

# Appendix 1C. Measuring violence: an introduction to conflict data sets

TAYLOR B. SEYBOLT

## I. Introduction

The systematic study of violent conflict was pioneered by Quincy Wright in *A Study of War*, first published in 1942.<sup>1</sup> The book brought quantitative analysis of conflicts to the attention of a wide audience for the first time by drawing on numerical and statistical research, in addition to historical and legal material.<sup>2</sup> Following Wright, the cornerstone of modern, systematic, quantitative studies of conflict was set in place in the mid-1960s by J. David Singer and Melvin Small with their Correlates of War project.<sup>3</sup> Since the 1980s, with the advent of the widespread use of computers, a multitude of conflict data-collection projects have emerged.<sup>4</sup> As the number of systematic data collections has increased, the field has become increasingly diverse and complex.<sup>5</sup> As a result, there is disagreement on some of the most basic questions. Is the world more or less violent today than in the past? Are wars more or less destructive than they used to be? Are modern violent conflicts different from earlier ones? What are the causes of conflict initiation, continuation and termination?

Most of the differences in opinion about these and other questions can be traced to differences in the collection and use of data. Variations in purpose, definitions and coding rules can lead to significant divergence on basic parameters, such as the number, frequency, duration and dispersal of armed conflicts in the world.<sup>6</sup> The purpose of this appendix is to introduce the users of conflict data to important methodological, theoretical and policy-related questions that face researchers in the systematic study of conflict. The world's primary English-language data-collection projects are presented, as are reasons for the differences between them.<sup>7</sup>

<sup>1</sup> Wright, Q., *A Study of War*, 2nd edn (University of Chicago Press: Chicago, Ill., 1965).

<sup>2</sup> Richardson, L. F., 'Generalized foreign politics', *British Journal of Psychology: Monograph Supplements*, vol. 23 (1934); and Sorokin, P., *Social and Cultural Dynamics*, vol. 3, *Fluctuation of Social Relationships, War and Revolution* (American Book Company: New York, 1937).

<sup>3</sup> Singer, J. D. (ed.), *The Correlates of War I: Research Origins and Rationale* (Free Press: New York, 1979). SIPRI published a list of 11 studies of international wars and armed conflicts in its first Yearbook. The list gave the definition of armed conflict used and publication information for each study. The studies were used to generate a table of conflicts, noting their type, size, parties, and beginning and end dates for the period from 1820 to 1968. Naidu, H., 'Conflicts', *SIPRI Yearbook of World Armaments and Disarmament 1968/69* (Almqvist & Wiksell: Stockholm, 1969), pp. 359–73.

<sup>4</sup> Two excellent central sources of databases on many aspects of social, economic and political life are the Inter-University Consortium for Social and Political Research, available at URL <<http://www.icpsr.umich.edu/>>, and the Harvard–MIT Data Center, available at URL <<http://data.fas.harvard.edu/>>. Access to a wide range of data collections on several aspects of international affairs is provided by Paul Hensel at URL <<http://garnet.acns.fsu.edu/~phensel/intldata.html>> and Richard Tucker at URL <<http://www.vanderbilt.edu/~rtucker/methods/data/>>.

<sup>5</sup> Identifying Wars: Systematic Conflict Research and its Utility in Conflict Resolution and Prevention, Conference held at Uppsala University, Uppsala, Sweden, 8–9 June 2001. The conference agenda, list of participants and papers are available at URL <<http://www.pcr.uu.se/>>.

<sup>6</sup> Gleditsch, N. P. *et al.*, 'Armed conflict 1946–99: a new dataset', Paper presented at the Identifying Wars Conference (note 5).

<sup>7</sup> Restricting this survey to English-language projects may under-represent the variety of work being done, but it captures almost all of the work in the sub-field. The vast majority of conflict data-collection

Section II reviews variations in purpose, definitions, methods and coding rules as sources of difference between data sets. Section III summarizes the primary data-gathering projects whose data are publicly available and provides contact information for them. Internet access to all the data projects reviewed here is available from the SIPRI Internet site at URL <<http://projects.sipri.se/conflictstudy/index.html>>. The last section summarizes the conclusions of this examination of conflict data.

## II. Variations in purpose, definitions and methodology

### Purposes

One of the primary explanations for the differences between data sets is that they are intended to shed light on different aspects of conflict. Data-collection projects are categorized in section III according to whether they are primarily concerned with the global pattern of conflict, processes of conflict initiation and termination, or conflict prediction and cost.<sup>8</sup>

The type of data-collection project that attempts to facilitate the understanding of the patterns of conflict occurrence, rather than the processes of conflict development, is the most well established. These projects measure the frequency, location and severity of conflicts. Most of them determine when a conflict should be included in the data set by the estimated number of people killed.<sup>9</sup> A few projects in this category use qualitative, rather than quantitative, determinants of what to consider as a conflict.<sup>10</sup> Data projects in this category cover both interstate and intra-state conflicts and distinguish between them. Data sets that focus on counting incidents are useful for investigating the prevalence and locations of various types of conflict. Some projects include data on the characteristics of the antagonists in violent conflicts, such as their geographical location, relative capabilities and ethnic identity.<sup>11</sup> This information allows for testing hypotheses on the causes of conflict. Data sets of this type also estimate the costs of conflicts in terms of the number of deaths. They can potentially assist cost-of-conflict studies by providing data on population displacement, economic performance, military expenditure and other factors, if they are combined with non-conflict data sets.<sup>12</sup> The geographical data offered are usually too general to allow

projects are conducted in the United States, probably because of the behavioural approach to studying social phenomena that is prominent in the USA but not elsewhere. In addition, some projects that reside in other countries publish some or all of their material in English. The only notable data project that provides data sets in another language is Arbeitsgemeinschaft Kriegsursachenforschung (AKUF). The AKUF Internet site provides data sets in German; some English-language papers are available at URL <<http://www.sozialwiss.uni-hamburg.de/Ipw/Akuf/home.html>>.

