Codebook for the Correlates of War Project's Trade Data Set, Version 3.0

Katherine Barbieri, (Katherine.barbieri@sc.edu)
Omar M. G. Keshk, (keshk.1@osu.edu)
February 2012

The following guide provides a description of the Correlates of War Project's Trade Data, Version 3.0. The data set includes annual dyadic and national trade figures for states within the international system, as defined by the Correlates of War Project. The data set covers the period 1870-2009 and includes three data files: (1) dyadic trade statistics; (2) national trade statistics; and (3) this Codebook. The data file names are dyadic_trade.csv and national_trade.csv. The update to Version 3.0 extends the data to an end date of 2009—as opposed to the earlier end date of 2006 associated with Version 2.1. It also includes some corrections of typographical errors.

The data are distributed in a flat text format, with variables separated by commas. Given the .csv extension, the files will open in Microsoft Excel by default. Users should not use Excel to read the dyadic file, since the number of observations far exceeds the maximum number of rows that Excel will read. You should be able to import the text file into most statistical software packages or into Microsoft Access. Missing values in the main data sets are assigned a value of -9. Variable names, procedures, source codes, and descriptive statistics are provided below for each data file.

The data are available at the COW website, http://correlatesofwar.org. If you use the data, obtained directly or indirectly, please read the documents below and include the following citations.

> Barbieri, Katherine and Omar Keshk. 2012. Correlates of War Project Trade Data Set Codebook, Version 3.0. Online: http://correlatesofwar.org.
>
> Barbieri, Katherine, Omar M. G. Keshk, and Brian Pollins. 2009. "TRADING DATA: Evaluating our Assumptions and Coding Rules." *Conflict Management and Peace Science*, 26(5): 471–491

This codebook has 6 sections:

1) Dyadic Trade Data
2) National (Monadic) Trade Data
3) Main References
4) Supplementary Information
5) Supplementary References
6) Version/Data History

Our data base builds upon and integrates data from our previous trade data projects. The majority of the post-WWII data were obtained from the International Monetary Fund's *Direction of Trade Statistics*. The most recent update to Version 3.0 used data from the IMF's April 2011 CD-ROM. Version 2.01 relied upon the 2007 CD-ROM Subscription and hard copy versions of multiple years. Version 2.0 also included supplementary data from BKP 1.0 (Barbieri, Keshk, and Pollins, 2005), Barbieri's International Trade Dataset, Version 1.0 (Barbieri 2002), and data from the Republic of China (ROC), Bureau of Foreign Trade. The pre-WWII data were obtained from Barbieri. [1]

The variables in the dyadic trade file are described in the Table 1 below.

Table 1. Variables in the Correlates of War Trade Data Set, Version 3.0 file

Variable	Description
ccode1	Correlates of War Country Code for State A
ccode2	Correlates of War Country Code for State B
year	Observation year
importer1	Name of country A
importer2	Name of country B
flow1	Imports of Country A from Country B in current US millions of dollars
flow2	Imports of Country B from Country A in current US millions of dollars
source1	Source of data for flow1 variable (see table below)
source2	Source of data for flow2 variable (see table below)
China_alt_flow1	Original The People's Republic of China trade values (see notes below)
China_alt_flow2	Original The People's Republic of China trade values (see notes below)
Bel_Lux_alt_flow1	Original Belgium and Luxembourg trade values (see notes below)
Bel_Lux_alt_flow2	Original Belgium and Luxembourg trade values (see notes below)
version	Data version = 3.0

Procedures for obtaining and reporting data

The update to Version 3.0 follows the same procedures as the original Correlates of War Dataset and new data were appended to Version 2.0. The procedures are described in Codebook 2.0 and are repeated here. We first generated all import and export trade data organized by pairs of economic entities, using the International Monetary Fund's *Direction of Trade Statistics on CD-*

[1] Separate supplementary files are provided on the COW Website for those interested in the pre-WWII and other data from Barbieri (2002). The data files contain information about trade figures reported in local currency, exchange rates, and source notes.

ROM. Monthly Subscription and *Direction of Trade Statistics Historical Data* CD-ROM for 1948-1980. The historical data were merged with the post-1980 data (dropping any redundant 1980 values). Details about the IMF's data's trade statistics data are available on line (see IMF, 1993, 2008).

For any given state, the IMF generates an entry of that state's imports from and exports to a list of its partners. The national trade figures include cases where a given state is partners with the world, for either imports or exports. The partner may be a state or non-state actor. The IMF data include, but are not limited to, sovereign states, non-state territories, regional trading groups, geographic regions, level of development grouping (e.g., developing states), and "the world." At this stage, for the dyadic data each observation consists of a given state matched with a given partner. For the national figures each observation consists of a given state matched with the world.

Next, we match the reports of the importer and exporter. For any directional flow of trade between two states (e.g., East to West flows or North to South flows), the importer and exporter should both report a given value. We then have two flows, Flow 1 and Flow 2, each with an importer and exporter report attached to that flow. We begin by matching the importer and exporter reports for Flow 1 and the importer and exporter reports for Flow 2.

