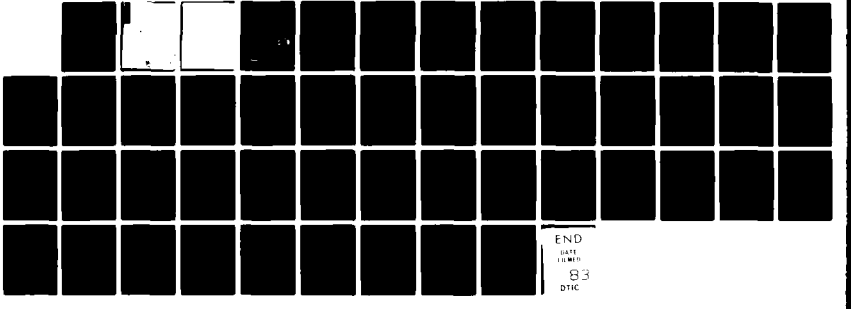
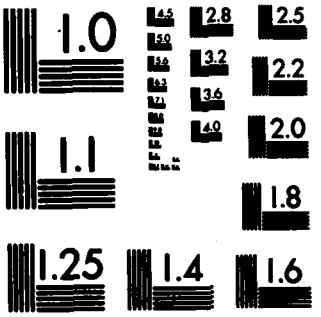
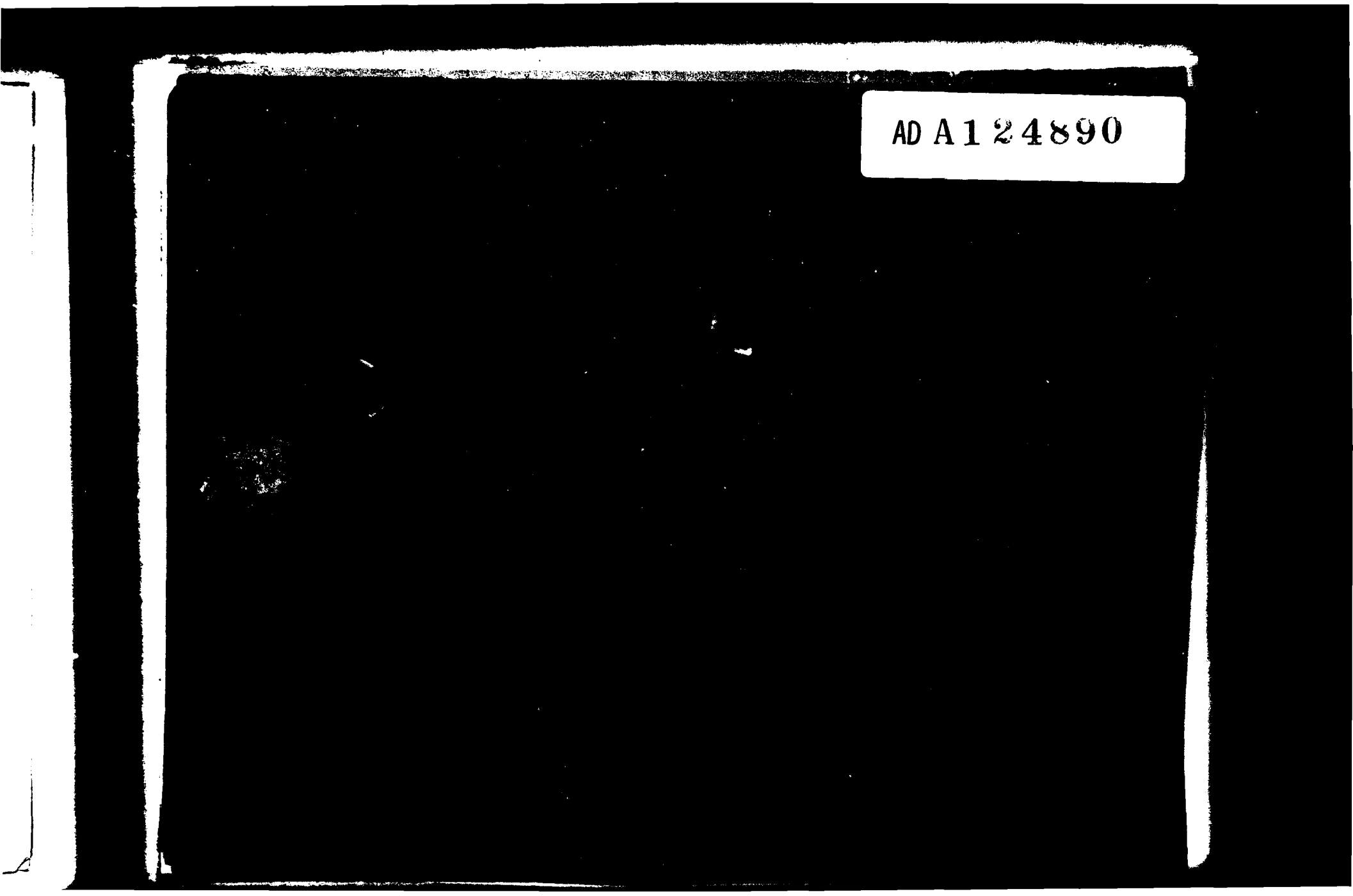
AD-A124 890 TERRORISM DATA BASES: A COMPARISON OF MISSIONS METHODS 1/1
AND SYSTEMS(U) RAND CORP SANTA MONICA CA W W FOWLER
MAR 81 RAND/N-1503-RC SBI-AD-E750 686

UNCLASSIFIED F/G 9/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



A RAND NOTE

TERRORISM DATA BASES:
A COMPARISON OF MISSIONS, METHODS, AND SYSTEMS

William Warner Fowler

March 1981

N-1503-RC

Rand
SANTA MONICA, CA. 90406

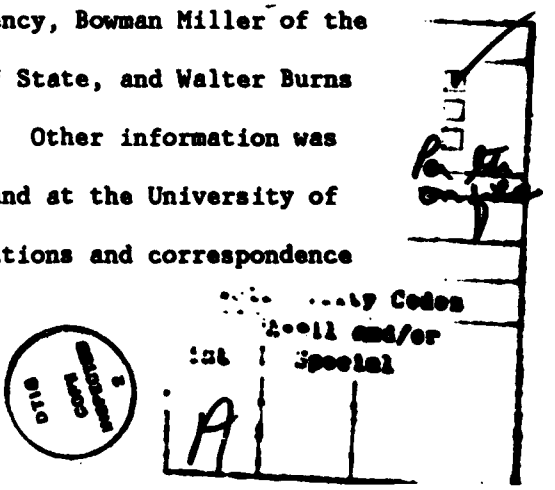
PREFACE

As part of its research on political violence, The Rand Corporation maintains a chronology of incidents of international terrorism. This data base now comprises over 1500 incidents from 1968 to 1980.