<sup>8</sup> These categories overlap considerably in practice. The projects are not monolithic and can be used to study several different aspects of conflict.

<sup>9</sup> These are the Correlates of War project, the Conflict Data Project and the Major Episodes of Political Violence project. Full information is provided in section III for all the data-collection projects named in the footnotes, unless otherwise noted.

<sup>10</sup> The Conflict Simulation Model (Konflikt-Simulations-Modell, KOSIMO) and AKUF projects.

<sup>11</sup> E.g., the Correlates of War project.

<sup>12</sup> Information on population displacement is available from the UN High Commissioner for Refugees at URL <<http://www.unhcr.ch>>, the US Committee for Refugees at URL <<http://www.refugees.org>> and the Norwegian Refugee Council at URL <<http://www.nrc.no/engindex.htm>>. Economic data are available from *inter alia* the World Bank at URL <<http://www.worldbank.org>> and the UN Development Programme (UNDP) at URL <<http://www.undp.org>>. Military expenditure data are available from the SIPRI Military Expenditure Project, URL <[http://projects.sipri.se/milex/mex\\_data\\_index.html](http://projects.sipri.se/milex/mex_data_index.html)>. Data on the dollar value of transfers of major conventional weapons are available from the SIPRI Arms Transfers Project at URL <<http://projects.sipri.se/armstrade/atfproj/html>>. Data on legal transfers

greater specificity than an entire country, which leads to a distorted picture of zones of conflict and zones of peace.

A second type of data-collection effort is concerned with studying conflict dynamics by examining the factors that are thought to have an impact on the outbreak, continuation and termination of conflicts. Data projects in this category tend to include either intra-state or interstate conflicts, but not both. Projects that are primarily concerned with the causes of conflict look at a range of factors, from historically common catalysts, to characteristics of groups vulnerable to crisis behaviour.<sup>13</sup> Some data sets that are concerned with the dynamics of ongoing conflicts focus on a particular kind of event, such as intervention by outside states.<sup>14</sup> Other projects attempt to capture the complexity of societies in conflict.<sup>15</sup> A few data projects are primarily concerned with when and how conflicts end.<sup>16</sup> Some data-collection projects that do not examine conflict are frequently used by researchers in combination with conflict data sets to test, for example, hypotheses on the relationship between the type of regime and violent conflict.<sup>17</sup> The diverse data sets that illuminate possible causes of conflict initiation, continuation and termination are useful for addressing questions related to conflict management and resolution. The current weakness of these data sets is that they are only just beginning to provide detailed information on the particular actions and motives of the antagonists.<sup>18</sup>

A third type of effort, to collect data known as 'events data', gathers very detailed information on the impact of violence and factors thought to influence the outbreak of violence. Event data projects attempt to record all interactions chronologically, without grouping them into distinct cases of conflict. When these data are combined with the knowledge of area specialists, the output can provide warning of potential outbreaks of violence.<sup>19</sup> The greatest shortcoming of events data projects is that they rely on a very limited number of media sources, so their content is skewed by the editorial decisions of media outlets, which are beyond their control. Also, owing to the quantity of material handled, coding is done by computers and must be checked by the user.<sup>20</sup>

of small arms are available from the Norwegian Initiative on Small Arms Transfers (NISAT) at URL <<http://www.nisat.org/>>. The Facts on International Relations and Security Trends (FIRST) project at SIPRI, available at URL <<http://first.sipri.org/>>, combines information from a number of databases to provide security-related country profiles. The Armed Conflict and Intervention Project (see section III) is currently working on the provision in a single location of information on conflicts, displaced populations, trade flows, political interactions, membership of international organizations and military interventions.

<sup>13</sup> These 3 topics are treated by the Issue Correlates of War project, the Minorities at Risk Project, and the International Crisis Behavior Project and Behavioral Correlates of War projects, respectively.

<sup>14</sup> The Third Party Interventions in Intrastate Conflicts project.

<sup>15</sup> The Armed Conflict and Intervention Project and the Minorities at Risk project.

<sup>16</sup> The Violent Intrastate Nationalist Conflicts project.

<sup>17</sup> Political Regime Characteristics and Transitions, 1800–1999 (Polity IV), available at URL <<http://www.bsos.umd.edu/cidcm/polity/>>.

<sup>18</sup> Collier, P. and Hoeffler, A., 'Data issues in the study of conflict', Paper presented at the Identifying Wars Conference (note 5).

<sup>19</sup> The Swiss Peace Foundation sponsors Früh-Analyse von Spannungen und Tatsachenermittlung (FAST). Its objective is early recognition of impending or potential crisis situations for the purpose of early action and conflict prevention. The project combines multiple methods, including events data analysis, field investigation and consultation of experts. URL <[http://www.swisspeace.ch/html/navigation/fr\\_program\\_fast.html](http://www.swisspeace.ch/html/navigation/fr_program_fast.html)>.

<sup>20</sup> There are also systematic, computerized, text-oriented projects intended to assist conflict prevention and management. They are distinct from data-collection projects in that they are organized around case studies. The University of Southern California provides access to 3 such projects (SHERFACS, a database compiled by Frank Sherman; Computer-Aided System for Analyzing Conflict (CASCON); and

## Definitions

Even data sets of the same type can present different views of conflict. For example, among the projects that study the global pattern of conflict, a comparison of the Correlates of War, Conflict Data Project, Conflict Simulation Model (Konflikt-Simulations-Modell, KOSIMO) and Arbeitsgemeinschaft Kriegsursachenforschung (AKUF) data sets demonstrates a low correlation between them on the basic matter of the number of interstate and intra-state wars between 1950 and 1999.<sup>21</sup> However, the pattern of rise and decline in the number of conflicts over time is similar in each of the four projects.<sup>22</sup> Disagreements among data sets of the same type are usually the result of different definitions of variables and coding rules.<sup>23</sup> As the definitions of the dependent variables for the 16 data projects presented in section III indicate, there are considerable differences between them. Depending on one's perspective, variation among definitions is either a significant problem or an opportunity.