The COW dyadic data set reports trade between sovereign states. The matrix of pairs of sovereign states was generated using EUGENE (Bennett and Stam, 2000, 2007). We compared the Correlates of War list of states with the IMF economy list (Correlates of War, 2005). The IMF uses multiple names, spellings, and mixed case combinations for the same country. This appears to be a product of which state is reporting the trade. It makes it extremely difficult to merge the COW state names with the IMF country names. We had to visually inspect many of the cases. In some cases, a country was divided after civil war, reunited, or transformed into a new entity. At times, the IMF retains all the different representations of a given economic/political entity, but typically assigns data to those years in which the state exists.

For example, the IMF includes the USSR in its data base for the period 1948-2006; the Russian Federation from 1980-2006, and Russia from 1980-2006. There are no data for the USSR after 1991 and there are no data for the Russian Federation and "Russia" until after 1991. The problem of multiple names is compounded when one state reports Russia to be its partner and another state lists the Russian Federation as its partner. Typically, Russia uses that term to refer to its own imports and exports to other states, but other states tend to use the term Russian Federation.

We identified similar problems with multiple or changed names for Yugoslavia and Serbia & Montenegro, Korea and South Korea, Germany and West Germany, and Yemen and Yemen Arab Republic. In each case, we isolated those countries and insured that the appropriate data were recorded for the appropriate entity in the appropriate year. We made sure our state list corresponded with the COW state membership list.

We generated a non-directed dyadic data set. Each combination of states has only one entry. In each case, two states report the same flow of goods in one direction. Thus, we have two states reporting Flow 1 and Flow 2. We have four trade values: Flow 1, Importer report (IR); Flow 1, Exporter Report (ER); Flow 2, Importer Report (IR); and Flow 2, Exporter Report (ER). The data are arranged, so the lower number in the COW country list appears in the first column when data for Flow 1 and Flow 2 are merged. For example, the US Canada dyad in 1989 would have one row, with four trade flows, flow1_IR would report the imports of the US from Canada as reported by the US, flow2_IR would report the imports of Canada from the US as reported by Canada. Flow1_ER would

be the exports of Canada to the US as reported by Canada and flow2_ER would be the exports of the US to Canada as reported by the US

The IMF contains many missing trade values. Once we exhausted the data from the importing state's reports, we substituted the missing import values with available figures provided by the exporting state. In other words, when the importer report for Flow 1 was missing, we relied upon the exporter report for Flow 1. Next, we followed the same procedure to replace observations with import values of zero.

After using the current IMF CD-ROMs, we relied upon trade data from earlier IMF tapes (1996) to fill in import values that were missing or coded as zero. Next, we used Barbieri's trade data set to fill in additional missing values. Barbieri relies upon both IMF and non-IMF sources for her trade data. There were a few special cases in the data set.

First, the IMF does not report trade data for Taiwan, so we acquired those data from multiple sources. For the years 1949-188, we obtained data when possible from United Nations *Yearbook of International Trade Statistics* (1951-1984*)*, APEC Study Center City University of Hong Kong Data Bank (2004), and ROC's Council for Economic Planning and Development (2002, 2004). For the years 1989-2009, data were obtained from the ROC's Bureau of Foreign Trade.

Second, the IMF reported one lumped value for Belgium and Luxembourg trade. This changed after 1996. For the pre-1996 values, we include the IMF's original values, plus disaggregated data based on the relative size of the GDP of each country. To disaggregate the Belgium-Luxembourg combined reports, we first obtained the nominal GDP for Belgium and Luxembourg, respectively (World Bank, 2005). We generated a ratio value of the smaller to higher GDP values (i.e., Luxembourg to Belgium). We multiplied Luxembourg's dyadic trade figure by this ratio to produce its trade values. For Belgium, we multiplied the total trade figures by one minus the ratio value above. The original figures are placed in separate columns in our data set (i.e., Bel_Lux_alt_flow1 and Bel_Lux_alt_flow2).[2] While this is not an ideal solution for disaggregating the trade figures, we would recommend using the adjusted figure rather than the same figure for Belgium and Luxembourg's trade.

We found that the trade figures for Macao and Hong Kong continued to be reported separately from Mainland China after these areas were unified in 1998. We combined these figures to produce our measures of China's dyadic trade. Once again, we include the original figures in separate columns in our data set (China_alt_flow1 and China_alt_flow2).

Finally, after following the procedures above, we code any remaining missing data points as missing data. This is one way in which our data set differs from that of some scholars. Some scholars opt to assume missing data indicate zero trade or that the trade conforms to a given linear trend that could be filled in through interpolation or extrapolation techniques. We believe this to be a dangerous and faulty assumption (See Barbieri, Keshk, & Pollins 2007; 2009b; Barbieri & Keshk, 2011).

In Table 2, we provide a summary of the procedures we used to produce each data point. The corresponding numbers appear in our data set. This gives scholars using the data set the opportunity to alter the decision rules and rely upon a different set of figures.

