**Copyright (c) 2021 Kenneth A. Wilson**

**All rights reserved.**

**USER’S MANUAL**

**CROSS-NATIONAL TIME-SERIES DATA ARCHIVE**

**Introduction**

The Cross-National Time-Series Data Archive (CNTS) was launched by Arthur S. Banks in the fall of 1968 at the State University of New York at Binghamton. The archive was, in part, the outcome of an effort initiated some years earlier to assemble, in machine readable, longitudinal format, certain of the aggregate data resources of *The Statesman's Yearbook*, an annual with a history of continuous publication since 1864, which had never been systematically mined for quantitative materials of potential utility for comparative social scientists. However, many of the data extracted from this source proved to be of questionable reliability (particularly for the earlier years) and a large number of additional sources were ultimately consulted (see Sources and Source Identification, below).

In establishing the archive, it was decided to assemble materials dating, insofar as possible, from 1815 (immediately after the Congress of Vienna and formation of the modern international system). It was also decided that all commonly recognized members of the international community would be represented, excluding a handful of quasi-states such as Andorra, Liechtenstein, Monaco, and Vatican City. In 1977, data for the latter were also introduced, with coverage extending from 1975.

The original file was punched and stored on IBM cards, but these quickly became too numerous for efficient utilization and, in the fall of 1969, were abandoned in favor of tape storage, for which various update, listing, and extraction procedures were concurrently developed.

In January 1971, 102 of the archive's variables were presented in a volume entitled *Cross-Polity Time-Series Data (*M.I.T. Press)*.* For some years thereafter, magnetic tape copies of the file were distributed from Binghamton. Internet access was initiated in December 1997.

Updating the file lagged somewhat in the two decades prior to the compiler's retirement in 1996, but has since been accelerated, with most variables relatively current, save for a few (such as Telegraph Mileage) whose measurement is now of little relevance, or others (such as Urbanization in smaller cities) for which data is no longer available.

The problem of missing data has been addressed as follows. Short-term gaps between "hard data" entries are remedied by means of an inverse compound interest procedure save for some of the early population data for which simple averaging was employed.

Given the wide variety of sources, varying degrees of reliability are to be expected. The file is, however, an open one, and corrections are constantly being made as they become known to the compiler.

The structure of the archive, its content, coding criteria and sources are detailed below.

**STRUCTURE OF THE ARCHIVE**

The archive has almost 200 variables and contains data for over 200 country units, with provision for entries from 1815 (excluding the two modern wartime periods, 1914-1918 and 1940-1945). The basic structure of the archive is that of a rectangular matrix of periodically augmented records, each encompassing data for one country-year.

**STRUCTURE OF THE DATA**

The data is contained in the file, **“20nn Edition CNTSDATA.xlsx”,** and may be categorized in a variety of ways.

First, all of the variables currently included in the file are longitudinal, rather than cross-sectional, in character. The temporal spans of the arrays vary, of course, depending on the availability of data and the relevance of an indicator at a given point in time. To cite the obvious, one would not expect to find telephone data for the first three-quarters of the nineteenth century; less obvious, perhaps, is the general lack of telegraph mileage data after 1939--attributable largely to the decline in relevance of the telegram as a means of communication in the contemporary era. Series terminated for reason of either source availability or relevance have the year of termination shown in the file, **“20nn Edition Codebook.xls”**.

Second, the overwhelming proportion of the data are *interval*-scaled, that is to say, expressed in true numeric units, be they dollars, miles, or what have you. The only *ordinal*-scaled data (ranked on a "more" or "less" basis without the implication of true numeric units) are certain of the political items in Legislative Process Data and Political Data. Only four variables, Type of Regime (**polit01**), Head of State (**polit05**), Premier (**polit06**) and Effective Executive (Type) (**polit07**) are *nominal*-scaled (ranked by qualitative category rather than on a "more"/"less" basis). While a variety of techniques have been developed for relatively sophisticated analysis of noninterval data, most of the readily accessible multivariate procedures remain regression-based, hence technically requiring an interval level of measurement.

Third, the file contains both *primary* and *secondary* (derived) data. The latter are calculated by mathematical manipulation of the primary data, most commonly by conversion of primary variables to per capita or per square mile form in order to achieve inter-nation comparability, and by recasting arrays on the basis of percent annual change.

Finally, most of the archive's interval-scaled arrays contain both *original* and *estimated* data. Each datum is an original entry, either taken directly or derived from an external source. The estimated data, on the other hand, are one of two principal types, depending on whether they were computer-generated (as described above) or supplied by the compiler, usually on the basis of indirect evidence contained in the literature (including instances where initial or terminal original data points fall in the periods 1914-1918 or 1940-1945), to remedy obvious discrepancies in report figures due to typographical or other error, or to "smooth" discontinuities resulting from longitudinal changes in external coding criteria. All such entries are referenced as Compiler’s estimate. Finally, a limited number of less reliable estimates are also included. These "Working estimates" were originally inserted for analytic purposes under circumstances where missing data could not be tolerated, and should be viewed with extreme caution, particularly where they are used as bases for computer-generated estimates.

Urbanization Data, largely in “Population, Cities of 25,000 & Over” (**urban05**) and “Population, Cities of 20,000 & Over” (**urban07**) contains some entries calculated according to a proportional estimation procedure described in Arthur S. Banks and David L. Carr, "Urbanization and Modernization: A Longitudinal Analysis," *Studies* *in* *Comparative* *International* *Development*, 9 (Summer, 1974), 26-45.

**HOW TO SUBSET THE DATA**

To see a subset of the data, in the spreadsheet menu select the down arrow of any column next to its variable name. More than one column may be subset simultaneously by selecting the down arrow of any other column(s).