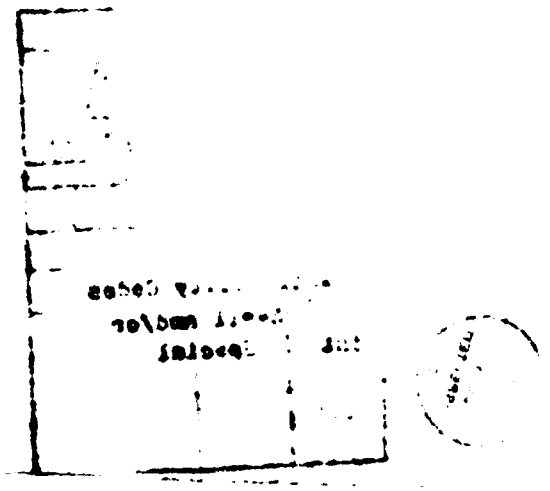
This Note describes the growing number of data bases devoted to the subject of terrorism. It examines the relationship between the use of terrorism data and the design of the data bases, the scope and content of the data bases, the structure of terrorism data, and the systems used for retrieval and analysis of the data.

The Rand chronology and similar information-gathering efforts directly or indirectly serve the interests of all sponsors of research in this area, including the Defense Nuclear Agency, the U.S. Department of Energy, and Sandia Laboratories. Periodic reporting on this aspect of its research is sponsored solely by The Rand Corporation. Future reporting will include periodic publication of the chronology itself as well as analyses of trends.

Most of the information contained in this Note was gathered during a trip to Washington, D.C., in mid-October 1979. Meetings were held with Michael Jaykub of the Counter Terrorism and Threat Analysis Branch of the Defense Intelligence Agency of the U.S. Department of Defense, Edward Mickolus of the Central Intelligence Agency, Bowman Miller of the Threat Analysis Group of the U.S. Department of State, and Walter Burns of Risks International in Alexandria, Virginia. Other information was derived from meetings with Dr. Charles McClelland at the University of Southern California and from telephone conversations and correspondence



with Russell William Mengel of BDM in McLean, Virginia, and Thomas Snitch of the American University. Donald Trees of Rand reviewed a draft version of this Note, made many useful suggestions, and drew out of the draft much of the explanatory material used in the present version. Any remaining errors of fact of interpretation are those of the author.



SUMMARY

This Note reviews eight terrorism data bases. It considers (1) the relationship between the use of terrorism data and the design of the data bases; (2) the scope and content of the data bases; (3) the structure of terrorism data; and (4) the systems used for retrieval and analysis of the data. Some conclusions are drawn about the state of the art in the design and implementation of terrorism data bases.

Terrorism data bases are primarily used in basic research on the underlying patterns and causes of terrorism, in making intelligence estimates of the probabilities of future terrorist activities, and in responding to specific queries about terrorist activities. While each of the data bases can be used for all of these applications, the present review classifies them according to their primary mission.

The major problems in the collection of terrorism data are the definition of terrorism itself and the determination of the scope and content of the data. Most of the data bases discussed here have implicit conceptual definitions of terrorism which were developed to meet the needs of particular missions or application domains. These definitions determine the scope and content of the data bases. The differences in scope are reflected in the wide variation of sample sizes.

Most data-collection efforts are centered on the development of chronologies of terrorist events. This means that the individual terrorist act is the unit of analysis. Various data bases include ancillary files in which locations, groups, and individuals are used as the units of analysis.

Terrorism data bases are implemented on a wide variety of access and analysis systems, from purely manual systems to sophisticated data-base management systems. Although the different systems do not necessarily reflect the state of the art in data-base system design, most users report a relatively high degree of satisfaction.

The greatest potential weakness of current data-collection efforts is the development of chronologies to the exclusion of other types of data bases and the lack of rigor in incident sampling. While chronologies do help analysts and researchers in addressing fundamental issues in the analysis of terrorism, they may not contain information relevant to many important questions. Data-collection efforts have matured to the point that the development of different kinds of data bases would be worthwhile.

CONTENTS

I. INTRODUCTION

Increasing amounts of resources are being committed to the collection of information about terrorism. Although social scientists often complain about the lack of empirical data relevant to the study of terrorism, the number of terrorism data bases is growing. However, this increase does not necessarily reflect a substantial increase in the amount of information available about terrorism. For the most part, data-gathering efforts are centered on the development of chronologies of recent terrorist activities. This means that there is an overlap in the information content of the currently available data bases. It has been estimated that there would be 90 percent agreement among the chronologies containing data on any given sample of incidents. The unique quality of most terrorism data bases is thus not the substance of the data, but the uses to which the information is put, the scope of the data, and the methods and tools used for collection, retrieval, and analysis.

The implementation of terrorism data bases on computer-based information systems is becoming more and more central to the analysis of terrorism data. Computers provide a powerful tool for developing relevant and interesting insights into important issues of subnational violence. The potential contribution of computer systems in the analysis of political, social, and historical data is significant and well recognized. They provide effective access to intricately structured information and analytical methods that would otherwise be impossible to implement. All the data bases discussed here are computer-based except the chronologies

of the Department of State; and the storage and retrieval of those chronologies will also be automated in the near future. Therefore, a significant part of this discussion will deal with issues related to the computerized retrieval and analysis of information. Unfortunately, detailed information describing the capabilities of these systems is not readily available.

Our working definition of the term "data base" is taken from Engles (1972): "A data base is a collection of stored operational data used in the application systems of some particular enterprise." An application system is a method (often, though not necessarily, computer-based) of accessing and analyzing data. This Note describes several data bases that support a variety of research interests and are used with a number of computer-based application systems.

The Note addresses some basic questions about the collection, maintenance, and use of terrorism data bases and illustrates them by describing eight data bases:

- The Chronology of International Terrorism of The Rand Corporation.
- The ITERATE data base (from International Terrorism: Attributes of Terrorist Events) compiled by Edward Mickolus.
- The FITE (File of International Terrorist Events) of the CIA.
- A terrorism chronology used by the Threat Analysis Group of the Department of State.
- A chronology used by the Counter Terrorism and Threat Analysis Branch of the Defense Intelligence Agency.

