BUILDING THE UNITED STATES EXTREMIST CRIME DATABASE (ECDB): LESSONS LEARNED

Joshua D. Freilich is the Director of the Criminal Justice Ph.D. program and a Professor in the Criminal Justice Department at John Jay College, the City University of New York. He is a lead investigator and a member of the Executive Board for the National Consortium for the Study of Terrorism and Responses to Terrorism (START), a Center of Excellence of the U.S. Department of Homeland Security. Dr Freilich is also a member of the Global Terrorism Database's (GTD) advisory board. He and Professor Steven Chermak have worked extensively on developing the Extremist Crime Database — the first of its kind National Database on criminal activities involving U.S. far right, far left, and supporters of Al-Qaeda, Hamas, Hezbollah and similar extremist movements. Dr Freilich's research focuses on the criminal activities of US extremists as well as criminological theory, especially neo-classical approaches and crime prevention. Recent projects have examined how lone wolf terrorists are similar and different from group-affiliated terrorists, county-level predictors of far right extremism and homicides in the United States, and how terrorist organizations use the Internet. He is the co-editor of a book that will be published in 2013 on transnational terrorism.

Steven Chermak is a Professor of Criminal Justice at Michigan State University and a lead investigator for the National Consortium for the Study of Terrorism and Responses to Terrorism (START), a Center of Excellence of the U.S. Department of Homeland Security. Dr Chermak's research includes activities in the following areas: far-right extremism, the effectiveness of strategies used to prevent terrorism and crime, and the media's role in relation to crime and terrorism issues. He and Professor Joshua Freilich have collaborated to create the Extremist Crime Database — the first of its kind National Database on criminal activities involving U.S. far right, far left, and supporters of Al-Qaeda, Hamas, Hezbollah and similar extremist movements. Specifically, the database includes data on the violent and financial crimes of these extremists, characteristics of violent groups, and the nature of foiled plots. Other research has focused on terrorism and media coverage of terrorist activities, including depictions of the militia movement and the September 11th attacks. Current research includes research on the characteristics of lone wolf terrorism, differences between violent and nonviolent extremist groups, and county-level predicators of far right violence.

ABSTRACT

This paper accomplishes two objectives. First, it provides an overview of the process used to create the United States Extremist Crime Database, the first of its kind national database to track violent and financial crimes committed by domestic extremists. Second, it discusses several critical lessons that have been learned in the process of creating the ECDB.

RÉSUMÉ

Cet article poursuit deux objectifs. En premier lieu, il offre un aperçu du procédé utilisé lors de la création d'une base de données américaine sur les crimes liés à l'extrémisme (United States Extremist Crime Database, ou ECDB), la première base de données en son genre au États-Unis qui trace le bilan des crimes violents et des crimes financiers qui sont commis par des extrémistes à l'intérieur du pays. En second lieu, nous discutons de plusieurs leçons cruciales qui ont étés apprises au cours de la création de la ECDB.

This essay explains how we created the United States Extremist Crime Database (ECDB). We outline nine important lessons we have learned from this process that could aid the development of a Canadian Extremist Crime Database.

INCLUSION CRITERIA

For a crime to be included in the ECDB it must satisfy a two-pronged test. The first prong is behavioral, and requires that an illegal violent act or financial scheme be committed inside the United States. Violent incidents include homicide events, incidents where extremists are killed by police, and in some cases arsons, bombings, attempted homicides or foiled/failed plots. Financial crimes are often committed in the context of larger criminal operations involving multiple perpetrators and jurisdictions over an extended period of time. Sometimes it is difficult to categorize financial crimes as distinct incidents because of the ambiguity that exists with spatial and temporal distinctions. To capture the nuances in financial crime cases, we developed the concept of "financial scheme" (Belli 2011). The financial scheme is defined as an illicit financial operation involving a set of activities (i.e. techniques) carried out by one or more perpetrators to obtain unlawful gain or other economic advantage through the use of deliberate deception.

The second prong is attitudinal, and requires that, at the time of the incident or scheme, at least one of the suspects who committed this act subscribed to an extremist belief system such as the far-right, eco/animal rights extremism, or Al-Qaeda affiliated/inspired ideologies. We conducted extensive literature reviews to craft descriptions of these extremist ideologies and our definitions are available upon request.

IDENTIFYING CRIMES[1]

The ECDB was developed in stages. The first stage identified relevant crimes committed by supporters of these extremist movements in the United States from a variety of source types. In fact, we identified and then reviewed over 50 different sources. These sources include specific volumes, databases, and general search strategies. Here we provide a sample of some of these sources. For example, we reviewed official sources such as the FBI's *Terrorism in the United States* and the National Counterterrorism Center's Worldwide Incidents Tracking System (WITS) database to identify violent crimes. Financial crimes were documented from various Department of Justice agencies that issue press releases and provide links to indictments, and convictions concerning financial crimes. We also looked at private watch-groups such as the *Southern Poverty Law Center* and the *Anti-Defamation League*.