**Note: The Country Code is invariant, while the country names change over time; thus select subsets based on the country code, not the country name.**

**VARIABLE DEFINITIONS AND CODING CRITERIA**

The variable names, definitions and coding criteria are discussed below, all of which are summarized in **“20nn Edition** **Codebook.xls”.**

*Identification Data*

Three fields are used exclusively for identification purposes: **code (CNTS country code), WBcode (World Bank country code)**, **country and year**. For a list of the Country IDs and Country Labels, see the file, **“20nn Edition** **Independent States Since 1815.xls”.**

Each country has a unique Country ID. Not all of the country labels are, however, invariant through time. Alternative labels are utilized, as follows, for the periods indicated:

|  |  |  |
| --- | --- | --- |
| Labels | Period | Country ID |
| Abyssinia  Ethiopia  Ethiopia PDR  Ethiopia FDR | 1898-1935  1946-1986  1987-1994  1995- | 0370 |
| Austrian Empire  Austria-Hungary | 1815-1866  1867-1913 | 0060 |
| Burma  Myanmar | 1948-1988  1989- | 0140 |
| Cambodia  Khmer Republic  Kampuchea  Cambodia | 1953-1970  1971-1974  1975-1989  1990- | 0160 |
| Central African Republic  Central African Empire Central African Republic | 1960-1975  1976-1978  1979- | 0190 |
| Ceylon  Sri Lanka | 1948-1970  1971- | 0200 |
| China  China Rep  China PR | 1815-1911  1912-1948  1949- | 0230 |
| Congo (Brazzaville)  Congo  Congo Republic | 1960-1970  1971-1996  1997- | 0250 |
| Congo (Kinshasa)  Congo Democratic Rep  Zaire  Congo Democratic Rep | 1960-1963  1964-1970  1971-1996  1997- | 0260 |
| Dahomey  Benin | 1960-1974  1975- | 0310 |
| Egypt  United Arab Republic  Egypt | 1951-1957  1958-1960  1961- | 1200 |
| Federation of Malaya  Malaysia | 1957-1962  1963- | 0750 |
| Ivory Coast  Cote d'Ivoire | 1960-1984  1985- | 0580 |
| Malagasy Republic  Madagascar | 1960-1974  1975- | 0730 |
| Ottoman Empire  Turkey | 1815-1913  1919- | 1170 |
| Persia  Iran | 1815-1913  1919- | 0540 |
| Rhodesia  Zimbabwe | 1965-1979  1980- | 1214 |
| Russia  USSR | 1815-1913  1919-1990 | 1190 |
| Russian Federation | 1991- | 0975 |
| Siam  Thailand | 1815-1913  1919- | 1130 |
| South Yemen  Yemen PDR | 1967-1969  1970-1989 | 1050 |
| Tanganyika  Tanzania | 1961-1963  1964- | 1120 |
| Upper Volta  Burkina Faso | 1960-1983  1984- | 1230 |
| Western Samoa  Samoa | 1968-1996  1997- | 1270 |
| Yemen  Yemen Arab Republic | 1921-1961  1962-1989 | 1280 |
| Yugoslavia  Serbia & Montenegro | 1919-2002  2003-2005, 2010- | 1290 |

*Domestic Conflict Event Data*

While no bibliographic references are utilized in connection with these data, most were initially derived from *The New York Times*. Every effort has been made in recent years to search the Internet for additional reliable news reports. Since 2011, a link to every news report from which an event is counted can be found in the optional LINKS files.

The eight variable definitions (adopted from Rudolph J. Rummel, "Dimensions of Conflict Behavior Within and Between Nations", *General Systems Yearbook*, VIII [1963], 1-50) are as follows:

*Assassinations* (**domestic1**)*.* Any politically motivated murder or attempted murder of a high government official or politician.

*General Strikes* (**domestic2**)*.* Any strike of 1,000 or more industrial or service workers that involves more than one employer and that is aimed at national government policies or authority.

*Terrorism/Guerrilla Warfare* (**domestic3**)*.* Any armed activity, sabotage, or bombings carried on by independent bands of citizens or irregular forces and aimed at the overthrow of the present regime. Such activity may take the form of sporadic attacks on police posts, small villages, government patrols, or military barracks. A country is also considered to have terrorism/guerrilla war when sporadic bombing, sabotage, or terrorism occurs.

*Major Government Crises* (**domestic4**)*.* Any rapidly developing situation that threatens to bring the downfall of the present regime - excluding situations of revolt aimed at such overthrow.

*Purges* (**domestic5**)*.* Any systematic elimination by jailing or execution of political opposition within the ranks of the regime or the opposition.

*Riots* (**domestic6**)*.* Any violent demonstration or clash of more than 100 citizens involving the use of physical force. For the 2021 Edition additional tags were added to the LINKS files to differentiate riots related to:

1. Black Lives Matter (BLM)
2. COVID-19 (COVID)

*Revolutions* (**domestic7**)*.* Any illegal or forced change in the top government elite, any attempt at such a change, or any successful or unsuccessful armed rebellion whose aim is independence from the central government.

*Anti-government Demonstrations* (**domestic8**)*.* Any peaceful public gathering of at least 100 people for the primary purpose of displaying or voicing their opposition to government policies or authority, excluding demonstrations of a distinctly anti-foreign nature. For the 2021 Edition additional tags were added to the LINKS files to differentiate demonstrations related to:

1. Black Lives Matter (BLM)
2. COVID-19 (COVID)
3. Gender Based Violence (GBV)

It should be noted that because these data are based on newspaper and on-line news reports, they are somewhat biased geographically and limited in comprehensiveness. Other distortions are attributable to venues not deemed clearly domestic, such, for example, as the Israel-Palestinian conflict. For these and other reasons, the contents of this segment should be used with extreme caution and, in general, only for macroanalytic purposes.