- o A chronology developed by Risks International, Inc., a private corporation engaged in terrorism analysis.
- o A data base maintained at the University of Southern California (part of which is devoted to terrorism).
- o A data base used for cross-national analysis of political assassinations developed by Thomas Snitch of the American University.

Terrorism data bases that were created to fulfill very narrow information requirements (such as a U.S. Treasury data base used to control entry and/or exit of known terrorists into or out of the United States) are not included. Some data bases are not included because very little information about them is publicly available.

The data bases discussed are not meant to represent a scientific sample of all terrorism data bases. They were selected on the basis of availability of their designers and/or documentation on their data. Nevertheless, the data and systems represent an interesting range of applications, tools, and data-collection methodologies. Table 1 presents basic information about each of the data bases. This Note examines (1) the relation between the use of terrorism data and the design of the data bases; (2) the scope and content of terrorism data bases; (3) the structure of terrorism data; and (4) the systems used for retrieval and analysis. Some conclusions will be drawn about the current status of terrorism data-base design. A description of each of the data bases discussed is given in the Appendix.

Table 1

BASIC FEATURES OF NINE TERRORIST DATA BASES

Institution, Agency, or Individual	Unit of Analysis	Number of Entries	Number of Variables	Sample	Time Period
BDM, Inc.	Incident	4700[a]	58	International/ domestic	1965-present
CIA (FITE)	Incident	3857[a]	115	International	1968-present (hijack file 1931-1977)
Edward Mickolus (ITERATE)	Incident	3329[a]	115	Transnational	1968 to present
Dept. of State (TAG)	Incident	Unknown	---	International/ domestic	---
DIA/CTTAB (STIF)	Incident	200[a]	32	International	1970-present
The Rand Corp. (Chronology of International Terrorism)	Incident	1500[a]	---	International	1968-present
Risks Int'l., Inc.	Incident	5000[a]	18	International/ domestic	1970-present
Thomas Snitch	Assassination	532	15	123 nations	1968-1978
DEMOCRATIC WORLDWATCH (DWW)	Incident	300[b]	5	International/ domestic	1978

[a] Number is approximate; the data base is updated regularly.

[b] Terrorist events only.

II. APPLICATIONS FOR TERRORISM DATA BASES

The uses to which terrorism data are put include basic research, intelligence estimates, and query responses. Basic research applications include those that deal with general issues concerning the nature of terrorist behavior as well as methodological issues in the collection and analysis of data. Two examples are the WORLDWATCH data base of Charles McClelland of USC and the Chronology of International Terrorism compiled by The Rand Corporation (hereinafter referred to Rand's Terrorism Chronology). Intelligence-estimate applications operate in a more self-consciously predictive mode. Attempts are made to draw inferences about future terrorist behavior on the basis of previous incidents. The CIA's ITERATE data base serves this application. A query-based application is oriented toward answering specific questions about particular terrorists or events in which they were involved. Risks International operates a query-driven data base.

This typology is somewhat artificial in that most data bases perform a variety of overlapping functions. (For example, the DIA Threat Analysis Group data base, from which intelligence estimates are derived, provides answers to queries from various constituents.) Nevertheless, the distinction is useful in understanding how data are used, how the information evolves into a coherent data base, and the reasons behind the adoption of specific computer-based facilities.

In one model of data-based basic research, an analyst develops theories about a subject and constructs statistically testable hypotheses which are applied to the data. While there have been rela-

tively few studies using formal hypothesis testing, examples of statistical analysis using ITERATE and Rand's Terrorism Chronology appear in Osmond (1979) and Midlarsky (1980). Rand's Terrorism Chronology is also currently being used to construct quantitative models of terrorism.

While most basic research on the causes and effects of terrorism is, in fact, statistically based, the studies have tended to be more descriptive than inferential. One reason for this is that until recently, the size of chronology-based data sets did not permit adequately reliable quantitative analysis, particularly in the case of interesting categories of terrorist incidents. The introduction to one quantitative analysis of barricade-and-hostage incidents (Jenkins, Johnson, and Ronfeldt, 1978) noted that

> It is hazardous to draw statistical inferences from so small a universe. The conclusions that may be suggested, or the predictions that seem inherent, are quite tentative. In our analysis of the [77 barricade-and-hostage] cases we have given the actual numbers and the percentages these represent . . . to give greater meaning to the imprecise words that are commonly used in place of numbers. . . . [Yet] the data do not permit a high level of confidence based upon rigorous quantitative analysis of a sufficiently large population of events.

As the universe of terrorist activities steadily increases, and as the terrorism data bases come to reflect this increase, more rigorous and comprehensive statistical analysis of data will become possible. Rand's Terrorism Chronology, the CIA's FITE data base, and others now have sufficient samples to permit more sophisticated statistical analysis. The usefulness of quantitative approaches to the analysis of terrorist behavior will increase as the amount of information about terrorism increases. As a consequence, data bases used for basic research

are implemented on statistical-analysis packages, which facilitate comprehensive and highly structured quantitative analysis of data. A version of ITERATE is available as an SPSS-based data file, and versions of Rand's Terrorism Chronology exist as several SPSS- and SCSS-based data files.[1] The DEMOCRATIC WORLDWATCH data-base system includes a number of computerized procedures for statistical analysis.

Computer-based statistical processing and data management tools also support statistic- and graphic-based analysis of data used in the development of intelligence estimates. Mickolus (1977) observed:

> An appreciation of ITERATE as a source of research is afforded by its uses in summarizing trends, comparing terrorist campaigns cross-nationally and over time, and evaluating policy prescriptions for crisis management and incident negotiations support.

In the trend-summarizing application, chronologies are used as indicator data to map trends through time. An example of regularized intelligence reporting of this kind is found in the CIA's annual report on terrorism (1977 and subsequent years). One advantage to using statistical packages in developing quantitative forecasts is the ability to use techniques borrowed from economics and other disciplines.

Most other methods of generating intelligence estimates, however, require more complex and less highly structured access to information. The intelligence analyst must derive data useful for a multitude of complex questions and therefore must have a number of methods for deriving that information. Most computer-based information systems require that

---

[1] SPSS is the Statistical Package for the Social Sciences; SCSS is a "conversational" version of SPSS.

queries be highly structured, in a way that could inhibit the ability to access other relevant information. Because there is no automated equivalent to manually browsing through files, a process useful in some analytical contexts, virtually all of the intelligence agencies discussed here maintain manual files instead of, or in addition to, automated data-base retrieval systems.