Existing terrorism databases such as the American Terrorism Study and the Global Terrorism Database were examined. Some sources like the RAND-MIPT database include rich data on indictments and other court proceeding documents. Relevant incidents were extracted for the ECDB. Scholarly and journalist accounts were also reviewed. Finally, media publications provide important open source materials and we conducted systematic searches for additional incidents in a variety of general newspapers and locally archived newspaper databases. Many of the watch-group and media sources also track the non-violent financial crimes committed by these extremists.

GATHERING OPEN SOURCE INFORMATION ON IDENTIFIED CRIMES

In the second step, each violent criminal incident, financial criminal scheme, and related perpetrators and victims were systematically searched in 26 web-engines and existing terrorism databases, official sources, and watch-group reports to uncover as much relevant open source information as possible. (A listing of these web-engines is available upon request).

Searchers used key information about the crime, suspect names, victim names, and names of the organizations or business entities linked to the schemes/incidents to conduct online searches. To insure that the searches are thorough, searchers use different spellings of suspects and victims' names, and various permutations. The searchers also systematically search by location to identify court documents. To capture the most relevant media accounts, the searchers focus on obtaining information from national outlets as well as the newspapers specific to the region where the event or scheme occurred or where the suspect or victims resided.

These searches uncover all published open source materials on each case, such as media accounts, government documents, court records, indictments, appellate court decisions, videos, blogs, books, watch-group reports, extremist movement produced materials, and scholarly accounts. This information is digitally archived and searchers organize it by source type starting with the most reliable.

CODING CRIMES

In the final stage, the open source information is provided to a research assistant coder. These coders search their assigned violent and financial cases to verify that the original searches are complete (they also note any other incident/scheme mentioned in the materials, compare that list to the master file, and add any missing cases). If

the original search materials are incomplete, the coder conducts targeted follow-up searches (e.g., searching specific names, group names, etc.) to fill in missing values.

The ECDB is relational and it collects data on multiple units of analysis by gathering information on the incidents/schemes, perpetrators, victims and target, the social ties between and among the suspects and victims, and the financial crimes, the business entities linked to the schemes. The coders use the open source materials to code variables found in these forms that are connected to an online database.[2]

RELIABILITY ISSUES

Scholars have raised concerns about the conclusions reached relying solely on open sources. These concerns include possible inconsistencies and gaps in available information, inaccuracies, and bias in some sources of information (LaFree 2010; Sageman 2004). It is surprising that there are few methodological pieces that evaluate the use of open source methodologies in studying terrorism. The ECDB has attempted to address some of these concerns by assessing reliability.

Again, our uncovered search materials contained documents from different source types that occasionally contained conflicting information. These discrepancies implicated reliability issues related to source type. In these situations greater weight is granted to the more "trusted" source. Similar to Sageman (2004) "in decreasing degrees of reliability… [we favour] court proceedings subject to cross examination, followed by reports of court proceedings, then corroborated information from people with direct access to information provided, uncorroborated statements from people with that access, and finally statements from people who had heard the information secondhand." Table 1 lists the source types in decreasing degrees of reliability.

Table 1: Ranking of source reliability

1. Appellate court proceedings
2. Court proceedings subject to cross examination (e.g., trial transcripts)
3. Court proceedings or documents not subject to cross examination (e.g., indictments)
4. Corroborated information from people with direct access to information provided (e.g., law enforcement and other key informants)
5. Uncorroborated statements from people with that access
6. Media reports
7. Watch-group reports
8. Personal views expressed in blogs, websites, editorials or Op-Ed, etc

Because the ECDB uses multiple coders, we addressed inter-rater (i.e., coder) reliability. Importantly though, unlike projects that are static and collect data at one point in time, the ECDB and other terrorism databases are large-scale ongoing efforts. The ECDB updates values as new information becomes available. Such a process requires substantial resources of money, time, and efforts to keep the databases current. It is difficult to engage in standard inter-rater reliability practices. Nevertheless, we address this important issue in a number of ways. First, coders are trained. New coders initially code previously coded cases and both sets of values are compared. We created a listserv of ECDB personnel and instruct coders to share difficult issues. In this way, inconsistencies are addressed early in the coding process. Second, coding abnormalities are continually checked across coders. Third, open source coding occurs in stages, which increases the chances that all available information from open sources is captured. Conducting targeted searches based on information uncovered during the initial search presents another opportunity for coders to recheck past work of fellow coders. Fourth, filling in values for certain ECDB variables requires little interpretation as the variables capture basic facts such as a perpetrator's race, age, or gender.