**Domestic9** is used for weighted conflict measures, the specific weights being variable. As of October 2007, the values entered were: Assassinations (25), Strikes (20), Guerrilla *Terrorism/Guerrilla Warfare* (100), Government Crises (20), Purges (20), Riots (25), Revolutions (150), and Anti-Government Demonstrations (10). Multiply the value for each variable times the specific weights; multiply that sum of products by 100 and divide the result by 8 to confirm the value for domestic9.

Prior to the October 2007 values, the weights were as follows: Assassinations (24), General Strikes (43), *Terrorism/Guerrilla Warfare* (46), Government Crises (48), Purges (86), Riots (102), Revolutions (148), Anti-Government Demonstrations (200).

A Note About Terrorism

Terrorism is defined in Title 22 Chapter 38 U.S. Code § 2656f:

premeditated, politically motivated violence perpetrated against noncombatant targets by subnational groups or clandestine agents.

It is the use of violence or threat of violence in the pursuit of political, religious, ideological or social objectives.

It can be committed by governments, non-state actors, or undercover personnel serving on the behalf of their respective governments.

It reaches more than the immediate target victims and is also directed at targets consisting of a larger spectrum of society.

It is both *mala prohibita* (i.e., crime that is made illegal by legislation) and *mala in se* (i.e., crime that is inherently immoral or wrong).

The following criteria of violence or threat of violence fall outside of the definition of terrorism:

Wartime (including a declared war) or peacetime acts of violence committed by a nation state against another nation state regardless of legality or illegality that are carried out by properly uniformed forces or legal combatants of such nation states. Reasonable acts of self-defense, such as the use of force to kill, apprehend, or punish criminals who pose a threat to the lives of humans or property.

Legitimate targets in war, such as enemy combatants and strategic infrastructure that are an integral part of the enemy's war effort.

Collateral damage, including the infliction of incidental damage to non-combatant targets during an attack on or attempting to attack legitimate targets in war.

*Area and Population Data*

Population Density (**pop2**) is calculated directly from Area in Square Miles (**area2**) and Population (**pop1**), while Population Density of Empire (**pop4**) is calculated directly from Area of Empire in Square Miles (**area3**) and Population of Empire (**pop3**). Area in Square Kilometers (**area1**) or Area in Square Miles (**area2**) is converted from one to the other on the basis of the factors .3861 (from K2 to M2) and 2.590 (from M2 to K2). As in a limited number of other original data fields (identified below), where an unusually large number of individual sources were consulted, no bibliographic references are provided for most of the area data. A substantial portion of the latter for the earlier years were, however, derived from the *Almanach de Gotha*, the *Journal of the Royal Statistical Society* (London), and *The Statesman's Yearbook*. Beginning in 2011 area and population data closely follow the CIA World Factbook figures.

Area and population of empire data are provided for only 13 countries: Austria-Hungary, Belgium, France, Germany, Italy, Japan, Netherlands, Portugal, Russia, Spain, Turkey (Ottoman Empire), United Kingdom, and United States, thus omitting a few marginal cases, such as the dual monarchies of Denmark-Iceland (to 1944) and Sweden-Norway (to 1905). For the Austro-Hungarian, Ottoman, and Russian Empires, the core territories and imperial domains are contiguous; hence the data in fields **area3**, **pop3**, and **pop4** duplicate those in fields **area1**, **area2**, and **pop1**, respectively. The other ten countries are more conventionally identified as "colonial" powers, most of whose possessions are noncontiguous "overseas" territories.

*Urbanization Data*

All fields give aggregate population figures for cities in the following categories: 100,000 and over, 50,000 and over, 25,000 and over, 20,000 and over, and 10,000 and over. Thus, Population, Cities of 50,000 & Over (**urban03**) includes cities of 100,000 and over (**urban01**), and so forth. Per capita data for the same classes of cities are also provided. Most of the externally derived data entries are *compiler summations* from the sources cited.

The inclusion of data for cities of 20,000 and over as well as for cities of 25,000 and over was originally mandated by a lack of uniformity in reporting categories in the sources utilized. Subsequent to preparation of the original version of the file, however, a series of missing data estimates, proportionally calculated *across* urbanization categories, was developed. The procedure for calculating these entries is discussed in Banks and Carr, *op. cit.*

In assembling the urbanization data, considerable difficulty was encountered in regard to the definition of "city" or "urban area". Insofar as possible, data for *core* cities or urban areas are employed, excluding greater metropolitan or suburban populations. It cannot be claimed, however, that the reliability problem is completely surmounted. Indeed, in some cases what UN sources term "municipios" (encompassing rural areas surrounding an urban center) are the only aggregations referenced.

Given the accelerated rate of global urbanization and an increasing dearth of data for smaller-sized localities, most summations for cities fewer than 100,000 have been truncated at 1980. Exceptions are countries with no cities of 100,000 or more; in these cases, lesser categories have been retained.

*National Government Revenue and Expenditure Data*

National Government Revenue and Expenditure (**revexp1**) is calculated directly from National Government Revenue (**revexp3**) and National Government Expenditure (**revexp5**). National Government Revenue and Expenditure Per Capita (**revexp2**) is a dependent (calculated) field based on National Government Revenue and Expenditure (**revexp1**).

National government revenue and expenditure data is reported exclusive of "extraordinary" expenditures financed by direct foreign aid or loans. **revexp4** and **revexp5** contain the same items on a per capita basis. **revexp7** contains the *ratio* of national defense expenditure to total national expenditure. The term "national government" should be construed as referring exclusively to *centraI* government. Thus, monies collected and dispersed locally by national government agencies (as in certain unitary systems) are, wherever possible, excluded.