The query-response application addresses relatively specific questions whose answers can be derived from the data in a simple and predictable manner. For example, a company designing a secure limousine for its executives overseas might query a data base about the types of weapons typically used in kidnapping attempts. Computer-based retrieval systems are ideal for this application for the same reason that they are sometimes less than ideal for less-structured information-searching problems. The more predictable and highly structured the information needs of the end-user, the easier it is to achieve comprehensiveness and relevance in a data base and its retrieval system. Moreover, simple queries can be readily translated into actual search commands in a computerized system.

III. SCOPE AND CONTENT OF DATA BASES

INTRODUCTION

A multitude of decisions affect the construction of a data base. Most important are those that focus the data-collection effort on a coherent set of operational rules for the selection and coding of information. The data bases considered here will be discussed in terms of several of these decisions. In decreasing level of generality, they are (1) definitions, (2) scope, (3) data selection, (4) time frames, and (5) variables.

DEFINITIONS

There is no universally accepted definition of terrorism. The word is used to describe a number of mostly illegal activities committed by a variety of political, criminal, and even psychopathic groups and individuals. "Terrorism" is used at once to describe the activities of states in the suppression of dissidents, the behavior of revolutionary groups against governmental authority, criminal actions carried out for personal profit, and the actions of obviously mentally disordered individuals (e.g., so-called "mad bombers"). The vague, nonrigorous, and usually pejorative use of the term has been described as "promiscuous" (Jenkins, 1978). The definitional problem has two aspects: developing conceptual definitions that delineate the interests of the researcher and the general needs of the application, and translating the conceptual notions into operational definitions with which information can be

selected and formed into a coherent data base.

Several researchers in the field have published definitions of terrorism: In the introduction to the first publication of Rand's Terrorism Chronology, Jenkins and Johnson (1975) wrote that

> International terrorism can be a single incident or a campaign of violence waged outside the presently accepted rules and procedures of international diplomacy and war; it is often designed to attract worldwide attention to the existence and cause of the terrorists and to inspire fear. The actual victim of terrorist attacks and the target audience may not be the same; the victims may be totally unrelated to the struggle.

Edward Mickolus (1978), who developed the ITERATE data base, defined terrorism as

> . . . The use, or threat of use, of anxiety-inducing extranormal violence for political purposes, which such action is intended to influence the attitudes and behavior of a target group wider that that of the immediate victim.

Risks International, Inc. (1980) defined terrorism as

> . . . the threatened or actual use of force and violence to attain a political goal through fear, coercion and intimidation.

These definitions provide a starting point from which operationally oriented decisions about terrorism data collection can be made.

SCOPE

None of the data bases discussed attempt to collect data on all activities defined as terrorism. The scope of data bases is limited in a number of important ways, many of which are dictated by the organiza-

tional mission for which the data are collected, but some of which are inherent in the availability of sources of information and budgetary constraints. Moreover, researchers may need to achieve a degree of methodological rigor that imposes special consistency and coherency requirements. In general, the scope of terrorism data bases differs primarily in the treatment of the nature of the perpetrators and their targets, the significance of the activities, and the time frame of the activities.

Mickolus (1978) suggested that political terrorism can be usefully categorized as transnational, interstate, domestic, or state, on the basis of the nature of the perpetrators and the targets of their activities. Transnational terrorism is defined as acts committed by "basically autonomous non-state actors" against territory or nationals of some foreign country. Interstate terrorism includes acts by "individuals or groups controlled by the governmental authorities of a sovereign state," when directed against nationals or the territory of another state. Domestic terrorism "confines itself to the nationals and territory of one state," and state terrorism includes "terrorist actions conducted by a national government within its own borders."

The ITERATE data base concentrates on instances of transnational terrorism, following the notion found in the international relations literature that an ever-increasing amount of activity between nations is carried out by non-state actors. Perhaps in the decades to come, terrorism could affect the relations between nation-state actors as much as the day-to-day transnational activities of, for example, multinational corporations. The term "international terrorism" has come to mean

instances of both transnational and interstate terrorism. There are operational difficulties in distinguishing the instances of government support (or even more to the point, instances where a government passively encourages terrorist activities) in the case of interstate terrorism. As Milbank (1976) noted

> Given the element of governmental patronage that is common to both, the border line between transnational and international terrorism is often difficult to draw. To the degree that it can determined, the key distinction lies in who is calling the shots with respect to a given action or campaign. Hence, groups can and do drift back and forth across the line. For example, even a one-time "contract job" undertaken on behalf of a government actor by a group that normally acts according to its own lights qualifies as international terrorism.

Rand's Terrorism Chronology and the CIA's FITE data base include incidents that in some way involve two or more nations, while all of the other data bases include both international and domestic incidents but generally exclude instances of state terrorism. The scope of the data bases, in these instances, relates directly to the mission of the agency or organization collecting the data. For example, the Threat Analysis Group of the Department of State has operational responsibility for the protection of certain State Department officials while in the United States; therefore, information about some purely domestic terrorist activities would be relevant to their mission.

One major difficulty in creating a rigorous and consistent operational definition of terrorist acts is the necessity of attributing certain motives to violent acts. What distinguishes a political bombing from a bombing by, say, someone who is simply criminally insane is the motivation of the bomber. Jenkins (1975) noted that

> The borderline separating political motives from highly personal motives and purely criminal motives is not always clear. For example, on May 3, 1970, a member of the British High Commission in Uganda faked his own kidnapping to dramatize the plight of Asians in that country. . . . We have tried to exclude those incidents where the motives were clearly personal profit, such as a hijacking in which the sole demands were cash ransom and a means of escape. Where all of the details of an incident were not known, it was included if it had the earmarks of an international terrorist incident.

One data base, that of BDM, does collect information on incidents of criminal activity. It includes

> Acts of criminal or terrorist robberies or thefts of arms, scientific equipment, military stores, or toxic materials [and] specific acts by criminals involving high degrees of intelligence, planning, organization, or level of force. (BDM, 1980)

The inclusion of these incidents is related to BDM's stated objective of providing data for the analysis of different kinds of threats, physical and personal security issues, and countermeasures and responses for private businesses and government.[1] In general, however, attempts are made to separate politically oriented acts from those carried out by criminal or psychopathic elements.

DATA SELECTION

While none of the data-collection efforts attempt to gather information on all forms of terrorism, these data bases should be not considered "samples" of terrorist incidents in the statistical sense. This

---

[1] The Rand Corporation is collecting and analyzing data on "high value" and sophisticated crimes as "analogs" to potential terrorist threats. These, of course, are not included in Rand's Terrorism Chronology.

is an important distinction. Within the scope of terrorist acts defined for each data base, and within technological and human limits, the data, in principle, comprise the actual "universe" of like terrorist acts. Terrorist data bases are more like census data bases (which attempt to collect data on every person) and less like public opinion polls (which generally select a specially defined "sample" of individuals that represent a larger universe). This means that sampling methods have not been used and missing incidents have been inadvertently omitted. Detailed data-selection criteria are not usually publicly available, so precise estimates of omission rates can not be calculated.