Fifth, and most significantly, we have begun validating our violent incidents and financial schemes by verifying that coders systematically applied the coding rules when creating relational records for perpetrators, victims, targets, and their networks. Where coding inconsistencies occurred, records are being updated and corrected so that coding procedures will be uniform across all research assistants and incidents. Similarly, we are verifying that each incident and scheme has the correct number of perpetrators and correct files in Access. We are fixing incorrectly coded IDs, and missing relational connections between codebooks.

NINE LESSONS LEARNED

Our experience in building the ECDB has identified a number of important lessons that are useful for endeavors to build similar databases:

(1) We conducted an initial measurement of inter-rater reliability for selected individual and situational characteristics of far-right homicides and found coder agreement between 89% and 98% of the time. When coders disagreed it was usually not because of differences in the values coded, but because one coder found a document that contained information that could be coded, while the second coder did not find it. *It is thus important to have multiple coders both search and code each incident when using open-source materials. Training, open discussion of discrepancies and updating cases is critical.*

(2) It is important to use resources to fill in missing values. Once data is preliminarily ready for analysis, time must be invested for additional cleaning and using resources to fill in missing data. For example, across several projects, we have been able to search additional databases (e.g., state, local, and federal inmate locators; online local court dockets; FBI's Supplemental Homicide Report; the social security death index; online national record aggregators such as Ancestry.com and Archives.com; and news aggregates), and were able to fill in 99-100% of certain variables' values (e.g., perpetrator/victim race, sex, age, their relationship; weapon used). *Thus, studies that use "subject matter experts" for key variables to conduct intensive targeted searches of selected data sources should be able to fill in the overwhelming number of values for these variables. Such efforts take time but the advantages are significant.*

(3) We examined selectivity bias. We looked at 10 sources (such as the FBI, the Anti-Defamation League, etc) that the ECDB used to identify far-right homicides (Chermak, Freilich, Parkin & Lynch, 2012). After examining these sources' similarities and differences, we normalized their criteria to accurately assess variations in the events they included. We used a "catchment-re-catchment" analysis and found that the inclusion of additional sources resulted in an increasing number of events that were identified in previous sources. Collectively the sources appeared to be approaching capturing the universe of eligible events. *Thus, using multiple sources- and ideally all relevant sources- to identify the cases you are interested in should minimize the danger of selectivity bias.*

(4) The ECDB does not limit itself to acts labeled terrorist by the FBI and prosecuted on the federal level. Most American terrorism databases and definitions, such as the ones used bythe FBI, require terrorist acts to use "force or violence" and exclude non-violent financial crimes. This is an important omission because the ECDB has identified over 700 financial schemes that were committed by far-rightists and Al-Qaeda, Hamas, Hezbollah, and similar extremists in the U.S. The total financial loss incurred by these schemes is over $650,000,000. Far-rightists most commonly committed tax avoidance crimes, while Islamic extremists committed money-laundering and provided material support to terrorists.

Similarly, over 35% of far-right and 48% of Al-Qaeda supporters' ideologically motivated homicides were committed by lone actors. Over 80% of tried far-right suspects in the ECDB who committed ideological homicide incidents were prosecuted on the state-level. Unlike the ECDB, the FBI and many American domestic terrorism studies exclude violent lone actor attacks and incidents prosecuted on the state-leve1. *Thus, data collection should not be limited to "terrorism" as defined by the FBI or other government bodies. Including financial crimes will allow scholars and law enforcement to investigate crimes that cost society hundreds of millions of dollars, and significantly also could be related to violent terrorist activity. Similarly, many claim that attacks by lone wolves are difficult to prevent and pose a greater challenge than attacks committed by organized groups. The first step to crafting effective policies is to provide law enforcement officials with the best knowledge and practices available from empirical data that collects this information.*

(5) The ECDB has identified over 370 homicide incidents committed by far-rightists in the U.S. since 1990. Over 150 of these homicides were ideologically motivated. During this same period, the ECDB has identified 30 homicides, 12 attempted homicides and 66 ideologically motivated foiled plots committed by Al-Qaeda supporters. Our preliminary analysis has found that these far-right and Al-Qaeda attacks differ in their spatial and temporal variation (i.e., they occur in different counties and states and in different years) and the characteristics (e.g., race) of the perpetrators who commit these crimes differ. *Thus, it is important to disaggregate terrorist and extremist criminal acts to uncover different patterns that might exist across ideological groupings. The comparative analysis provides opportunities for devising intervention strategies. In addition, our inclusion of foiled plots provides insights into effective strategies that prevent terrorist acts.*