Lack of consistency is a problem if the reader wants to consult a single authoritative source: there is none. More importantly, some analysts are concerned that the variation indicates basic theoretical differences. For example, does a state have to be one of the antagonists for a violent event to have political significance? These differences are often not made explicit, so data projects cannot be compared on the basis of full information. This undermines the possibility of the progressive accumulation of knowledge, which is the essential purpose of systematic studies of human behaviour.<sup>24</sup>

Variation among data sets of the same type is an asset if one wants to test the robustness of a hypothesis by subjecting it to two different sets of data. If a hypothesis yields the same answer with two different data sets, it can be considered stronger than if it is tested against only one. If the two tests yield different answers, then both the hypothesis and the data sets need to be examined more closely.<sup>25</sup> This process can lead to stronger hypotheses as well as to more refined and accurate data-collection projects.

It is instructive to look at fatality thresholds in this regard.<sup>26</sup> Several projects use a fatality count to determine the degree of violence.<sup>27</sup> Once a specified fatality threshold is crossed, the event is counted in the data set. This method of operationalizing the definition of violent conflict raises several questions. How many people must be killed before an event is considered significant enough to be included? Should the count be on an annual or overall basis? The Correlates of War project requires

HAAS, a database compiled by Ernst Haas) through its Prototype Action Recommenders' Information Support (PARISinLA) project, available at URL <<http://www.usc.edu/dept/ancntr/Paris-in-LA/>>.

<sup>21</sup> Eberwein, W.-D. and Chojnacki, S., 'Scientific necessity and political utility: a comparison of data on violent conflicts', Discussion paper P01-304, International Politics Working Group, Wissenschaftszentrum-Berlin (Sep. 2001), available at URL <<http://www.wz-berlin.de/ip/pub.en.htm>>.

<sup>22</sup> Gleditsch *et al.* (note 6).

<sup>23</sup> The use of different sources is another likely reason for disagreements. The limited amount of information about the sources used by projects is discussed in the next section.

<sup>24</sup> Eberwein and Chojnacki (note 21).

<sup>25</sup> Collier and Hoeffler (note 18).

<sup>26</sup> Sambanis, N., 'A note on the death threshold in coding civil war events', Conflict Processes Newsletter online (June 2001), available at URL <<http://mailer.fsu.edu/~whmoore/cps/newsletter/june2001/Sambanis.html>>.

<sup>27</sup> The Conflict Data Project, whose data have been published in the SIPRI Yearbook since 1987, uses a threshold measure of fatalities. Among the data projects summarized in section III, fatality thresholds are used by 3 out of 5 projects in the 'patterns of conflict occurrence' category and by 2 out of 8 projects in the 'causes and processes of conflict' category.

1000 battle-related deaths over the course of the conflict for interstate wars and 1000 battle-related deaths in a single year for intra-state and extra-systemic (imperial and anti-colonial) wars. The Major Episodes of Political Violence data set uses a fatality measure of 500 'directly related' fatalities over the course of the conflict, together with substantial destruction of infrastructure and population displacement. The Conflict Data Project distinguishes three levels of violence that combine annual and total battle-related deaths. Minor conflicts cause at least 25 battle-related deaths in a year, but fewer than 1000 overall; intermediate conflicts cause more than 1000 battle-related deaths overall, but fewer than 1000 in any single year; and wars cause at least 1000 battle-related deaths in a single year. The lack of direct comparability between the three projects is obvious.

The seemingly simple question of what to count as a conflict fatality is also open to interpretation. These three projects count only 'battle-related' deaths, which excludes civilian massacres as well as any other deaths that were not directly caused by battle, even if they were brought on by military actions.<sup>28</sup> Other projects use a broader definition of deaths but sometimes do not clearly specify what they include.<sup>29</sup> Differences over what to count as a conflict death have a direct bearing on whether and when a fatality threshold is reached. Then there is the question of whether the threshold should be defined in absolute or relative (for example, per capita) terms. Violence that killed hundreds of people in the early 1980s shook the foundations of the state in Surinam (population *c.* 500 000). The violent death of hundreds of people in India (population *c.* 1008 million), as has occurred many times, is a tragedy but not a political crisis. The point is that absolute measures fail to capture the cost differences for small and large countries.<sup>30</sup>

All of this demonstrates that data sets are open to the challenge that the number of deaths may not be the best way to determine the severity of a conflict.<sup>31</sup> The context of the violence may be critically important for determining its severity. For example, the approximately 3000 people killed by the terrorist attack on the World Trade Center in New York City on 11 September 2001 had a far larger impact on world politics than the 60 000 killed in Sri Lanka's civil war over the past 18 years. An inability to incorporate the contextual dimension of behaviour is a fundamental problem for all quantitative research, and conflict data projects can do little to resolve it. Researchers who use data sets would be well advised to keep in mind the inherent limitations of such projects. As an illustration, nearly all data projects, even those that use several measures of severity, treat war as a distinct phenomenon recognizable by the absolute number of people killed. Yet this begs the question of whether war really is a phenomenon distinct from other types of political violence. Wars usually begin as lesser conflicts that escalate. In what ways are the dynamics of war and war termination different from the dynamics and termination of lower levels of violence? Does treating war as distinct reduce our ability to understand it? The projects summarized in the section below on the causes and processes of conflict attempt to address these questions.

<sup>28</sup> In an extreme example of the distortion that a definition can impose, the Conflict Data Project did not register the deaths from the genocide in Rwanda, even though it registered the existence of a major armed conflict in that country in 1994. The genocide deaths were not considered to be battle-related.

<sup>29</sup> The KOSIMO project.

<sup>30</sup> Collier and Hoeffler (note 18).

<sup>31</sup> The data projects located at the Center for International Development and Conflict Management (CIDCM), University of Maryland, are particularly innovative in using other measures of severity in addition to or in place of fatalities.