The primary data-selection criterion is the importance of the terrorist act. All the data bases except FITE and ITERATE select only "significant" terrorist incidents. FITE and ITERATE explicitly attempt to be comprehensive, given the information resources at their disposal. The sample sizes for those data bases containing only important incidents vary considerably, which indicates that the criteria of significance must also vary considerably. Rand's criteria (the only ones available to the author) include (1) an incident's availability (which usually means that information on it appears in a public domain source), (2) some evidence of premeditation, if not actual planning (this eliminates some apparently random fire-bombings and acts akin to vandalism such as "trashing"), and (3) some indication that there is a real threat or possibility of violence and harm (which usually means there must be weapons involved). This is not to suggest that these criteria are the only valid ones for deciding the significance of a terrorist incident, or even that the application of these would lead to consistent selec-

tion, given independent data collectors (i.e., that they would necessarily lead to a high level of intercoder reliability). Selection criteria, like all other operational decisions in data-base development, must be matched to the application that the data will serve.

The usefulness of a data base depends heavily on the existence of consistent and reliable data-collection rules and procedures. The results of quantitative analyses of data items that were substantially different would be of questionable value. Absolute replicability in terrorism data collection is impossible to achieve because even the most precisely defined concepts in social science are subject to judgments when applied to empirically derived information. Operationalization of important, but necessarily judgmental concepts (such as those that imply, for example, goal-oriented behavior) is a source of bias, as are reporting errors. And the amount of systematic bias caused by judgments in the data-collection process is only partially measured by tools such as intercoder reliability tests.[2] The point is that a certain amount of systematic bias should be assumed in the collection of terrorism data. In fact, the exclusion of certain kinds of acts from a data base consciously introduces systematic bias, which may be desired. More important than intercoder reliability tests or other tools to measure "validity"--and certainly more useful from a research standpoint--are the precise definition of the sources and methods of collection and the documentation of the criteria for inclusion of information.

---

[2] To the author's knowlege, there are no published intercoder reliability figures for the selection of incidents in any of the data bases. Mickolus (1980) notes that tests have been made for the coding of variables in ITERATE and reports 99 percent reliability.

TIME FRAME

Although acts of terrorism have probably occurred since the beginnings of organized government, the present interest has stemmed from the attention focused upon such acts by the dramatic increase in mass-media coverage. As a consequence, most of the data bases have very recent time boundaries. The most significant characteristic of spectacular terrorist incidents--namely, worldwide media coverage--may well suggest that modern terrorist activities are fundamentally different from those of previous decades. Nevertheless, historical information about terrorist activities and related forms of subnational conflict could prove useful to current research and analytical interests. Unfortunately, there are no widely available computerized data bases that have comprehensive coverage of events that occurred prior to 1965.[3]

VARIABLES

Most terrorism data are collected and coded in the conventional manner. That is, an incident is selected for entry into the data base and a number of data items or variables are abstracted or coded from the available information. Table 2 shows the level of detail of the most common (and important) of these variables in selected data bases. In general, information on four variables--date, location, target, and type of incident--is mandatory to assure, with some confidence, that an accurate and relatively complete picture of the event is available.

---

[3] Previous interest in politically oriented non-state violence centered mostly on guerrilla warfare and, more recently, the concept of civil strife developed by Ted Robert Gurr. The distinctions, if any, between these concepts are beyond the purview of this Note. Data bases on these topics have not been included.

Table 2

LEVEL OF DETAIL OF IMPORTANT DATA ELEMENTS
IN SELECTED COMPUTERIZED DATA BASES

Data Element	BDM, Inc.	CIA	Rand	Risks, Int'l.	DWW
Date	Time of day	Time of day	Date	Time of day	Date
Location	City-area	City	Nation	City	Nation
Type of act	Type files	Type files	Codable	Type files	In text
Perpetrators	Group files	Group files	Codable	Group files	In text
Targets	---	---	Codable	---	---

NOTE: Elements are coded unless otherwise noted. The level of detail for a variable (such as time of day) depends on whether or not the information is available. Some data bases require the existence of certain data items for an incident to be entered in the data base. Group and/or type files indicate separate data files with group or type of act as the unit of analysis. "Codable" indicates the ability to use available computer software to code incidents.

- Date. Day, month, and year are usually mandatory, although precise days are sometimes difficult to identify. Many incidents take place over time. Time of day is sometimes included.
- Location. The nation in which the incident took place is usually mandatory. Some data bases include city and area of city (when available).
- Target. An individual, corporate, or governmental name is usually mandatory.
- Type of act/incident. The terrorist act must be classifiable according to some specific selection criteria.

The following variables are also important but are not usually mandatory, since the required information is far less likely to be available:

- o Name of perpetrator(s). For some types of events, such as bombings, this information is rarely available.
- o Descriptors of perpetrator(s). These are not typically available for many types of events.

Finally, some variables are conditional on the type of incident and are coded only for incidents of a particular kind. For example,

Bombings

- o Amount of damage (in dollars).
- o Type of explosive.

Hijackings

- o Aircraft make and model.
- o Type of weapon used.

Kidnappings

- o Name and position of victim.
- o Amount of ransom demanded.

Although no studies or reports exist on the availability of information for different types of variables, data bases gathered from public domain sources are thought to have a lot of missing data.

Text-based systems, such as Rand's Terrorism Chronology, do not code specific variables before they are entered into the data base. The information in the report is placed into a descriptive text and is then

retrieved through text-processing and retrieval tools in the data-base-management system (DBMS). In general, text-based systems attempt to make available as much information as possible, without restricting the types that are allowed. For example, a report may contain the phrase "a great many people were injured." While this statement could be stored and retrieved in a text-based system, many conventional data bases could not handle this information unless it was coded with a specific number. Text-processing and retrieval tools make it possible to retrieve all the variables indicated above.