Revenue and expenditure data, particularly when expressed, as here, in U.S. dollar equivalents, are particularly susceptible to error and should be used with appropriate caution. The possibility of error could, of course, have been substantially reduced had conversion to a common currency unit not been attempted, but the resultant lack of comparability would severely limit the utility of the data in question.

Prior to 1973, official rates of exchange were employed only when deviations therefrom were presumed to be minimal. Otherwise, free (occasionally black) market rates were employed, except in cases of such extreme fluctuation as to preclude the assembly of meaningful series. Needless to say, the overwhelming proportion of data omitted for this reason occurs in the 1919-1939 period.

Since the British pound sterling was the principal basis of international exchange prior to World War I, most data for the period were assembled accordingly, then converted into dollar equivalents at the rate of 4.87 dollars per pound. Some data for 1919-1939 and most data for the post-World War II period were assembled by means of direct conversion to dollar equivalents. It should be noted that here, as elsewhere, there are no "base-year" figures; in other words, there is no adjustment for inflation/deflation in either the British pound (before 1919) or the U.S. dollar (after 1919).

Since 1973 IMF average period market rates have been utilized wherever feasible.

*Trade Data*

All trade data is exclusive of transshipments and bullion transfers. **Trade1** and **trade3** contain import and export data respectively, while **trade2** and **trade4** contain the same items on a per capita basis. Both imports and exports are f.o.b.

**Trade5** is a periodic update of the proportion of world trade (imports and exports) for each country for each year. Since the denominator employed is simply a summation of imports and exports for all *independent* nations included in the archive, it falls somewhat short of being a total summation of world trade. It may be assumed, however, that the proportion contributed by nonindependent territories for most years is relatively small. As in the case of revenue and expenditure data, conversion to U.S. dollar equivalents involves a certain degree of risk as regards the introduction of error, but without such conversion the data would be largely worthless for comparative purposes.

*Energy Data*

Energy production and consumption are provided in these variables as measured in metric tons of coal equivalent (converted to oil equivalent through 1993) and from 1994 measured in metric tons of oil equivalent. The shift from coal to oil equivalents was necessary because of a shift by the UN Statistical Office, whose figures are utilized; standardization is achieved by using the conversion factors of .700 for coal to oil and by 1.43 from oil to coal. (See listings in **“Codebook.xls”***).* United Nations definitions are employed throughout.

Production data are based on the production of coal, lignite, crude petroleum, natural gas and hydro and nuclear energy; where peat used as fuel is important, it is included with coal and lignite. Consumption data are based on the apparent consumption of coal, lignite, petroleum products, natural gas and hydro and nuclear energy. Coke, manufactured gas and electricity internally traded are considered to have been consumed by the importing country. Bunkers supplied to foreign-going ships are excluded from consumption. (UN Statistical Yearbook: 1971, p. 44).

**Energy1** contains data on overall energy production in metric tons of oil.

**Energy2** contains data on overall energy production in kilograms per capita.

**Energy3** contains data on overall energy consumption in metric tons of oil.

**Energy4** contains data on overall energy consumption in kilograms per capita.

*Military Data*

National Defense Expenditure (**military1**) is calculated from National Government Expenditure (**revexp5**) and the ratio National Defense Expenditure/National Government Expenditure (**revexp**7). While deriving the data in this way unquestionably results in some loss of precision, it was not considered sufficiently consequential to offset the added labor required to assemble collateral data directly from external sources.

**Military2** contains **military1** data in per capita form.

**Military3** is the size of military, while **military4** contains the same information on a per capita basis. The "military" is defined as embracing all active-duty members of a nation's armed forces (army, navy, air corps) and excludes all semi- or paramilitary forces, save in a limited number of cases (such as Japan and Panama) where, for some or all reporting years, military establishments are not formally acknowledged. In the case of Switzerland, which does not maintain a continuously active military establishment, estimates of active-duty reserves are utilized.

*Industrial and Labor Force*

**Industry1** is the Percent GDP Originating in Industrial Activity, while **industry2** is the same information on a per capita basis. "Industrial activity" is defined as embracing categories 2-4 of the revised (1958) International Standard Industrial Classification of all Economic Activities (ISIC), which includes mining and quarrying; manufacturing; and electricity, gas and water.

**Industry3**, **industry4** and **industry5** contain percent workforce engaged in agriculture, industry, and other activity, respectively. "Industry" is here defined as embracing revised ISIC categories 2-3 and 5, which include mining and quarrying; manufacturing; and construction, while "agriculture" is defined in terms of revised ISIC category 1, which includes agriculture, forestry, and fishing. "Other activity" is simply the sum of the foregoing subtracted from 100%.

It should be noted that some sources report on "civilian labor force employed", while others report on "number of employees" (based on statistics of establishments). The latter normally encompass only a limited portion of the labor force and, for that reason, have not been utilized.

*Railroad Data*

**Railroad1** embraces railroad mileage, defined as miles of *line* (both public and private), rather than as miles of *track*. Thus, ten miles of a single track line would be counted as equal to ten miles of double track line. Tramway (e.g., streetcar) and lift lines are excluded, but not cog railways if of a non-tramline character. **Railroad2** contains the same data on a per square mile basis.

**Railroad3** and **railroad4** deal with rail passenger-miles and rail passenger-kilometers, respectively, the first being a calculated variable derived from the second. These data refer, of course, to the sum of miles or kilometers traveled by each individual rail passenger. Similarly, **railroad5** and **railroad6** are based on rail-ton miles and rail-ton kilometers, respectively, of freight carried. **Railroad7** records rail-ton miles per capita.

Given the recent decline in importance of rail transportation, all of the series in this segment are terminated as of 1981.