[2] Another strategy to adjust these figures would be to use available disaggregated data for the post 1996 period and generate a ratio for each dyad and apply those to the pre-1996 period.

Table 2. Sources and Corresponding Codes

Source Codes	Source
1	Barbieri Version 1 (see 2002; appendix A)
2	IMF c.i.f. import reports (2007/2011 CD ROMs)[3]
3	Missing import values replaced with IMF exporter's trade (2007/2011 CD ROMs)
4	Zero trade values replaced with the IMF's exporter's trade (2007/2011 CD ROMs)
5	Missing import values replaced with IMF c.i.f. import reports 1992 tapes
6	Zero trade values replaced with IMF c.i.f. import reports 1992 tapes
7	Missing import values replaced with IMF exporter's trade reports 1992 tapes
8	Zero import values replaced with IMF exporter's trade reports 1992 tapes
9	Missing values replaced with Barbieri trade values
10	Zero values replaced with Barbieri trade values
11	Belgium-Luxembourg Data 1948-2006 (see notes below)
12	Taiwan Data from Multiple sources: United Nations (multiple years), APEC Study Center (multiple years), and ROC, Council for Economic Planning and Development (multiple years)
13	Taiwan Data from ROC, Bureau of Foreign Trade
-9	Missing

Table 3 provides a list of the variables in the national trade data file.

Table 3. Variables in the national trade file and their description

Variable	Description
ccode	Correlates of War Country Code
country	Name of country
year	Observation year
imports	Total Imports of Country in current US millions of dollars
exports	Total Exports of Country in current US millions of dollars
source1	Source of data for imports variable (see table 2)
source2	Source of data for exports variable (see table 2)
alt_imports	Original People's Republic of China total import values and Belgium/ Luxembourg to0tal import values (see discussion above)

[3] The 2011 CD ROM was used for the years 2007-2009 and the 2007 CD ROM was used for earlier years.

alt_exports	Original People's Republic of China total export values and Belgium/ Luxembourg to0tal export values (see discussion above)
version	Data version = 3.0

References

APEC Study Center, City University of Hong Kong Data Bank. 2004. http://fbweb.cityu.edu.hk/hkapec/DataBank/Taiwan-Trade-Balance.pdf

Barbieri, Katherine. 2002. *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor, University of Michigan Press.

Barbieri, Katherine and Omar M. G. Keshk. 2011. "Too Many Assumptions, Not Enough Data." Conflict Management and Peace Science 28(2): 168–172.

Barbieri, Katherine and Omar Keshk. 2009a. Correlates of War Project Trade Data Set Codebook, Version 2.1. Online: http://correlatesofwar.org.

Barbieri, Katherine, Omar M. G. Keshk, and Brian Pollins. 2009b. "TRADING DATA: Evaluating our Assumptions and Coding Rules." *Conflict Management and Peace Science*, 26(5): 471–491.

Barbieri, Katherine, Omar Keshk, and Brian M. Pollins. 2003. "The Correlates of War Trade Data Set: A Progress Report." Paper presented at the Annual Meeting of the Peace Science Society (International), Ann Arbor, Michigan, November 13-16.

Bennett, D. Scott, and Allan Stam. 2000. "*EUGene*: A Conceptual Manual." *International Interactions 26:179*-204. Website: http://eugenesoftware.org.

Bennett, D. Scott, and Allan Stam. 1997-2008. EUGene: Expected Utility Generation and Data Management Program. Version 3.204 http://www.eugenesoftware.org/.

Correlates of War Project. 2005. "State System Membership List, v2004.1." Online, http://correlatesofwar.org.

International Monetary Fund. 1993. "A Guide to Direction of Trade Statistics" on CD-Rom and at http://www.imf.org/external/pubs/cat/longres.cfm?sk=154.0.

International Monetary Fund. *Direction of Trade Statistics: Historical Database and Browser* on CD- ROM.

International Monetary Fund. 2007-2011. *Direction of Trade Statistics on CD-ROM.* Monthly Subscription.

International Monetary Fund. 2008. "Data and statistics: Direction of Trade Statistics Consistency between partner country data." http://www.imfstatistics.org/dot/DOTConsi.htm

Republic of China (ROC) Bureau of Foreign Trade Website. 2008. Data source, Directorate General

of Customs, Ministry of Finance. http://cus93.trade.gov.tw/english/FSCE/FSC0011E.ASP

Republic of China (ROC), 2002, 2004. Council for Economic Planning and Development, *Taiwan Statistical Data Book*. http://www.cepd.gov.tw/english/file/2004DataBook.pdf?CFID=87002&CFTOKEN=50209707

STATA 9.2. (2007) Statistics/Data Analysis. College Station, TX: StataCorp

United Nations, Statistical Office, Department of Economic Affairs. *Yearbook of International Trade Statistics*. New York: United Nations, 1951-1984.

World Bank. 2005. World Development Indicators, On-Line Version. World Bank, Washington DC.