SOME OBSERVATIONS

Virtually every researcher we interviewed expressed concern about the difficulty of defining a set of consistent criteria for inclusion of incidents in terrorism data bases. The problem is one of balancing the desire for comprehensiveness with the necessity for rigor and relevance of the data. Data bases that apply clearly defined criteria were criticized for being overly restrictive--in some cases, the restrictions are imposed by law (as in the case of FITE data); in others, they are dictated by mission (as in the State Department's collection) or by the desire to achieve methodological rigor (as in Rand's Terrorism Chronology). On the other hand, some data bases include information that might be excluded under a strict interpretation of data-base definition. This information would include data of special interest--for example, the DIA data base (by design) oversamples minor terrorist incidents for groups that are underrepresented in the data base. The data bases are clearly divided on this issue: Those used for basic research strive for more

rigorous definitions and operationalization of concepts, while those used mostly for intelligence estimates contain data that seem more directly related to policy questions, whether or not this results in consistent data selection.

IV. DATA-BASE STRUCTURE

Related to definitional problems in the creation of terrorism data bases is the selection of the basic organizational features, i.e., the unit of analysis, of the data base itself. The unit of analysis has been defined as the observation to which specific information in the data base relates. The units of analysis found in terrorism data bases include the terrorist incident, the terrorist group, the individual terrorist, and the location of the incident. Most chronologies use the incident as the unit of observation. Data bases using the group or individual terrorist as the unit of analysis contain dossier-type information. Many intelligence-oriented data bases have separate nation files organized by area of interest.

The structure of the data is defined by the relationships between entities, variables, values, and relations. An entity is any object represented in the data base, i.e., a unit of analysis, the most important of which is the terrorist incident. Variables are sets of values that characterize or define each entity. For example, terrorist entities are usually characterized by such variables as type of incident (hijacking, kidnapping, bombing, etc.), location, and target. Values are the specific characterizations of the variable. Relations are groups of entities that are related in some way. For example, a series of bombings in which a particular type of explosive device was used might be defined as a relationship.

Most data bases are structured around the operational necessities of computer systems and methods of analysis. For example, statistical

analysis and computer-based statistical packages require a matrix structure where each row of data represents all values of variables defining an entity and each column represents all values of variables across all entities. This format requires the development of a well-defined process for extracting and coding individual variables from usually unstructured textual material. The extraction and coding process has the potential disadvantage of eliminating information that may be relevant or interesting as unanticipated questions are generated by the research process. Manual data bases, on the other hand, have no requirement for the coding of variables or even the identification of entities. The disadvantages of these data bases are the lack of accessibility of information and the potential lack of methodological rigor.

There are alternatives to these extremes, however. One is the use of a generalized DBMS for the storage and retrieval of data on terrorism. While DBMSs, like simpler statistical packages, generally require the coding of variables, they have the ability to operationalize complex structures of data. These might appear as structural representations of the relationship between a number of terrorist groups, as exemplified in Heyman's (1979) paper on diffusion of terrorist activities. A more abstract example was Waterman and Jenkin's (1977) development of a heuristic analysis system, where queries and responses form the structure of the data. Two data bases presently exist under DBMSs: the FITE data base (which is implemented under RAMIS III) and the DIA chronologies that exist under DIAL. As far as we know, however, complex structures of data have not yet been implemented.

Another possibility is the development of computerized text-based retrieval systems with extended analytical and organizational capabilities. This approach has been taken in Rand's thesaurus-based Terrorism Chronology, with the goal of being able to access terrorist incidents (or other units of information) in a variety of ways. The present operationalization allows for access to and retrieval of incidents on the basis of words, phrases, sentences, or other language units as well as pre-coded variables. The definition of complex structures of entities has yet to be realized, although McClelland's WORLDWATCH data base is implemented on a system with complex retrieval and text-manipulation features.

V. DATA-BASE RETRIEVAL TOOLS

The most difficult task in implementing effective information retrieval and analysis tools is matching the needs of the application with the functional capabilities of available systems. To begin with, it is difficult to clearly define the needs of the application on an a priori basis. Not only do the needs of the data user change over time, the functional capabilities of data-base systems, to a large extent, determine the way the application approaches its data access and analysis problems. Moreover, there are both obvious and subtle differences in the needs of different approaches within each of the application types that have been defined. Given these caveats, the following is a list of functions carried out in the three types of applications defined for terrorism data:

- Identification/retrieval. Accessing individual incidents by defining the value of any number of variables given for each case.
- Linkage. Defining groups of incidents (or other units of analysis, such as terrorist groups) by defining values for each member of the group.
- Variable creation/coding. Defining variables and their values for each incident (or other entity) in the data base.
- Report generation. Formatting and outputting information retrieved and/or derived from the above functions.
- Statistical analysis. The aggregation of numeric data using specific mathematical techniques.

Table 3 presents a tentative definition of the functional requirements of the specific applications we have defined. Each function/application combination specifies either a primary or secondary need, depending on the function's importance to the application. One major difference between the applications is the way in which information is generated for use. Report generation is the primary method for intelligence estimates and query response, while statistical analysis is mostly used for basic research. There are other differences as well. Query-response applications are mostly served by identification/retrieval functions, while both basic research and intelligence estimates need effective linkage capabilities. Because of its extensive use of statistical analysis, variable creation and coding is an important requirement of basic research.

Table 3

A COMPARISON OF FUNCTIONAL ELEMENTS AND NEEDS OF SPECIFIED APPLICATIONS

	Function				
Application	Identification/ Retrieval	Linkage	Variable Creation/ Coding	Report Generation	Statistical Analysis
Basic research	Primary	Primary	Primary	Secondary	Primary
Intelligence estimates	Primary	Primary	Secondary	Primary	Secondary
Query response	Primary	Secondary	Secondary	Primary	Secondary

Attempting to relate the functional characteristics of the applications to specific data analysis and retrieval systems is an even more hazardous task. Table 4 presents a preliminary attempt to make this association. The functional emphasis of true DBMSs--DIAL, ORBIT IV, and RAMIS III--are in the identification/retrieval, linkage, and report-generation areas. The functional emphasis of SPSS/SCSS (and, by implication, other statistical and data-management packages) is on variable creation/coding and aggregate mathematical analysis.

A comment should be made on the practice (common both in terrorism research and in other applications) of adopting the data-retrieval system most readily available. The ability to select DBMSs or other types of file-management systems on the basis of a comprehensive study of the requirements of the application is an expensive luxury not often found in the real world. Such requirements analyses will prove useful, how-

Table 4

THE FUNCTIONAL EMPHASIS OF FOUR DATA-RETRIEVAL SYSTEMS USED IN RESEARCH ON TERRORISM

	Function				
System	Identification/ Retrieval	Linkage	Variable Creation/ Coding	Report Generation	Statistical Analysis
DIALS	Primary	Primary	Secondary	Primary	Secondary
ORBIT IV	Primary	Primary	Secondary	Primary	Secondary
RAMIS III	Primary	Primary	Secondary	Primary	Secondary
SPSS/SCSS	Primary	secondary	Primary	Primary	Primary

ever, even if they are ultimately exercises in clarifying notions about how the data will be organized and used.