The definitions a data project uses, like the project's purpose, can determine the value of any particular data set for answering a research question. For reasons of consistency over time, data-collection projects must adhere to a predetermined set of variables and definitions for those variables. As a consequence, they reveal a consistent view (or frame) of conflict, with variation occurring within the frame.<sup>32</sup> They do not reveal changes outside the frame. One example is that most projects use definitions of conflict that require a state to be one of the antagonists.<sup>33</sup> Violent conflicts in which a state is not an antagonist, such as inter-communal violence, are excluded. If one wants to investigate whether the role of the state in violent conflict is in decline, it would be useful to know the relative frequencies of conflicts involving states and not involving states. Whether or not a data set can provide the information depends on the definition of conflict used.

### Methods and coding rules

In broad terms, all data producers follow the same method. They gather and examine copious amounts of information on actual or potential conflicts and conflict actors, pick out material related to their study variables and code the information according to a set of rules. Exactly where project researchers get their information is somewhat of a mystery, since virtually none of them offers a specific listing of sources used to create its database.<sup>34</sup> Given the limited resources available to them and the global scope of all the projects listed here, it is safe to assume that project personnel do not have a network of colleagues in the field who feed them first-hand information. It is likely that they rely primarily on news media reports and secondary sources.<sup>35</sup> In some cases this type of source is supplemented by consulting regional experts and using field reports produced by non-governmental and international organizations. This situation raises a potential problem when combined with the fact that the vast majority of data projects are located in the United States, and apparently all the projects are Western. In all likelihood, different projects use largely the same set of sources, such as major US and European newspapers and reports from respected organizations. Any bias in news reporting will influence the information in every data project, making the entire enterprise a less reliable reflection of reality. This situation is in the process of changing with the advent of the Internet. Lexis-Nexis Academic Universe is a widely used Internet-based service that provides subscribers with access to several hundred news sources from around the world.<sup>36</sup>

<sup>32</sup> In practice, data projects occasionally adjust their definitions and refine their categories. Sarkees, M. R., 'The Correlates of War data on war: an update to 1997', *Conflict Management and Peace Science*, vol. 18, no. 1 (fall 2000), pp. 123–44. SIPRI adjusted the way it uses data collected by the Conflict Data Project, starting in 2000 (in this volume, see appendices 1A and 1B). A major armed conflict is now defined as one that caused at least 1000 deaths in a year rather than over the entire course of the conflict.

<sup>33</sup> Exceptions are the Major Episodes of Political Violence project, the Minorities at Risk project and the Violent Intrastate Nationalist Conflicts project. The Correlates of War project is currently expanding its typology to include conflicts with no state party. Sarkees, M. R. and Singer, J. D., 'Armed conflict past and future: a master typology?', Paper presented at the Identifying Wars Conference (note 5).

<sup>34</sup> Events data projects explicitly rely on news media and do reveal their original sources.

<sup>35</sup> Some of the projects reviewed here combine data from other projects. The Major Episodes of Political Violence project is an example, as is the Conflict Data Project's extension of its data set from 1989 back to 1946. In cases such as these, the projects cite the data collections they use, but the sources used for the original data sets are still not known.

<sup>36</sup> Hensel, P., Private communication with the author, 6 Mar. 2002.

Discussion of the exact methods used by each project and of the differences between them is beyond the scope of this survey. Instead, the remainder of this section makes three brief points, all of which carry the message that quantitative analytical understanding of conflict is highly subject to coding rules.

First, the duration of conflicts is a topic of interest to many researchers and policy makers. Are the durations of conflicts longer or shorter than in the past? Why do some conflicts last longer than others? Are the factors that explain the outbreak of a conflict the same as those that explain its duration? If a conflict appears to subside and then flares up again, is this a recurrence of the same conflict or a new conflict? Analyses of duration and related questions require that the start and end of conflicts be dated.<sup>37</sup> Some projects mark the beginning of a conflict from the initiation of sustained fighting that leads to the fatality threshold. Projects that do not use thresholds count the start of a conflict from the point of observable incompatibility. Still others wait until a conflict has crossed a threshold and then backdate it to the point of stated incompatibility.<sup>38</sup> In similar fashion, various coding rules are used to determine the end of a conflict and whether additional fighting is a new conflict or the continuation of an old one.<sup>39</sup>

Second, increased attention has been paid to conflict management and resolution during the past decade. A common analytical approach divides conflicts into phases of escalation and de-escalation. What factors lead from one phase to the next? Are some types of conflict more likely to escalate than others? Which actions by outside actors lead to de-escalation (or escalation) of the violence and under what circumstances? Analysis of conflict dynamics requires the demarcation of phases.<sup>40</sup> Some projects do not distinguish levels of intensity at all, so a conflict is either 'on' or 'off'.<sup>41</sup> Projects that do measure changes in intensity use additive coding rules so that, once a conflict has accumulated enough fatalities to cross into the next category of violence, it cannot be re-coded in a lower category if it diminishes in intensity.<sup>42</sup> Some projects use an ordinal scale to incorporate several measures to indicate the impact of a conflict on society.<sup>43</sup> These have great potential for the study of conflict management, but they are often used to mark the highest level of violence reached.<sup>44</sup>

Third, a fundamental principle of systematic, or scientific, research projects is that they are objective, transparent and consistent. These standards are not always maintained by conflict data-collection projects since they work to maintain the balance between validity (relevance) and reliability. The definitions of some variables require personal interpretation by the coder. For example, when determining the number of battle-related deaths, projects that count civilians must identify the line between civilians killed by cross-fire (battle-related) and massacres (not battle-related).<sup>45</sup> Some projects attempt to avoid definitional restrictions by setting

<sup>37</sup> Collier and Hoeffler (note 18).

<sup>38</sup> An example of each approach is the Correlates of War project, the KOSIMO project and the Conflict Data Project, respectively.

<sup>39</sup> The Violent Intrastate Nationalist Conflicts project pays special attention to conflict endings.