*Highway Vehicle Data*

**Vehicle1** and **vehicle3** are based on the total number of passenger and commercial vehicles, respectively, while **vehicle2** and **vehicle4** contain the same two items in per capita form. **Vehicle5** (all highway vehicles) is the sum of **vehicle1** and **vehicle3**, while **vehicle6** is based on all highway vehicles per capita. Motorcycles and motorized construction equipment are excluded from these categories. Taxis (though technically "commercial vehicles") are counted as passenger cars. Buses, vans, lorries, etc., are all classified as commercial vehicles, even though some may be privately owned and not used for commercial purposes.

*Phone Data*

**Phone1** is a summation of **phone2** and **phone3** data, thus referencing all telephones, including cellular. The number of telephones and telephones per capita are located in **phone3** and **phone4**, and, for many years, exhibit a high degree of reliability because of their ultimate source: the reasonably accurate local telephone directory. It should be noted, however, that there is some likelihood of underreporting in the early years of telephonic communication, when a disproportionate number of instruments were owned or operated by private businesses and government offices. An equally serious source of underreporting stems from contemporary reports based on number of *lines*, which may service a number of instruments.

**Phone3** and **phone4** exclude mobile cellular telephones. The number of such instruments (since 1989) is given in **phone2**, while (as noted above) the total number of telephones, both cellular and noncellular, is given in **phone1**.

**Phone5** and **phone6** contain dependent (calculated) telephone entries: mobile cellular telephones per capita and all telephones, including cellular, per capita.

*Telegraph Data*

**Telegraph1** deals with telegraph mileage, defined as miles of *line* (both public and private), rather than as miles of *wire*. Telegraph mileage per square mile is given in **telegraph2**. **Telegraph3**and **telegraph4** contain the number of telegrams and telegrams per capita, respectively. For these entries, every effort has been made to report purely *domestic* telegraphic activity, excluding foreign sent and received, as well as in-transit messages. However, in some case (particularly in the pre-World War I period) the sources do not adequately distinguish between the several message categories, and occasional over-reporting may be expected. The result is a serious reliability problem for certain Latin American countries during the latter years of the nineteenth century, when an unusually high proportion of telegrams fall into the foreign-sent and foreign-received categories.

Since virtually no data for telegraphic mileage could be located for the post-World War II, entries for **telegraph1** and **telegraph2** are discontinued after 1939, while entries for **telegraph3** and **telegraph4** terminate in 1980.

*Mail Data*

**Mail1** contains first class mail and **mail2** first class mail per capita. **Mail3** is all letter-post mail, while **mail4** contains all letter-post mail per capita.

As in the case of telegraphic communication, the coding criteria call for the exclusion of foreign sent/received and in-transit items, although in some case where official government figures are used, at least some foreign items appear to be included.

Newspapers carried by mail are included as bona fide (non-first class) postal matter, but since figures for the latter are occasionally lacking, some discrepancies are to be expected in the all-mail category. Post cards are, of course, construed as first class items and prior to World War I constituted a large part of the latter class of mail in many European countries (most notably Germany).

Most of the post-World War II mail figures are from the Universal Postal Union (UPU), which does not distinguish between first-class and all mail. For this reason, the first-class series are, for the most part, terminated as of 1939.

*Media Data: Radio and Television Set Data; Newspaper Data; Book Production Data*

**Media1** and **media3** contain data on radio and television sets, respectively, while **media2** and **media4** deal with the same items in a per capita basis. **Media5** is devoted to newspaper circulation per capita, **media6** concerns book production by number of titles published, and **media7** deals with the latter on a per capita basis. All media data are for comparatively recent years (the earliest, number of radio receivers, goes back only to 1938, while the most recent, television receivers, dates from 1960).

There is a tendency for news circulation to be underreported, since data for weekly and biweekly publications are not included. It should also be noted that book production figures generally include children's and school text books, and are not restricted to either first edition or hardbound titles. It should be emphasized, however, that the data reference only the number of *titles*, not copies in print.

*School Enrollment Data*

**School01** and **school03** contain data on primary and secondary enrollment, respectively, while **school02** and **school04** deal with the same items on a per capita basis. **School05** aggregates **school01** and **school03**, yielding primary *and* secondary enrollment, while **school06** presents the same data in per capita form. **School07** offers primary enrollment as a *proportion* of primary and secondary enrollment.

Although significant improvement has been registered over the years regarding standardization of reporting categories in educational statistics, many difficulties remain in attempting to assemble truly comparable data, particularly of a longitudinal character. Insofar as possible, data on preprimary, vocational or technical, part-time, and adult education students have been omitted from the archive listings. With these exceptions, every effort has been made to assemble data on the basis of relevant UNESCO criteria:

*First level:* Education whose main function is to provide basic instruction in the

tools of learning (e.g., at elementary school, primary school). Its length may vary

from 4 to 9 years, depending on the organization of the school system in each

country;

*Second level:* Education based upon at least four years of previous instruction at

the first level, and providing general or specialized instruction, or both (e.g., at

middle school, secondary school, high school . . .);

*Third level:* Education which requires, as a minimum condition of admission, the

successful completion of education at the second level, or evidence of the

attainment of an equivalent level of knowledge. . . (UN *Statistical* *Yearbook:*

*1973*, p. 781).

Regrettably, the UN criteria for categorizing second-level instruction changed during 1964-65. In general, 1964 "secondary level" figures are equated with 1965 and later "second level: general" figures, but not uniformly so. Also the omission of vocational education introduces an element of bias, particularly in socialist countries, because of the inclusion of many students under this rubric.

**School08** and **school10** deal with university and total school enrollment, respectively, while **school09** and **school04** report the same items on a per capita basis.

**School12** contains literacy data, calculated, wherever possible, on the basis of nonliterates, 15 years of age and over. Literacy is defined in the UN *Demographic Yearbook* (from which most of the post-World War II data are extracted) as "ability both to read and to write". While this is not an entirely adequate definition, it is unrealistic to assume that the caliber of most reporting agencies could sustain a more precise one. Indeed, for the limited amount of pre-World War I literacy data that is included in the file, overall reliability must be assessed with extreme caution.

