years later, State B invades A and takes control of P. The answer would be “yes” (1).

**Example 3:** In cases involving transfers from a non-state entity to a state, the answer is “yes” (1), because entities that are not recognized as states cannot have recognized state sovereignty over territory.

**Example 4**: The COW may list the wrong losing entity. For example, it is possible that the COW lists State A as the losing state, but in fact State A was did not have independent legal personality—it instead is properly understood as a sub-entity of State B and therefore State B is properly understood as the losing state, not State A. (Indicia of independent legal personality include the ability to make laws that cannot be overruled, the capacity to enter into treaties, possessing a military (that is, capacity to project force outside the territory).) In such cases, code as if State B were listed as the losing state.

Note: If a small portion of the territory (less than 1%) is in dispute, this variable should be coded “no” (0).

**Q2. Reversion.** Was the transfer a reversion of the same land previously exchanged between the same countries? (For example, in 1945, Ethiopia regained its territory from Italy. This was a reversal of the conquest of Ethiopia by Italy.) The available answers are (0) no; (1) yes; (2) independence; (9) unclear.

Note: If the area involved is within 10% or 10 square kilometers of the original transfer between the two countries, the prior transfer is a reversion (1), but if less than that, it is not (0).

**Q3. Multinational organization.** Was the intervention carried out by, or with the approval of, a multinational organization (e.g., the United Nations, NATO, the League of Nations)? The available answers are (0) no; (1) yes; (9) unclear.

**Q4. International recognition.** Was the transfer recognized by the international community? Since the project is concerned with the legitimacy of conquest under international law, it is important to note whether the international community recognized a territorial exchange. The available answers are (0) there is no evidence that any state other than those involved recognized the transfer (or the only evidence regarding recognition is of *refusal* to recognize); (1) one country other than those involved recognized the transfer; (2) less than a majority of countries other than those involved recognized the transfer; (3) the majority of countries recognized the transfer; and (4) nearly all countries recognized the transfer (if there was a treaty between the states involved in the transfer in which the losing state consented to the transfer, the default expectation is that the transfer should be placed in this category unless there is clear evidence to the contrary). If there is no formal or affirmative act of recognition, but all states appear to accept the transfer (for example, by continuing normal trading relations with the territory in a way that suggests acceptance of the transfer) and no state rejects the transfer or refuses to recognize it, then the episode should be coded a “4”.

Notes: Where there is a treaty in which the state losing territory agrees to the transfer of territory, the presumption is that all states will recognize the transfer. It should therefore be coded a “4.” An armistice is not a treaty for these purposes.

**Q5. Manifesto or Declaration**. In the course of their research, coders attempted to locate any manifesto or declaration issued during the conflict. Code options are (1) Manifesto or Declaration located and (0) not located. The manifestos located are posted here, together with a large number of additional manifestos: <http://documents.law.yale.edu/manifestos>.

**Q6. Fission.** Fission occurs when one state divides into two (or more) distinct states, each with its own non-overlapping territory, each politically independent from the other. Code options are (0) "no" if there is not fission, (1) "yes” if there is, and (2) “independence” if a part of the prior state wins its independence from the continuing parent state. Category (2) includes both colonial and non-colonial independence. Note that independence is a subcategory of fission.

For our purposes, independence is defined as any fission in which the pre-fission entity continues to exist in largely recognizable form. This is distinct from a non-independence fission, in which none of the post-fission entities can be characterized as the pre-fission entity (this can be true even if one or more of the new entities accepts the legal obligations of the parent entity).

Example 1: The 1991-1992 breakup of the Socialist Federal Republic of Yugoslavia is fission. Code “yes” (1).

Example 2: The breakup of Korea in 1948 into North and South Korea is fission. Code “yes” (1).

Example 3: The 2011 recognition of the Republic of South Sudan as an entity separate from Sudan in independence. Code “independence” (2).

Example 4: In 1939, Mussolini invaded Albania and declared the country part of the Italian Empire. Albanians resisted the Italian occupation, until it ended in 1943 and was replaced by German occupation. Resistance continued and was led by the growing Communist movement within the country. By the end of 1944, Albanian forces had regained control of most of the country and a provisional Communist government was set up. In the following years, Albania was able to regain the territory that had fallen to Italy. Code “independence” (2).