The difficulty in developing comprehensive data-base accessing tools is only partly the result of budgetary limits. Even if unlimited resources were available for data-base development, the use of advanced information processing techniques would provide only part of the solution. Matching information processing tools and complex data-accessing requirements involves understanding the underlying structure of information, the ability to predict with some confidence the information needs of the end-user, and the capability of understanding the ongoing development of information needs and data-base content. The nature of the applications that have been defined as well as the nature of terrorism data make these tasks especially difficult.

VI. CONCLUSIONS

The current academic and policy interest in terrorism is nearly matched by efforts to develop relevant information and data bases on which to perform analyses. The discussion of several terrorism data bases that have resulted from this intense interest has brought to light some weaknesses in current data-collection efforts:

- Data-collection efforts are concentrated on incident-oriented chronologies, while a number of different related subjects and methods need to be addressed.
- The sampling frames and methods governing the data-collection process are poorly defined.
- The capabilities of computer-based information processing systems are not being exploited.
- There is a reliance on variable coding schemes that may restrict access to relevant information and the development of future applications.

Most current data-collection efforts are devoted to the development of chronologies of terrorism. Everyone seems to be collecting the same data over and over again. The chronologies do address fundamental questions about terrorism (such as whether terrorism is increasing or decreasing and by how much), but they may not contain information that is relevant to many important questions. This is not to say that efforts at developing chronologies should stop; rather, we are suggesting that the data-collection movement has matured to the point that the

development of a number of different kinds of data bases would be worthwhile.

The difficulty of establishing a widely accepted conceptual and operational definition of terrorism prevents a comprehensive comparison of the substance of currently available data bases. The current definitions have been derived to suit a number of purposes, the most important of which are the application the data are to serve and the analytical approach of the end-user. While the ultimate value of data, however defined and collected, is their ability to serve the application they were designed to serve, a comprehensive quantitative-based comparison of data bases would be informative and relevant. The goal of such a comparison would not be to judge the efficacy of one data base over another, but to develop an understanding of the effect (if any) of the quantity and substance of terrorism data bases.

The wide range in the sophistication of the application systems used to access and analyze terrorism data make generalizations about the state of the art difficult. It is clear that greater use of more sophisticated and more powerful DBMSs will facilitate increasingly complex data bases (perhaps derived from the much simpler chronologies that now exist). For example, chronologies could easily be reorganized into new data bases with completely different entity-variable-value relationships. Moreover, with increasingly powerful text-processing tools, greater flexibility could be achieved in the creation of derived variables. Together with advances in content-analysis techniques, terrorism data bases could broadly benefit from the greater use of the capabilities of computer-based information systems.

Mickolus (1977) noted the existence of ". . . a small but growing 'invisible college' attempting to apply the tools of systematic empirical inquiry to the analysis of terrorist behavior." The development of comprehensive and relevant data bases is a fundamental aspect of this effort, in terms of both supporting basic research on terrorism and providing policymakers with relevant and accurate intelligence.

Appendix

DESCRIPTIONS OF INDIVIDUAL DATA BASES

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

PERSON TO CONTACT

C.1 A. International Security Issues

DATA BASE NAME (ACRONYM):

File on International Terrorist Events (FITE)

DATA BASE DESCRIPTION

FITE is a coded chronology of international terrorist events. Terrorism is defined as the "use, or threat of use, of anxiety-inducing extranormal violence for political purposes." The data are extensively coded, including subfiles on various aspects of terrorist incidents.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Terrorist incident	3857 (approximately) (regularly updated)
TIME PERIOD	**NUMBER OF VARIABLES**
1968 to present	150 (approximately)
SAMPLE	**SOURCES**
International only	Public domain; Intelligence reports
DATA BASE SYSTEM	**APPLICATIONS**
RAMIS III	Basic research; intelligence estimates

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE	PERSON TO CONTACT
Edward Mickolus	Edward Mickolus

DATA BASE NAME (ACRONYM):

International Terrorism: Attributes of Terrorism Events (ITERATE)

DATA BASE DESCRIPTION

ITERATE is a coded chronology of "transnational terrorism." Terrorism is defined as the "use, or threat of use, of anxiety-inducing extranormal violence for political purposes." Transnational terrorism, as opposed to other forms, is not government controlled, but involves more than one nation-state. The data are extensively coded, including subfiles on various aspects of terrorist incidents.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Terrorist incident	3329 (approximately) (regularly updated)
TIME PERIOD	**NUMBER OF VARIABLES**
1968 to present	150 (approximately)
SAMPLE	**SOURCES**
Transnational only	Public domain
DATA BASE SYSTEM	**APPLICATIONS**
SPSS	Basic research

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

Risks International, Inc.

Alexandria, Virginia

PERSON TO CONTACT

Walter Burns

DATA BASE NAME (ACRONYM):

N/A

DATA BASE DESCRIPTION

This is primarily a coded terrorism chronology. Terrorism is defined as "the threatened or actual use of force and violence to attain a political goal through fear, coercion and intimidation." The data base is limited to "significant actions carried out by terrorist groups operating within the United States and overseas, excluding communist countries." The data base does not include what Risks International refers to as activities of criminal elements.

UNIT OF ANALYSIS:

Terrorist incident

NUMBER OF ENTRIES

5000 (approximately)
(regularly updated)

TIME PERIOD

1970 to present

NUMBER OF VARIABLES

18

SAMPLE

Significant incidents

SOURCES

Public domain

DATA BASE SYSTEM

Microprocessor based

APPLICATIONS

Query response

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

PERSON TO CONTACT

The BDM Corporation
7915 Jones Branch Drive
McLean, Va 22102

R. William Mengel

DATA BASE NAME (ACRONYM):

Terrorism Data Base

DATA BASE DESCRIPTION

This is a coded terrorism chronology. The purpose of the data base is to "provide extensive and systematically collected material upon which evaluations of a multitude of private business and governmental concerns might be undertaken." The data base includes: Terrorist and criminal thefts of arms, military stores, scientific equipment, or toxic materials; criminal acts requiring high degrees of planning, organization, or intelligence; armed attacks; bombings; arsons; kidnapping by terrorists or those with heavy demands; hijackings; the use of psychological terror; acts of destruction against power systems, protected facilities or other high-technology targets; acts of sabotage; acts of potential mass destruction (such as the use of toxic chemicals); and assassinations.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Terrorist incident	4700 (approximately)
TIME PERIOD	NUMBER OF VARIABLES
1965 to present	58
SAMPLE	SOURCES
Terroristic and other violent events	Public domain
DATA BASE SYSTEM	APPLICATIONS
Computer-stored	Basic research

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

Department of State
Threat Analysis Group

PERSON TO CONTACT

Bowman Miller

DATA BASE NAME (ACRONYM):

N/A

DATA BASE DESCRIPTION

The Department of State maintains a multisubject manual data base. The Office of Security Command Center, which has responsibility for the data base, plans computerization in the near future. The data base contains a chronology of important terrorist incidents in each country, personality profiles of terrorists who might be a threat to U.S. personnel overseas, and information on terrorist groups. The data base also includes country threat assessments.