*Physician Data*

**Physician1** deals with inhabitants per physician, while its reciprocal (physicians per capita) appears in **physician2**. The latter is deemed a somewhat more useful cross-national indicator than the former (which appears in the UN *Statistical Yearbook*), since the direction of the array, for most countries, accords with that of other "developmental" indicators (tending to yield positive rather than negative correlation coefficients).

*Economic Data*

**Economics1** is devoted to national income per capita, **economics2** to gross domestic product (at factor cost) per capita, and **economics3** to gross national product (at market prices) per capita. These three basic components of aggregate product are defined as follows:

*Gross national product at market prices* is the market value of the product, before

deduction of provisions for the consumption of fixed capital, attributable to the

factors of production supplied by normal residents of the given country. It is

identically equal to the sum of consumption expenditure and gross domestic capital

formation, private and public, and the net exports of goods and services plus the net

factor incomes received from abroad.

*Gross domestic product at factor cost* is the value at factor cost of the product,

before deduction of provisions for the consumption of fixed capital, attributable to

factor services rendered to resident producers of the given country. It differs from

the gross domestic product at market prices by the exclusion of the excess of

indirect taxes over subsidies.

*National income* is the sum of the incomes accruing to factors of production

supplied by normal residents of the given country before deduction of direct taxes.

(UN *Yearbook of National Accounts Statistics, 1969,* v. 1, p. xi.)

The interrelationships of the three aggregates are as follows: GNP at market prices *less* net factor income from abroad and indirect taxes net of subsidies *equals* GDP at factor cost. The latter, in turn, *less* depreciation, *plus* net factor income from abroad, *equals* national income (*ibid*, p. 819). *All data for these three indices for the period 1970-1973 are estimated* because of definitional changes in 1970, which make aggregate product figures somewhat inconsistent with earlier figures. It should be noted that as of 1999 the World Bank reports GNP as gross national income (GNI).

Largely because of the abandonment by the UN of national income figures in US dollars, the series terminates as of 1985, although data in national currency continues to be reported by the IMF. As of 1998, the UN Statistical Yearbook figures are used.

**Economics4** deals with per capita currency in circulation, expressed in U.S. dollars at the free market rate, save in a limited number of cases where the free rate closely approximates the official rate. Data are from *Pick's Currency Yearbook*, whose reports terminated as of 1984.

**Economics5** gives the age of a nation's currency in months. "Age" is defined in terms of the number of months that have elapsed since the introduction of a new monetary system or since an upward or downward revaluation of 5% or more. In cases of *multiple* revaluations totaling 5% or more during a given year, the count is from the last such revaluation. Because of the general abandonment of artificially pegged and multiple rate systems, the series is discontinued after 1970.

**Economics6** gives a nation's official exchange rate at year's end, expressed in local currency per U.S. dollar. After 1971 the effective rate (usually the IMF market or principal rate) is used if the official rate is inoperative.

**Economics7** gives the free or black market rate in local currency per U.S. dollar, primarily as reported until 1985 by *Pick's Currency Yearbook.*

*Electoral Data*

The percent turnout in the most recent (lower house) legislative election is given in **electoral1**; **electoral2** gives the number of registered voters (in some cases, such as the United States, those eligible to register and vote) for the year in question; **electoral4** contains the number of *valid* votes cast. (For the overall turnout, including those whose ballots were disallowed, **electoral2** should be multiplied by **electoral1**). In situations involving runoff balloting, the figures are based on first-round results.

*Legislative Process Data*

**Legis01** contains the number of seats held by the largest party in the lower house of each country's national assembly. In one-party systems with legislative membership in excess of 999, then 999 is entered.

**Legis02** contains the total number of seats in the lower house, *except in cases where no parties exist (or did not exist at the last election),* where a zero is entered (in such cases, the *absence of a legislature* is indicated by zero entries in **legis03** and **legis04)**. In one-party systems with legislative membership in excess of 999, then 999 is entered from legis02.

**Legis03**-**legis06** contain ordinal-scaled data, coded as follows:

**Legis03***. Effectiveness of Legislature*

(3) Effective

(2) Partly Effective

(1) Largely Ineffective

(0) No Legislature

**Legis04***. Nominating Process*

(3) Competitive

(2) Partly Competitive

(1) Essentially Non-Competitive

(0) No Legislature

**Legis05***. Legislative Coalitions*

(3) More than one party, no coalitions

(2) More than one party, government coalition, opposition

(1) More than one party, government coalition, no opposition

(0) One party or no parties

**Legis06***. Party Legitimacy*

(3) No parties excluded

(2) One or more minor or "extremist" parties excluded

(1) Significant exclusion of parties (or groups)

(0) No parties, or all but dominant parties and satellites excluded

It may be noted that the data in **legis03** are substantively similar to the data in **polit13**, below. The two data sets are not, however, *identical*. For the earlier years they were coded at different times and incorporated into the file as components of different subfiles. In recent years, they tend to converge.

**Legis07** is an index of seats held by the largest party, obtained by dividing **legis02** by **legis01**. The principal reason for calculating the index in this manner (rather than as a percentage of seats held) is to ensure that the entries for countries with no parties (or no legislatures) and countries with one-party systems will be adjacent, rather than at opposite extremes of the array. Thus, a country with no parties has a score of 0, a one-party system has a score of 1.0, a system with 40 out of 100 seats held by a majority party has a score of 2.5, etc.