**Q7. Claim of Sovereignty.** States sometimes occupy another state's territory without claiming sovereignty over that territory. Article 42 of Hague IV (1907) states: “Territory is considered occupied when it is actually placed under the authority of the hostile army. The occupation extends only to the territory where such authority has been established and can be exercised.” An occupying power does not claim to have acquired sovereignty over the occupied territory. Coding options are as follows: (0) "yes" if there is a claim of sovereignty; (1) "no" if there is no claim of sovereignty; this includes cases where a state is occupied or partially occupied (except *de minimis* partial occupation); (2) “Multinational organization” if a state is occupied pursuant to a mandate from a multinational organization (League of Nations mandates, UN trust territories, and similar juridical institutions).

# **Tracking Reversions in the TC Data**

In an effort to understand nature of territorial conquest over time, we sought to understand the permanence of conquests. We tracked this information in two ways. First, coders were asked to track whether a particular territorial transfer is a reversion in Q2 (this is backward looking, tracking whether the current transfer is *itself* a reversion of an earlier transfer). In addition, we employ an algorithm to match transfers of territory longitudinally, to track whether a given transfer is later reverted (this is forward looking, tracking whether the current transfer is reverted *in the future*). The algorithm primarily relies on the following information:

1. the names of the gaining and losing entities to identify all subsequent conflicts in which a transfer may have been reversed.
2. the area of the territory transferred to determine whether the two transfers are similarly sized.

**Automatic assignment—**the algorithm considers a conflict to be reverted (reverting) if the gaining (losing) entity of the first matches the losing (gaining) entity of the second, and the area involved is within 10% or 10 square kilometers of the original transfer.

**Manual assignment—**following the automatic assignment, we use the “10 and 10” rule to create a list of conflicts without matching political entities to be considered for manual assignment.

For example, COW reports that the United Kingdom obtained 389,362 square kilometers from an unspecified entity in 1893 (transfer 340). COW also reports that, in 1980, Zimbabwe gains 390,662 square kilometers in its independence from the United Kingdom (transfer 819). Because Zimbabwe was not part of the state system in 1893, it would be inappropriate to code the original transfer as coming from Zimbabwe (which would make it a candidate for automatic assignment). However the independence in 1980 is clearly a reversion of the first transfer. Therefore the reversion is coded manually.

Additionally, we use the Q2 (reversion) codes from section 2 to check the type-I and type-II error of this procedure. This information was used to develop the “10 and 10” rule, but we also used it to correct information when the best version of the algorithm was reached. Conflicts between the reversions identified by the algorithm were reconciled, either by (1) correcting Q2 information in the legal categories data, (2) correcting apparent errors in the TC data, or (3) by creating exceptions to the “10 and 10” rule.

**Results of the matching.** Tables 1 and 2 show the results of the procedure. Automatic assignments make up the bulk of the identified cases. As noted above, all reversions were benchmarked against the reversion flag created by the coders, as described in section 2.

|  |  |  |
| --- | --- | --- |
| Table 1: Distribution of Reverted Assignments |  |  |
|  |  |  |
| Method by which transfer was identified as reverted |  | Count |
| Manual: Based on land match |  | 5 |
| Manual: Based on logical checks |  | 6 |
| Automatic: Land and names match |  | 34 |
| Not assigned: no reversion |  | 209 |
|  |  |  |
| Total |  | 254 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Table 2: Distribution of Reversion Assignments |  |  |
|  |  |  |
| Method by which transfer was identified as a reversion |  | Count |
| Manual: Based on land match |  | 2 |
| Manual: Based on logical checks |  | 9 |
| Automatic: Land and names match |  | 21 |
| Not assigned: no reversion |  | 221 |
|  |  |  |
| Total |  | 254 |
|  |  |  |

# **Categorizing Conquest in the TC Data**

We consider transfers to be military conquest if the following criteria are met.

1. The transfer was not a reversion of a previously unrecognized transfer (Q2=0 & Q4>2 for the reverted prior transfer).
2. The intervention not carried out by, or with the approval of, a multinational organization (e.g., the United Nations, NATO, the League of Nations) (Q3=0).
3. The transfer was not a result of independence. (Q6=0).
4. The gaining country claims sovereignty over the territory (Q7=1)

Additionally, we often distinguish transfers based on whether or not they were recognized by the international community (Q4>2), and whether or not they were later reversed, as described in section 3.