<sup>40</sup> A project that explicitly addresses conflict escalation in terms of phases is SHERFACS (note 20).

<sup>41</sup> The Correlates of War project.

<sup>42</sup> The Conflict Data Project allows movement from 'war' down to 'intermediate' but not to 'minor' conflict.

<sup>43</sup> The Minorities at Risk and the Major Episodes of Political Violence projects.

<sup>44</sup> The Violent Intrastate Nationalist Conflicts project.

<sup>45</sup> The various Correlates of War projects do not count civilian deaths, even if they occur in a battle.

qualitative rather than quantitative parameters for variables.<sup>46</sup> However, variables that are not strictly defined require personal interpretation and may result in the establishment of categories that overlap or are not mutually exclusive. In another example, grievances are often viewed as a cause of conflict, but few objective measures of grievance are available.<sup>47</sup> Even when there is a precise definition for a variable, it may be difficult in practice to gather reliable information. The number of people killed in a conflict is a central piece of information for many projects, but it is notoriously hard to determine with certainty.<sup>48</sup> This is not to say that systematic data collections are inherently unreliable, only that they are not foolproof. Most researchers are aware of these problems, and the projects listed below appear to do an admirable job of collecting, coding and verifying the information they use.

### III. The leading data sets

The 16 data sets described below are grouped according to whether they focus primarily on the patterns of conflict occurrence, the causes and processes of conflict, or conflict early warning. The categories are not exclusive and are intended only as a rough guide. This list is not comprehensive, but an attempt has been made to make it complete within certain parameters. Every data set is directly concerned with conflict, provides worldwide coverage, is publicly available in English and is widely judged to be reputable. A large number of data sets created for specific publications and made available by authors are not included here because they constitute data use rather than data collection.<sup>49</sup>

#### **Patterns of conflict occurrence<sup>50</sup>**

##### *Correlates of War (COW and COW<sup>2</sup>)*

*Location:* University of Michigan, Ann Arbor, Michigan, USA, and Pennsylvania State University, State College, Pennsylvania, USA.

*Principal investigators:* J. David Singer, University of Michigan; and Stuart Bremer, Pennsylvania State University.

*Purpose:* To promote and support the scientific study of the causes of war and the conditions of peace by collecting and processing large quantities of historical information in an attempt to identify and explain empirical regularities that lead to war.

<sup>46</sup> The KOSIMO project, e.g., defines a conflict as consisting of 'some duration' and 'magnitude'.

<sup>47</sup> Collier and Hoeffler (note 18).

<sup>48</sup> The 1999 NATO bombing in the Federal Republic of Yugoslavia is a case in point. It was a short-duration event in a small territory with a large number of independent observers. Nevertheless, a careful study could only estimate the number of people killed by all sides during the period as in the range of 7449–13 627 (with 95% confidence). Projects that use a single number are likely to adopt the study's best estimate of 10 500, but there is a small chance that it could be incorrect. American Bar Association and American Academy for the Advancement of Science, *Political Killings in Kosova/Kosovo, March–June 1999* (ABA Central and Eastern European Law Initiative: Washington, DC, 2000), pp. 7–8. The mortality in most conflicts has not been so exhaustively analysed.

<sup>49</sup> A notable source for research on conflicts that uses data extensively but is not a data-collection project is the World Bank project *The Economics of Civil Wars, Crime and Violence*, available at URL <<http://www.worldbank.org/research/conflict/index.htm>>.

<sup>50</sup> In the interest of accuracy, the wording of the descriptions of the purpose and dependent variables of each project is as close as possible to that of the project's own presentation. The information in this section was gathered from project responses to a questionnaire, supplemented with information from project Internet sites.



Interstate conflict is the special focus of the project, with emphasis on those conflicts that involve the threat, use or display of force. Intra-state and extra-systemic conflicts are also studied.

*Current coverage:* 1816–1997.

*Dependent variables:* (a) Interstate war is sustained combat between the regular military forces of two or more state members of the international system in which there is a total of at least 1000 battle-related fatalities. (b) Intra-state war is sustained armed combat between two armed forces within the boundaries of a state, in which there are at least 1000 battle-related fatalities per year. (c) Extra-systemic war is sustained armed combat between a state member of the international system and a non-system-member political entity outside its territorial boundaries, in which there are at least 1000 battle-related fatalities per year.

*Availability:* URL <<http://cow2.la.psu.edu>> provides access to the most recent data sets, code books, history and contact information. URL <<http://www.umich.edu/~cowproj/>> provides earlier data sets, code books, publications and contact information.

### *Conflict Data Project*

*Location:* Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden.

*Principal investigators:* Peter Wallensteen and Margareta Sollenberg.

*Purpose:* To collect information on selected variables relating to armed conflict, primarily to be used in research on various aspects of the origins, dynamics and resolution of conflict. Data have been collected on a global and yearly basis since 1989.

*Current coverage:* 1989–2001. The project has recently collaborated with others to extend the coverage from 1946 to 1988.

*Dependent variables:* Armed conflict is a contested incompatibility that concerns government or territory or both, over which the use of armed force between the military forces of two parties results in battle-related deaths. At least one of the parties is the government of a state. (a) Minor armed conflict results in at least 25 deaths per year and fewer than 1000 deaths over the course of the conflict. (b) Intermediate armed conflict results in more than 1000 deaths during the course of the conflict, but fewer than 1000 in any given year. (c) War results in more than 1000 deaths in any given year.

*Availability:* URL <<http://www.pcr.uu.se/research/data.htm>> provides access to a data set, code book, definitions, summary table and contact information. The data set covering the extended period of 1946–2001 will be available in late 2002 on the Internet site of the International Peace Research Institute, Oslo (PRIO) at URL <<http://www.prio.no>>.

### *Conflict Simulation Model (Konflikt-Simulations-Modell, KOSIMO)*

*Location:* Heidelberg Institute of International Conflict, University of Heidelberg, Heidelberg, Germany.

*Principal investigator:* Frank R. Pfetsch.

