

## REVIEWS OF DATASETS

Ahmet Izmirliglu

### The Correlates of War Project

**Review of:** The Correlates of War Project

**Author/Compiler:** Zeev Maoz, D. Scott Bennett

**Repository:** <http://www.correlatesofwar.org/data-sets>

**Date of Posting:** 2017

**Size:** n/a

Political scientist J. David Singer founded The Correlates of War Project (COW) in 1963 to collect empirical data on large state conflicts and to systematically build scientific knowledge of war. Armed with a Carnegie Corporation grant, Singer and historian Melvin Small led an exploratory study of conditions leading to the outbreak of war. Their collaborative work was presented in two books, *The Wages of War* (1972) and *Resort to Arms* (1982). Despite some difficulties owing to the retirement of Singer in the late 1990s, the project has since thrived, becoming the largest and most influential source of data on war for the period 1815–2016. The dataset is currently administered in multiple repositories under the leadership of Director Zeev Maoz and Associate Director D. Scott Bennett.

COW consists of thirteen individual datasets that provide detailed information on diverse factors that may shed light on the nature of conflict among interstate and non-state entities.<sup>1</sup> Examples of topics addressed by the datasets include geographic location of interstate disputes in latitude-longitude coordinates, instances of threats and aggression displayed by states, national material capabilities, national and bilateral trade flows, and diplomatic representation.<sup>2</sup> The temporal range of the data is the post-Napoleonic era, with the majority of the datasets covering the early nineteenth century to the early twenty-first century. The tabular data in each dataset are organized into subcategories and presented in comma-separated values (CSV) files (and in some cases as MS Excel files), with supplementary information presented in PDF and RTF files. The variables in the CSV files—such as population, conflict duration, number of fatalities in the conflict, and geographic region of conflict—are organized into columns in the tables. There are 395 different variables available within the datasets.<sup>3</sup> As may be apparent from the list of topics and variables, the information is provided in a manner for researchers to easily discern possible explanatory variables that may be pertinent to their research. The inclusion of such additional data that can be readily adopted as explanatory variables is one of the definitive ways in which COW has evolved from its early iterations.

The tabular data in the CSV files are all encoded numerically. For example, the geographic location for a conflict may appear in the dataset as “2”, necessitating researchers to take care in identifying the correct supplementary file to discern between multiple usages of the same numerical value. This practice of numerical encoding of the information ties back to Singer’s original intent to study war empirically. Thus, the dataset suits the needs of quantitative research far more than those of qualitative research. Consequently, the works that cite COW tend to favor quantitative methodologies and modelling, and it is important to note that the dataset so far has mainly attracted researchers in international relations and policy analysis. For examples of such works that have cited COW, one can find bibliographies at two locations. The first is a very thorough list of works published by COW personnel from 1963 to 1990, available at [http://cow.la.psu.edu/COW\\_bibliographic\\_essay.htm](http://cow.la.psu.edu/COW_bibliographic_essay.htm).<sup>4</sup> This list is well organized into four categories that focus on theoretical, data generation, model and analysis, and practical implications of the research. The second resource is a general, less comprehensive list of works that have used COW data, which covers the years from 1963 to 2016; it is available at <https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00232>.<sup>5</sup>

Most of the thirteen datasets that comprise COW are currently hosted in U.S. academic institutions, with the sole exception of the Diplomatic Exchange dataset, hosted by the Koç University in Istanbul, Turkey. The decentralization of the datasets is a deliberate strategy by COW that they refer to as “coordinated decentralization.” This strategy is meant to economize the effort and costs of maintaining a massive dataset, while also encouraging interdisciplinary and international collaboration. The “home” (hosting institution) and “host” (individual who collects and maintains data) each agree to maintain their adopted dataset for 3–5 years. Both the “home” and “host” agree to abide by COW data-hosting guidelines and standards for the duration of their duties.<sup>6</sup> These procedures

ensure that even such a large amount of decentralized data remains subject to strict controls to maintain data integrity, as well as to ensure that all changes to datasets are diligently documented.

The guidelines clearly state the duties of hosts: regular revision and update of data, routines to keep track of errors and questions regarding data, procedures for publishing analytical results based on COW datasets, and maintenance of all documentation associated with datasets. Data Hosting Standards further delineate procedures to sustain data integrity. These include maintaining units of analysis across dataset versions, procedures for introducing new variables into datasets, and documentation of data sources. Additionally, Data Hosting Review Procedures add another level of quality control for internal dataset consistency, comparability and compatibility. While these extensive regulations may initially appear excessive, they are vital to the success of a large, decentralized operation like COW.

Although it was the first of its kind, COW is no longer the only game in town. The success of COW, as well as some perceived limitations in its definitions, inspired researchers worldwide to collect and maintain empirical data on war. For instance, the Uppsala Conflict Data Program (UCDP) began collecting data on low-level conflicts in the 1980s by focusing on conflicts with an annual fatality limit of 25 (rather than the COW's higher threshold of 1000 annual fatalities). Thus, UCDP began as an effort to improve on COW by stratifying conflicts among minor and major categories based on annual fatality rates. Interested readers will find links to a number of datasets on war and conflict in the COW web site, at <http://www.correlatesofwar.org/external-links/correlates-of-war-external-links>. Still, despite the increasing number of conflict datasets, COW stands out as a unique benchmark—both in terms of its broad scope and its decentralized structure. It should continue to inspire and support international collaboration in creative, and hopefully more widely interdisciplinary research for years to come.

## NOTES

---

<sup>1</sup> Although the project originally only included data for armed conflict among members of the interstate system, some datasets within the project now include data on armed conflicts involving non-state entities. A detailed overview of definitions and typology is provided at <http://www.correlatesofwar.org/data-sets/COW-war>, (accessed 8/24/17).

<sup>2</sup> <http://www.correlatesofwar.org/data-sets> (accessed 8/24/2017).

<sup>3</sup> <https://www.icpsr.umich.edu/icpsrweb/ICPSR/ssvd/series/232/variables> (accessed 8/24/2017).

<sup>4</sup> Accessed 8/24/2017.

<sup>5</sup> Accessed 8/24/2017.

<sup>6</sup> The guidelines and standards may be found at <http://www.correlatesofwar.org/data-hosting>, (accessed 8/24/2017).



Articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.