We did not distinguish between conquests of territory based on whether the sovereignty was in dispute at the time (Q1), though we provide the data in case others wish to do so. It may initially seem appropriate to exclude transfers of territory where sovereignty is in dispute from “conquest.” But doing so means excluding almost all colonial conquests, including the scramble for Africa. An actor like the Zulu Kingdom, for example, did not possess recognized sovereignty over its territory, since as a non-state entity it wasn’t recognized as a member of the state system. Indeed, the colonial empires justified their conquests by suggesting that areas controlled by non-state actors were *res nullius* and free for the taking.

# **Mapping the Evolution of State Size**

In addition to tracking territorial conquest, we also show the changing landscape of state size as independences undid the colonization and conquest and seizing territory via military conflict became unacceptable. In order to accomplish this, we use the territorial change data along with data from the World Bank, which reports the territorial size of states in 2014. We perform a match of the World Bank data to the Territorial Change data, and use the transfers to infer the size of all states in all years back to 1816.

Again, we perform logical checks to ensure that the results of the matching and backward territory calculation are sensible. For example, we check to ensure that no states have negative territory in any year. The results of these checks form the basis for some of the corrections to the TC data noted above.

# **Description of Software**

This section lists the code (Stata do-files) provided to replicate the results. All code is controlled by and can be run using KB\_main.do. In order to replicate the data and figures from the draft:

1. Download and unzip the provided files to a local directory
2. Change line 28 of KB\_main.do to match the local directory where the unzipped files are located
3. Run KB\_main.do.

The main function of each file is described briefly here. For more detail, see the comments provided within the code.

*KB\_main.do* – This is the main control for the entire project. To reproduce all results, simply edit the path directory and run.

*KB\_infsetup.do* – This file ensures that all directories exist and sets some simple macros for use throughout the project.

***Data Preparation***

*KB\_wbarea.do* – Cleans the World Bank territory data

*KB\_sfdata.do* – Cleans the data on state failures

*KB\_cleancow.do* – Control file for cleaning all COW data

*KB\_cleancow\_wars.do* – Cleans raw COW data on wars

*KB\_cleancow\_states.do* – Creates state system roster and data on state counts

*KB\_cleancow\_tc.do* – Cleans COW Territorial Change data

*KB\_cleancow\_stockwars.do* – Creates a dataset containing stock data on wars active in a given year

*KB\_riskdata.do* –Control file for creating conquest data

*KB\_riskdata\_combinetclc.do* – Combines territorial change data on conquests with legal categories

*KB\_riskdata\_idrev.do* – Controls algorithm to longitudinally identify reversions

*KB\_riskdata\_idrev\_matchmatrix.do* – Creates a matrix of all transfers to determine which might be reversion of the other

*KB\_riskdata\_idrev\_setup.do* – Creates empty reversion tags in the TC data to be filled by the algorithm

*KB\_riskdata\_idrev\_manuals.do* – Manually assigns reversions that are too complicated to be picked up by the algorithm

*KB\_riskdata\_idrev\_automatics.do* – Automatically detects reversions based on the 10 and 10 rule

*KB\_riskdata\_idrev\_checks.do* – implements logical checks to for type I/type II error

*KB\_riskdata\_idrev\_remainders.do* – Generates a list of conflicts that look close but are not assigned (candidates for manual assignment)

*KB\_riskdata\_conqflags.do* – Creates conquest flags as described in the documentation

*KB\_riskdata\_instate.do* – Creates a flag for whether the loser of a conquest was a state at the time of conquest

*KB\_ssdata.do* – Matches World Bank data to Territorial Change data to infer state sizes in a given year

***Analysis***

*KB\_statecount.do* – Generates a table and figure describing the number of states in the state system over time

*KB\_indcount.do* – Creates a figure and table showing the number of independences per year and decade

*KB\_sffig.do* – Generates a table and figure showing the number of state failures over time

*KB\_intrafig.do* – Generates a table and figure showing the number of active intra state wars over time

*KB\_riskfig.do* – Generates tables and figures describing the rate of conquest and area conquered over time

*KB\_ssfig.do* – Generates data and a figure plotting the evolution of state size over time

1. For more detailed information on the Territorial Change data, see the documentation at <http://www.correlatesofwar.org/data-sets/territorial-change>. [↑](#footnote-ref-1)