*Purpose:* To provide a searchable database of political conflicts including crises, wars, insurrections, negotiation, mediation and peace settlements.

*Current coverage:* 1945–99.

*Dependent variables:* Political conflict is defined as the clashing of overlapping interests around national values and issues between at least two parties, at least one of which is the organized state. The conflict has to be of 'some duration' and 'magnitude'. The intensity ranges from 'latent conflict' to 'non-violent crisis' to 'violent crisis' to 'war'. Possible instruments used in the course of a conflict are negotiations, authoritative decisions, threat, pressure, passive or active withdrawals, or the use of physical violence.

*Availability:* URL <<http://www.hiik.de/en/kosimo/kosimo.htm>> provides access to a searchable database, summary graphs, publications and contact information.

### *Major Episodes of Political Violence (MEPV)*

*Location:* Center for Systemic Peace, University of Maryland, College Park, Maryland, USA.

*Principal investigator:* Monty G. Marshall.

*Purpose:* To list all episodes of major political violence of any type. Categories include all forms of interstate, intra-state and inter-communal warfare. This data set is one of six data sets that comprise the Armed Conflict and Intervention Project at the Center for Systemic Peace and the Center for International Development and Conflict Management (CIDCM), University of Maryland. The larger project attempts to capture the deeper qualities and complexities of violent social conflict. It collects global information on the security context, membership of international organizations, displaced populations, direct military interventions, political interactions and bilateral trade flows.

*Current coverage:* 1946–2000.

*Dependent variables:* Major episodes of political violence involve the systematic use of lethal violence and terror by organized groups and/or states that substantially affect the society or societies that directly experience the armed conflict (resulting in at least 500 directly related fatalities, substantial destruction of infrastructure and population displacements). Episodes may involve states, a state and non-state group, or non-state groups only, including interstate and independence war, ethnic and revolutionary (civil) war, inter-communal warfare, genocide and communal massacres. Each episode is rated on a 10-point scale according to its total impact on the society or societies that are directly affected by the violence.

*Availability:* URL <<http://members.aol.com/CSPmngm/warlist.htm>> provides access to a data set, code book, summary table and contact information. In addition to the MEPV data set, the Armed Conflict and Intervention Project provides access to data sets on membership of international organizations and displaced populations. Data sets on military interventions, political interactions and bilateral trade flows are scheduled to be made publicly available in the future.

## **Causes and processes of conflict**

### *International Crisis Behavior Project*

*Location:* University of Maryland, College Park, Maryland, USA and McGill University, Montreal, Quebec, Canada.

*Principal investigators:* Michael Brecher, McGill University, and Jonathan Wilkenfeld, University of Maryland.

*Purpose:* To investigate 20th century interstate crises and the behaviour of states under externally generated stress. The data describe the sources, processes and outcomes of all military–security crises involving states.

*Current coverage:* 22 December 1917 to 31 December 1994.

*Dependent variables:* Part 1: All international crises occurring during the coverage period, characterized by: (a) a distortion in the type and an increase in the intensity of disruptive interactions between two or more adversaries, with an accompanying high probability of military hostilities or, during a war, an adverse change in the military balance; and (b) a challenge to the existing structure of an international system—global, dominant or sub-system—posed by the higher-than-normal conflictual interactions. Part 2: All foreign policy crises experienced by states as a result of their involvement in the international crises defined above. A foreign policy crisis is defined as a situation in which three conditions, deriving from a change in a state's external or internal environment, are perceived by the highest-level decision makers of the state: (a) a threat to basic values, (b) an awareness of finite time for response to the external threat to basic values, and (c) a high probability of involvement in military hostilities.

*Availability:* URL <<http://www.missouri.edu/~polsjhh/ICB>> provides access to data sets, code books, summary tables, papers and contact information.

### *Correlates of War–Militarized Interstate Disputes (MID 3)*

*Location:* Pennsylvania State University, State College, Pennsylvania, USA.

*Principal investigators:* Stuart Bremer, Jim Ray, Dan Geller, Paul Diehl, Doug Gibler, Paul Hensel, Chuck Gochman, Glenn Palmer, Brian Pollins, Ric Stoll, Pat Regan and Zeev Maoz.

*Purpose:* To identify for all militarized interstate disputes the participants, start and end dates, fatality totals, hostility levels, revision sought, outcome and method of settlement.

*Current coverage:* 1816–1992. Currently being updated to 2001.

*Dependent variables:* A militarized interstate dispute involves the threat, display or use of force short of war by one member state, explicitly directed towards the government, official representatives, official forces, property or territory of another state. The outcome variable is recorded on a five-point ordinal scale ranging from non-reciprocated action, to the threat, display or use of force, to interstate war.

*Availability:* Data for 1816–1992 are available at URL <[http://pss.la.psu.edu/MID\\_DATA.HTM](http://pss.la.psu.edu/MID_DATA.HTM)>. Update to 2001 available from late 2002 at URL <<http://mid3.la.psu.edu/>>; this site provides access to papers, operational definitions and coding procedures (in the paper by Daniel M. Jones, Stuart A. Bremer and J. David Singer), progress reports, a discussion forum and contact information.

### *Behavioral Correlates of War (BCOW)*

*Location:* Middlebury College, Middlebury, Vermont, USA.

*Principal investigator:* Russell Leng.

*Purpose:* To analyse the behaviour of states engaged in potential pre-war disputes. Data on 47 cases focus on crisis dynamics by generating descriptive data on the actions and interactions of states. The cases are selected as a representative sample of crises since 1816 and they permit the testing of a number of theories of interstate crisis behaviour.

*Current coverage:* 1838–1980.

*Dependent variables:* Interstate crises in which the principal protagonist on each side is a member of the interstate system. A crisis is an event on the continuum of belligerence that extends from a simple dispute, to a militarized dispute, to a crisis, to war. A crisis is a militarized dispute that requires protracted bargaining, defined as when there are at least 50 exchanges between the two major participants.