(6) The ECDB tracks ideologically motivated and non-ideological crimes committed by extremists. Most terrorism databases exclude crimes committed for non-ideological reasons. Only around 40% of fatal far-right strikes were ideologically motivated and 20% were non-ideological, but were related to the extremist movement. These non-ideological but movement-related homicides include incidents involving internal organizational disputes (e.g., killing an informer or fatal attacks over drugs or women). Another approximate 40% of attacks were not ideologically motivated, but were committed for personal motivations such as greed. *Thus, focusing only ideologically motivated crimes misses important information.*

(7) Most existing terrorism databases use rules that label an incident or perpetrator terrorist and include it or label it non-terrorist and exclude it. The reality, however, may be more complex. In response, the ECDB created strength of certainty variables for perpetrators (based upon the open source information, how certain are we the suspect adheres to an extremist ideology) and incidents (based upon open sources, how certain are we the act is ideologically motivated). Both variables are coded on a scale from 0-4 (0=non-ideological; 4=undisputed evidence of ideology). This scale captures

if the perpetrators committing ideologically motivated crimes exhibit different levels of commitment to their ideology, and if ideologically motivated acts exhibit different levels of motivation for their etiology. A preliminary examination indicates that nearly 10% of the suspects committing these ideologically motivated homicides were not extremists and close to 20% only received a "1" or a "2" regarding our certainty of their association to the movement. ECDB data thus undermine the traditional distinction between political extremism and non-ideological offenders. These results may support the convergence thesis that methods and motives driving political extremists and opportunistic offenders sometimes coincide. *Thus, focusing only on binary measures (extremist versus non-extremist) could miss important nuance that could be useful to both policymakers and scholars.*

(8) Most terrorism databases focus on the event-level. A few focus on the perpetrator level, but they usually only collect demographic or network information. The ECDB, on the other hand, collects information on target and victim characteristics and includes individuals who were injured, killed, or targeted. We have found that far-right homicide attacks claimed well over 600 lives (over 435 excluding victims of the Oklahoma City bombing). Over 9% of these victims were representatives of law enforce-ment, correctional officers, or private security guards killed in the line of duty. Interestingly, among the universe of all homicide victims, law enforcement victims usually account for less than 1% of this total in a given year. *Thus, ignoring victims and assuming they are randomly selected, and a representative sample of the population with no risk patterns to uncover misses important information.*

(9) Again, most terrorism databases collect information on one unit of analysis (e.g., event or perpetrator) and are flat files. The ECDB is relational and collects data on incidents/schemes, perpetrators, victims and target, the social ties between and among the suspects and victims, and, for financial crimes, the business entities linked to the schemes. *A relational database allows for analysis across the variables found in the various codebooks. For example, a relational database allows a researcher interested in the characteristics of all Hamas perpetrators involved in a specific type of scheme (e.g., Ponzi) to merge variables from these two different codebooks to create a new distinct dataset. A non-relational database does not have the capabilities to answer such questions.*

NOTES

[1] This section and the ones that follow that describe how we built the ECDB draw heavily from Freilich, Chermak, Belli, Gruenewald & Parkin's (in press) piece.

[2] While the open source search files are primarily used by our coders to input values for the variables in our codebooks, they also can be used for qualitative research such as case studies and discourse analysis, etc (see for e.g., Freilich and Chermak, 2009; Freilich, Chermak and Caspi, 2009).

REFERENCES

Belli, R. [2011] *Where political extremists and greedy criminals meet: A comparative study of financial crimes and criminal networks in the United States.* Graduate Center/John Jay College; CUNY.

Chermak, S.M., Freilich, J.D., Parkin, W. & Lynch, J.P. [2009] American terrorism and extremist crime data sources and selectivity bias: An investigation focusing on homicide events committed by far-right extremists. *Journal of Quantitative Criminology,* 28(1): 191-218.

Freilich, J.D. & Chermak, S.M. [2009] Preventing deadly encounters between law enforcement and American far-rightists. *Crime Prevention Studies,* 25: 141-172.

Freilich, J.D., Chermak, S.M. & Caspi, D. [2009] Critical events in the life trajectories of domestic extremist white supremacist groups: A case study analysis of four violent organizations. *Criminology and Public Policy,* 8(3): 497-530.

Freilich, J.D., Chermak, S.M., Belli, R., Gruenewald, J. & Parkin, W.S. Introducing the United States Extremist Crime Database (ECDB). *Terrorism and Political Violence.* Forthcoming.

LaFree, G. [2010] The Global Terrorism Database: Accomplishments and challenges. *Perspectives on Terrorism,* 4: 24-46.

Sageman, M. [2004] *Understanding terror networks.* Philadelphia: *University of Pennsylvania Press.*