**Legis08**-**legis10** contain secondary data derived from items appearing above. **Legis08** is a total of the ordinal scores contained in **legis03**-**legis06** and, as such, may be construed as a simple, nonfactoral, measure of political polyarchy or pluralism. **Legis09** contains seven-year averages of the data in **legis07**, while **legis10** contains seven-year totals of the data in **legis08**.

*Political Data*

**Polit01** is a party fractionalization index, based on a formula proposed by Douglas Rae in "A Note on the Fractionalization of Some European Party Systems", *Comparative Political Studies*, 1 (October 1968), 413-418. The index is constructed as follows:

m

F = 1 - Σ (ti)2

i=i

where ti = the proportion of members associated with the

ith party in the lower house of the legislature

(where there are no parties, a zero is entered)

In calculating the Index entries, independents are disregarded and legislative changes between elections are not taken into account. It should also be noted that sources vary on the distribution of seats (and even the overall number of seats) for many countries; thus figures calculated by different researchers may vary.

**Polit02**-**polit15** embrace 14 nominal and ordinal political variables coded as follows:

**Polit02***. Type of Regime*

*(1) Civilian.* Any government controlled by a nonmilitary component of the

nation's population.

*(2) Military-Civilian.* Outwardly civilian government controlled by a military

elite. Civilians hold only those posts (up to and including that of Chief of

State) for which their services are deemed necessary for successful conduct of

government operations. An example would be retention of the Emperor and

selected civilian cabinet members during the period of Japanese military

hegemony between 1932 and 1945.

*(3) Military.* Direct rule by the military, usually (but not necessarily)

following a military coup d'état. The governing structure may vary from

utilization of the military chain of command under conditions of martial law

to the institution of an ad hoc administrative hierarchy with at least an upper

echelon staffed by military personnel.

*(4) Other.*  All regimes not falling into one or another of the foregoing

categories, including instances in which a country, save for reasons of

exogenous influence, lacks an effective national government. An example of

the latter would be Switzerland between 1815 and 1848.

**Polit03***. Coups d'État*

The number of extraconstitutional or forced changes in the top government

elite and/or its effective control of the nation's power structure in a given

year. The term "coup" includes, but is not exhausted by, the term "successful

revolution". Unsuccessful coups are not counted.

**Polil04***. Major Constitutional Changes*

The number of basic alterations in a state's constitutional structure, the

extreme case being the adoption of a new constitution that significantly alters

the prerogatives of the various branches of government. Examples of the

latter might be the substitution of presidential for parliamentary government

or the replacement of monarchical by republican rule. Constitutional

amendments which do not have significant impact on the political system are

not counted.

**Polit05***. Head of State*

*(1) Monarch.*  Chief of state is a monarch (either hereditary or elective) or a

regent functioning on a monarch's behalf.

*(2) President.* Chief of state is a president who may function as a chief

executive or merely as titular head of state, in which case he will possess

little effective power. The presiding officer of a legislative assembly or state

council may qualify for the coding, even though the formal title may be that

of "chairman".

*(3) Military.*  A situation in which a member of the nation's armed forces is

recognized as the formal head of government. In case of conflict between

(2) and (3), coding is determined on the basis of whether the incumbent's

role is intrinsically military or civilian in character.

*(4) Other.*  This category is generally used when no distinct head of state

can be identified; it also includes individuals not included in (1-3), such as

theocratic rulers, as well as nonmilitary individuals serving in a collegial

capacity.

**Polit06***. Premier.*

(1) Formal executive is premierial, including "Chairman, Council of

Ministers"

(2) Formal executive is non-premierial

**Polit07***. Effective Executive (Type)*

Refers to the individual who exercises primary influence in the shaping of

most major decisions affecting the nation's internal and external affairs. The

"other" category may refer to a situation in which the individual in question

(such as the party first secretary in a Communist regime) holds no formal

governmental post, or to one in which no truly effective national executive

can be said to exist.

(1) Monarch

(2) President

(3) Premier

(4) Military

(5) Other

**Polit08***. Effective Executive (Selection)*

*(1) Direct Election.* Election of the effective executive by popular vote or

the election of committed delegates for the purpose of executive selection.

*(2) Indirect Election.* Selection by an elected assembly or by an elected but

uncommitted electoral college. This coding is also used when a legislature is

called upon to make the selection in a plurality situation.

*(3) Nonelective.* Any means of selection not involving a direct or indirect

mandate from an electorate.

**Polit09***. Parliamentary Responsibility*

Refers to the degree to which a premier must depend on the support of a

majority in the lower house of a legislature to remain in office.

*(0) Irrelevant.*  Office of premier or legislature does not exist.

*(1) Absent.* Office of premier exists, but there is no parliamentary

responsibility.

*(2) Incomplete.* The premier is, at least to some extent, constitutionally

responsible to the legislature. Effective responsibility is, however, limited.

*(3) Complete.* The premier is constitutionally and effectively dependent

on a legislative majority for continuance in office.

**Polit10***. Size of Cabinet (end of year)*

Refers to the number of ministers of "cabinet rank", excluding

undersecretaries, parliamentary secretaries, ministerial alternates, etc.

*Includes the president and vice-president under a presidential*

*system, but not under a parliamentary system*.

In many cases, counts are approximate, since sources often differ

(particularly in regard to "ministers of state") as to what constitutes

cabinet status.

Generally, the count is of ministries, not of individuals holding multiple

offices (the most extreme recent case being that of New Zealand).

**Polit11***. Major Cabinet Changes*

The number of time in a year that a new premier is named and/or 50%

of the cabinet posts are assumed by new ministers.

**Polit12***. Changes in Effective Executive*

The number of times in a year that effective control of executive power

changes hands. Such a change requires that the new executive be

independent of his predecessor.

**Polit13***. Legislative Effectiveness*

*(0) None.* No legislature exists.