*Availability:* URL <<http://community.middlebury.edu/~leng>> provides access to the project description, data sets, software needed for working with the data, a users' manual and contact information.

### *Issue Correlates of War (ICOW)*

*Location:* Florida State University, Tallahassee, Florida, USA.

*Principal investigators:* Paul Hensel and Sara McLaughlin Mitchell.

*Purpose:* To collect systematic data on contentious issues between states, with a focus on identifying the issues regardless of any particular action that may or may not have been taken to resolve them. Presently covering territorial, riverine and maritime claims.

*Current coverage:* 1816–2000.

*Dependent variables:* Contentious issues which involve explicit statements of disagreement by official governmental representatives of at least two states. ICOW identifies each issue with reference to the involved states, the object of the claim (such as the specific river or territory) and the time frame over which it endures. ICOW then collects data on the salience of each issue and on attempts to settle it through peaceful bilateral negotiations, binding or non-binding third-party activity or militarized conflict.

*Availability:* URL <<http://www.icow.org>> provides access to the project description, data sets, code books, publications and contact information.

### *Rivalry Data Set*

*Location:* Department of Political Science, University of Illinois, Urbana, Illinois, USA.

*Principal investigators:* Paul Diehl and Gary Goertz.

*Purpose:* This data set provides a comprehensive overview of 1166 rivalries, 63 of which are enduring. The data set is designed to provide the basis for the analysis of the initiation, dynamics and termination of international rivalries.

*Current coverage:* 1816–1992.

*Dependent variables:* Rivalry is defined by the frequency of militarized interstate disputes between the same pair of states. The existence of a militarized rivalry is indicated by the occurrence of militarized disputes as defined by the COW–MID data set. Disputes which occur within 10–15 years of each other are considered to be part of the same rivalry. A dispute is considered part of the same rivalry if it involves the same two states and occurs within 11 years of the first dispute of the sequence, within 12 years of the second dispute of the sequence, and within up to 15 years of the fifth dispute in the sequence.

*Availability:* URL <<http://www.pol.uiuc.edu/faculty/Diehl/diehl3lnk.htm>> provides access to data sets, explanatory information and contact information.

*Internal Wars and Failures of Governance: State Failure Data Set*

*Location:* Center for International Development and Conflict Management (CIDCM), University of Maryland, College Park, Maryland, USA.

*Principal investigators:* Monty G. Marshall, Ted Robert Gurr, Jack A. Goldstone and Barbara Harff.

*Purpose:* To provide the dependent variable (state failure) in the US Government-sponsored State Failure Task Force quantitative analyses of structural indicators of failure, the purpose of which is to create 'early warning' (two-year) models of state failure situations. Independent variables used in published Task Force analyses are being prepared for public release and should be available soon to complement the dependent variable data on state failures.

*Current coverage:* 1955–2000.

*Dependent variables:* The data set includes all cases of internal wars (ethnic war, revolutionary war, or genocide–politicide) and failures of governance (substantial reversion to more autocratic rule or collapse of central authority) that began between 1955 and 2000 in independent countries with populations greater than 500 000. A war is defined as an armed conflict involving state authorities and a challenger group that results in at least 1000 directly related deaths over the course of the episode and at least one year during which there were more than 100 directly related deaths. The case begins with the first year during which the 100-death threshold is reached and ends when deaths fall below that threshold for at least five years. An episode of genocide or politicide (politically motivated mass murder) is defined by the merits of the case (for instance, an established intent to eliminate non-combatant group members). A failure of governance is defined generally as a six-point decrease in the state's Polity IV regime score (that is, towards greater institutional autocracy) or a Polity IV 'interregnum' (a collapse of central regime authority through failure, revolution or involuntary state disintegration).

*Availability:* URL <<http://www.bsos.umd.edu/cidcm/inscr/>> provides access to data sets, code books, summary tables and contact information.

*Minorities at Risk*

*Location:* Center for International Development and Conflict Management (CIDCM), University of Maryland, College Park, Maryland, USA.

*Principal investigators:* Ted Robert Gurr, Monty G. Marshall and Christian Davenport.

*Purpose:* To monitor and analyse the status and conflicts of politically active ethno-political groups in countries with a population of at least 500 000. Coverage of 275 contemporary and 65 historical groups.

*Current coverage:* 1945–2000.

*Dependent variables:* Ethno-political groups, defined as communal groups that: (a) are disadvantaged by comparison with other groups in their societies, usually because of discriminatory practices, or (b) have organized politically to promote or defend their collective interests. Only ethno-political groups with populations greater than 100 000 or 1 per cent of the population are included.

*Availability:* URL <<http://www.bsos.umd.edu/cidcm/mar/>> provides access to data sets (registration required), a code book, publications, project history and contact information.

*Violent Intrastate Nationalist Conflicts (VINC)*

*Location:* University of Indianapolis, Indianapolis, Indiana, USA.

*Principal investigator:* Bill Ayres.

*Purpose:* To measure and study conflict outcomes in violent, intra-state nationalist conflicts—those involving ethnic and other forms of secessionism—and their antecedents and correlates. It seeks to mark starting and ending points of these conflicts, to measure the characteristics and behaviour of the actors in them, and to answer questions about how and why conflicts end the way they do.

*Current coverage:* 1945–96.

*Dependent variables:* Each conflict episode is coded for the highest level of violence or rebellion reached during that episode, using a seven-point ordinal scale adapted from the Minorities at Risk data set project. Each conflict episode is coded for estimated number of deaths caused by that conflict (rounded to the nearest 100 for estimates less than 10 000, rounded to the nearest 1000 for estimates greater than 10 000). Conflict ends when both sides are no longer either fighting or talking with each other about what the solution to the conflict should be. Different episodes of the same conflict must be separated by a 12-month lull in both fighting and negotiating. The data also specify four types of ending and four types of agreement.

*Availability:* URL <<http://facstaff.uindy.edu/~bayres/vinc.htm>> provides access to data sets, a code book, summary tables, papers and contact information.