UNIT OF ANALYSIS:

Terrorist incident;groups; individuals; countries

NUMBER OF ENTRIES

Unknown
(regularly updated)

TIME PERIOD

N/A

NUMBER OF VARIABLES

N/A

SAMPLE

International and domestic

SOURCES

Public domain; intelligence sources; diplomatic cable traffic

DATA BASE SYSTEM

N/A

APPLICATIONS

Intelligence estimates; query response

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

PERSON TO CONTACT

D.I.A. Counter Terrorism
and Threat Analysis Branch of
Counter Intelligence Division

DATA BASE NAME (ACRONYM):

Significant Terrorist Incident Files (STIF)

DATA BASE DESCRIPTION

D.I.A. maintains a computerized chronology of terrorist incidents as well as manually stored files of data on groups and individuals. The computerized data base includes textual material of general interest as well as the coded data. An incident must be of international significance and must have created some problem relevant to the agency's mission. The data base may include other incidents created by groups that would not otherwise appear in the chronology.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Terrorist incident	200 (approximately) (regularly updated)
TIME PERIOD	NUMBER OF VARIABLES
1970 to present	32
SAMPLE	SOURCES
International only	Public domain; intelligence
DATA BASE SYSTEM	APPLICATIONS
DIAL	Intelligence estimates; query response

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

Thomas Snitch
The American University
Washington D.C., 20016

PERSON TO CONTACT

Thomas Snitch

DATA BASE NAME (ACRONYM):

N/A

DATA BASE DESCRIPTION

This is a coded data base containing records of assassinations and attempts between 1968 and 1978. The data were derived from an original study reported by the National Commission on Violence and subsequently added to. The data base includes data about 123 nations with a population of 1 million or more in 1975. Data is derived from public domain sources.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Assassination incident	532
TIME PERIOD	NUMBER OF VARIABLES
1968 to 1978	15
SAMPLE	SOURCES
All	Public domain
DATA BASE SYSTEM	APPLICATIONS
SPSS	Basic research

INSTITUTION, AGENCY,OR INDIVIDUAL RESPONSIBLE

PERSON TO CONTACT

The Rand Corporation
1700 Main Street
Santa Monica, Ca 90406

Brian Jenkins

DATA BASE NAME (ACRONYM):

Chronology of International Terrorism

DATA BASE DESCRIPTION

This data base is an uncoded textual chronology of significant terrorist incidents. Terrorism is defined as "violence waged outside presently accepted rules and procedures of international diplomacy and war. . . . designed to attract worldwide attention to the terrorist and to inspire fear." Incidents are included if information is publicly available, there is a real threat of violence or harm, and there is apparent premeditation to the act. The data base is computerized, computerized, and incidents are retrievable by word and sentence searching. Versions also exist as files of statistical packages.

UNIT OF ANALYSIS:	NUMBER OF ENTRIES
Incident	1166 (approximately) (regularly updated)
TIME PERIOD	NUMBER OF VARIABLES
1968 to present	4 plus textual descriptions
SAMPLE	SOURCES
International	Public domain
DATA BASE SYSTEM	APPLICATIONS
ORBIT IV, SPSS/SCSS	Basic research

BIBLIOGRAPHY

The BDM Corporation, The EDM Corporation's Terrorism Data Base, McLean, Virginia, 1980.

Engles, R. W., "A Tutorial on Data Base Organization," Annual Review in Automatic Programming, Vol. 7, Part 1, Pergamon Press, July 1972.

Fowler, W. W., The Semantics of International Relations Data: Mapping the Structure of Complex Social Systems, Ph.D. dissertation, University of Southern California, 1979.

-----, and H. E. Purkitt, "Temporal Trends in International Terrorism 1968 to 1979," Paper delivered to the Annual Convention of the International Studies Association, Los Angeles, March, 1980.

Heyman, E., Monitoring the Diffusion of Transnational Terrorism: A Conceptual Framework and Methodology, Technical Notes, Clandestine Tactics and Technology, International Association of Chiefs of Police, 1979.

Jenkins, Brian M., International Terrorism: Trends and Potentialities, The Rand Corporation, P-6117, 1978.

-----, and Janera Johnson, International Terrorism: A Chronology, 1968-1974, The Rand Corporation, R-1597-DOS/ARPA, 1975.

-----, Janera Johnson, and David Ronfeldt, "Numbered Lives: Some Statistical Observations from 77 International Hostage Episodes," Conflict: An International Journal, Vol. I, No. 1, 1978, pp. 71-111.

Mickolus, E. F., "Statistical Approaches to the Study of Terrorism," in Y. Alexander and S. M. Finger (eds.), Terrorism: Interdisciplinary Perspectives, John Jay Press, New York, 1977.

-----, "An Events Data Base for Analysis of Transnational Terrorism," in R. J. Heuer (ed)., Quantitative Approaches to Political Terrorism, Westview Press, Boulder, 1978.

-----, "Studying Terrorist Incidents: Issues in Conceptualization and Data Acquisition," Paper delivered to the Annual Convention of the International Studies Association, Los Angeles, March 1980.

-----, Transnational Terrorism: A Chronology of Events, 1968-1979, Greenwood Press, Westport, 1980.

Midlarsky, M. I., M. Crenshaw, and Y. Fumihiko, "Why Violence Spreads: The Contagion of International Terrorism," International Studies Quarterly, Vol. 24, No. 2, June 1980, pp. 262-311.

Milbank, D. L., International and Transnational Terrorism: Diagnosis and Prognisis, Central Intelligence Agency, Washington, D.C., 1976.

Osmond, L. O., Transnational Terrorism 1968-1974: A Quantitative Analysis, Ph.D. dissertation, Department of Political Science, Syracuse University, 1979.

Waterman, D. A., and Brian M. Jenkins, Heuristic Modeling Using Rule-Based Computer Systems, The Rand Corporation, P-5811, 1977.