*(1) Ineffective.*  There are three possible bases for this coding: first,

legislative activity may be essentially of a "rubber stamp" character;

second, domestic turmoil may make implementation of legislation

impossible; third, the effective executive may prevent the legislature

from meeting, or otherwise substantially impede the exercise of its

functions.

*(2) Partially Effective.* A situation in which the effective executive's

power substantially outweighs, but does not completely dominate, that

of the legislature.

*(3) Effective.* The possession of significant governmental autonomy by

the legislature, typically including substantial authority in regard to

taxation and disbursement, and the power to override executive vetoes

of legislation.

**Polit14***. Legislative Selection*

*(0) None.*  No legislature exists.

*(1) Nonelective.*  Examples would be the selection of a majority of

legislators by the effective executive, or by means of heredity or

ascription.

*(2) Elective.* A majority of legislators (or members of the lower house in

a bicameral system) are selected by means of either direct or indirect

popular election.

**Polit15***. Legislative Election*

The number of elections held for the lower house of a national legislature

in a given year. A limited number of by-elections are included, but most are

not.

*International Status Indicators*

**Instat1**-**instat8** embrace, *for the period 1817-1935,* eight international status indicators developed by J. David Singer and Melvin Small in "The Composition and Status Ordering of the International System: 1815-1940," *World Politics,* 18 (January 1966), 236-282. Singer and Small provide entries, in each case, for every fifth year. *Yearly* estimates were calculated and are provided in the present file for the basic variable, "International Status, Composite Score", which appears in **instat3**. The full set of entries is as follows:

**Instat1***. International Status: Ranking*

**Instat2**. *International Status: Case Size*

**Instat3***. International Status: Composite Score*

**Instat4***. International Status: Composite Standardized Score*

**Instat5***. International Status: Quintile*

**Instat6***. International Status: Weighted Rank*

**Instat7***. International Status: Weighted Status Ordering*

**Instat8***. International Status: Weighted Quintile*

For a discussion of these data and the coding criteria employed, see Singer and Small, *op. cit.*

Given structural changes in the international system, the Singer-Small coding criteria became increasingly irrelevant in the late 1930's and no attempt was made to continue the series beyond 1935.

*Computer Indices*

B*eginning in 1999* a set of computer indices, as follows are reported:

**Computer1***. Internet Hosts*

**Computer2***. Internet Hosts Per Capita*

**Computer3***. Internet Users*

**Computer4***. Internet Users Per Capita*

**Computer5***. Estimated Personal Computers*

**Computer6***. Estimated Personal Computers Per Capita*

*Industrial Production*

**Indprod1**gives electric power production. Insofar as possible, the data include production for both public and private purposes, and cover both thermal and hydroelectric output, thus reflecting total gross generation of electricity, excluding station use and transmission losses. **Indprod2** gives the same information in per capita form.

As in the case of data in *Energy Production* (see above), conversion factors are linked to time spans adopted by the UN Statistical Office. Here there are three such spans: kilowatt hours for 1919-1980, metric tons of coal equivalent for 1981-1993, and metric tons of oil equivalent from 1994. The conversion factors used are .123 from 1000000 kwh to 1000 coal; .086 from 1000000 kwh to 1000 oil; 8.13 from 1000 coal to 1000000 kwh; .700 from 1000 coal to 1000 oil; 1.43 from 1000 oil to 1000 coal; and 11.6 from 1000 oil to 1000000 kwh (see listings in *“***Codebook.xls***”).*

**Indprod3** contains data on crude steel production, including, insofar as possible, both ingots and steel for castings, whether obtained from pig iron or scrap. Wrought (puddled) iron is generally excluded. **Indprod4** gives the same data in per capita form.

**Indprod5** contains data on the total production of hydraulic cements used for construction purposes (Portland, metallurgic, aluminous, natural, etc.). **Indprod6** gives the same data in per capita form.

*Percent Annual Increase Data*

All of the fields in this segment are defined as **Delta** variables and each contains derived data of a percent annual increase (decrease) character. Values are calculated only on the basis of entries >0 for consecutive years. A value entered at year Y is (B-A)/A, where A is the original entry for year Y-1 and B is the original entry for year Y. For the items included in this subset, see **“2016 Edition** **Codebook.xls”.** Delta values compare increase (decrease) for a particular country for a particular year. If there is an increase it is a positive value, if it decreases it is a negative value compared to the previous year.

**SOURCES AND SOURCE IDENTIFICATION**

*Cited Sources*

In the case of serials every effort has been made to utilize the most recent editions, in order to benefit from revisions of earlier information. By way of example, the following entry appears in the file, **“20nn Edition** **Bibliography.xls”**, which is sorted by code, variable and ranges:

|  |  |  |  |
| --- | --- | --- | --- |
| code | variable | ranges | Source |
| 0061 | electoral2 | 1919-1971 | Stein Rokkan & Jean Meyriat, Eds., International Guide to Electoral Statistics. The Hague: Mouton, 1969, p. 45. |

This entry tells us that the nonofficial source for country 0061 (Austria), variable **electoral2**, from 1919 through 1971 are taken from page 45 of the volume edited by Rokkan and Meyriat.

*Uncited Sources*

Prior to 2011 all of the conflict data citations were virtually impossible because of the large number of sources consulted. However, since 2011 the LINKS (hypertext links) are available to support each entry in the conflict data.

Databanks International has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites or articles referred to in the LINKS files and does not guarantee that any content on such websites is or will remain accurate or appropriate. However, you may contact us in the case of a link which no longer works, and we may be able to help you.

email:   [support@databanksinternational.com](mailto:support@databanksinternational.com)  
mobile phone: +972-52-521-1523  
phone: +972-2-566-9340  
contact person: Kenneth A. Wilson  
[https://www.cntsdata.com](https://www.cntsdata.com/)