*Third Party Interventions in Intrastate Conflict*

*Location:* University of Binghamton, Binghamton, New York, USA.

*Principal investigator:* Patrick M. Regan.

*Purpose:* To identify all military, economic and diplomatic interventions in civil conflicts, primarily to help determine the relationship between military, economic and diplomatic interventions and the duration of conflict. Current coverage includes only military and economic interventions.

*Current coverage:* 1945–99.

*Dependent variables:* Third-party interventions in intra-state conflicts are convention-breaking military and/or economic activities in the internal affairs of a foreign country targeted at the authority structures of the government with the aim of affecting the balance of power between the government and opposition forces. Intrastate conflicts are organized military hostilities between two groups in conflict in which there were at least 200 fatalities over the course of the conflict.

*Availability:* URL <<http://bingweb.binghamton.edu/~pregan/>> provides access to a data set, user's manual and contact information.

**Costs of conflict and conflict early warning**

*Protocol for the Assessment of Nonviolent Direct Action (PANDA)*

*Location:* Program on Nonviolent Sanctions and Cultural Survival (PONSACS), Harvard University, Cambridge, Massachusetts, USA.

*Principal investigators:* Doug Bond, Joe Bond, J. Craig Jenkins, Churl Oh and Charles Louis Taylor.

*Purpose:* PANDA is an automated early warning system that is combined with on-the-ground research of conflict regions provided by anthropological insights. These two strands of research at PONSACS work to identify conflict regions before they erupt into violence and to actively promote non-violent alternatives to armed conflict. The project's premise is that by monitoring and examining interaction events with a 'data lens' that is sensitive to non-violent direct action, it can track and compare the evolution of conflict manifest in both violent and non-violent behaviour. The project also seeks to help make the costs of conflict transparent by providing a longitudinal series of social, political and economic events gleaned from news reports, and to facilitate independent testing and peer review. The PANDA software protocol for parsing news stories has been superseded by the Integrated Data for Events Analysis (IDEA) protocol.

*Current coverage:* 1991–2000.

*Dependent variables:* The basic parameters of the data include the source actor and target actor of social, political and economic events, the events themselves, as well as their date, location and a selection of attributes of the same. In more common terms, each data record represents the 'who does what to/with whom, when, where, why and how' of an event reported in the news. Any of the events data variables may be treated variously as an independent or dependent variable, depending upon the specific research questions being asked.

*Availability:* The code book, data files, protocol files, publications and contact information are available at URL <<http://www.wcfia.harvard.edu/ponsacs/panda.htm>>.

### *The Global Event-Data System (GEDS)*

*Location:* Center for International Development and Conflict Management (CIDCM), University of Maryland, College Park, Maryland, USA.

*Principal investigator:* John Davies.

*Purpose:* To allow computer-assisted identification, narrative description and analytical coding of daily international and intra-national events, describing the day-to-day actions of all states and the major non-state communities and international organizations. GEDS includes and expands on Edward Azar's Conflict and Peace Data Bank (COPDAB, covering 1948–78), updating it for selected countries. Near-real-time tracking, as needed for early warning, can be generated on request using COPDAB scales or more specific 'accelerator' models for anticipating ethno-political conflicts, genocides or politicides.

*Current coverage:* 1948 to the mid-1990s.

*Dependent variables:* Events are operationally defined as reports from reputable sources which specify who did or said what to whom, when and where. Conflict and cooperation are operationally defined and coded using Azar's 15-point COPDAB scales either as categorical variables, as ordinal variables, or as ratio variables.

*Availability:* URL <<http://geds.umd.edu/geds/>> provides access to data sets, code book and contact information. The software is currently off-line. Updated data for some countries can be generated on request (at cost). This can include near-real-time tracking.

*The Kansas Events Data System (KEDS)*

*Location:* University of Kansas, Kansas City, Kansas, USA.

*Principal investigator:* Philip A. Schrodtt.

*Purpose:* To generate political event data through automated coding of English-language news reports. These data are used in statistical early-warning models to predict political change. Building on the World Event/Interaction Survey (WEIS) Project, the KEDS project has three major research concentrations: software development for the machine-coding of political event data, production of events data sets and development of early-warning methods.

*Current coverage:* 1979–99.

*Dependent variables:* Event data are nominal or ordinal codes recording the reported interactions between international actors at specific points in time.

*Availability:* URL <<http://www.ku.edu/~keds/>> provides access to software, data sets, papers and contact information.

## IV. Conclusions

*Caveat emptor*—let the user beware. In an ironic twist on the presumption of objectivity that underlies these quantitative research projects, the diversity of systematic data collection appears to support the constructivist argument that reality lies in the eye of the beholder.<sup>51</sup> To what extent do the data problems reviewed here affect researchers' ability to do good work? The core issue is the balance between reliability and validity, that is, between accuracy in recording information and appropriateness of the information for addressing theoretical concepts of interest. The balance confronts both quantitative and qualitative attempts to simplify the world in order to understand it and elicits different types of solutions from different types of researcher. Quantitative researchers place primary importance on reliability. To fulfil the requirement of systematically recording a series of events in a consistent manner (reliability), conflict data projects need to delimit complex phenomena through definitions and coding rules. In the process, they limit the range of their validity.

The problem of limited validity is partially resolved by the wide variety of data-collection projects that now exist. The primary purpose of this appendix is to point researchers towards the data sets that are appropriate for the questions they seek to answer. Some projects provide data that are appropriate for studies of the patterns of conflict occurrence and the structural features of the international system and its members that make violent conflict more or less likely. Other projects operate at a different level of analysis by focusing on the issues at stake and the comparative characteristics of the antagonists. The data produced allow for the development and testing of theories on the processes of conflict initiation, sustainment and resolution. A third type of project allows for analysis of foreign policy interactions by providing detailed information on events over time. These data-collection projects offer researchers a vast array of good data with which to develop academic theories and policy-related arguments.

<sup>51</sup> Eberwein and Chojnacki (note 21).