

The National Academies

Title:

Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment (2008)

Authors:

Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals, National Research Council

Executive Summary

In a democratic society it is vitally important that citizens and their representatives be able to make an informed judgment on how to appropriately balance privacy with security. This report seeks to contribute to that informed judgment.

September 11, 2001, provided vivid proof to Americans of the damage that a determined, fanatical terrorist group can inflict on our society. Based on the available information about groups like Al Qaeda, most importantly their own statements, it seems clear that they will continue to try to attack us. Further attacks by such groups, and indeed by domestic terrorists like Timothy McVeigh, could be as serious as, or even more serious than, September 11 and Oklahoma City. Because future terrorist attacks on the United States could cause major casualties as well as severe economic and social disruption, the danger they pose is real, and it is serious. Thus, high priority should be given to developing programs to detect intended attacks before they occur so that there is a chance of preventing them.

At the same time, the nation must ensure that its institutions, information systems, and laws together constitute a trustworthy and accountable system that protects U.S. citizens' rights to privacy.

In this report, the Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals examines the role of data mining and behavioral surveillance technologies in

counterterrorism programs,¹ and it provides a framework for making decisions about deploying and evaluating those and other information-based programs on the basis of their effectiveness and associated risks to personal privacy.

The most serious threat today comes from terrorist groups that are international in scope. These groups make use of the Internet to recruit, train, and plan operations, and they use public channels to communicate. Therefore, intercepting and analyzing these information streams might provide important clues regarding the nature of the terrorist threat. Important clues might also be found in commercial and government databases that record a wide range of information about individuals, organizations, and their transactions, movements, and behavior. But success in such efforts will be extremely difficult to achieve because:

- The information sought by analysts must be filtered out of the huge quantity of data available (the needle in the haystack problem); and
- Terrorist groups will make calculated efforts to conceal their identity and mask their behaviors, and will use various strategies such as encryption, code words, and multiple identities to obfuscate the data they are generating and exchanging.

Modern data collection and analysis techniques have had remarkable success in solving information-related problems in the commercial sector; for example, they have been successfully applied to detect consumer fraud. But such highly automated tools and techniques cannot be easily applied to the much more difficult problem of detecting and preempting a terrorist attack, and success in doing so may not be possible at all. Success, if it is indeed achievable, will require a determined research and development effort focused on this particular problem.

Detecting indications of ongoing terrorist activity in vast amounts of communications, transactions, and behavioral records will require technology-based counterterrorism tools. But even in well-managed programs such tools are likely to return significant rates of false positives, especially if the tools are highly automated. Because the data being analyzed are primarily about ordinary, law-abiding citizens and businesses, false positives can result in invasion of their privacy. Such intrusions raise valid concerns about the misuse and abuse of data, about the accuracy

¹ In this report, the term “program” refers to the system of technical, human, and organizational resources and activities required to execute a specific function. Humans—not computers—are always fully responsible for the actions of a program.

of data and the manner in which the data are aggregated, and about the possibility that the government could, through its collection and analysis of data, inappropriately influence individuals’ conduct. Intruding on privacy also risks ignoring constitutional concerns about general search, as reflected in the Fourth Amendment. The committee strongly believes that such intrusion must be minimized through good management and good design, even if it cannot be totally eliminated.

The difficulty of detecting the activity of terrorist groups through their communications, transactions, and behaviors is hugely complicated by the ubiquity and enormity of electronic databases maintained by both government agencies and private-sector corporations. Retained data and communication streams concern financial transactions, medical records, travel, communications, legal proceedings, consumer preferences, Web searches, and, increasingly, behavioral and biological information. This is the essence of the information age—it provides us with convenience, choice, efficiency, knowledge, and entertainment; it supports education,

health care, safety, and scientific discovery. Everyone leaves personal digital tracks in these systems whenever he or she makes a purchase, takes a trip, uses a bank account, makes a phone call, walks past a security camera, obtains a prescription, sends or receives a package, files income tax forms, applies for a loan, e-mails a friend, sends a fax, rents a video, or engages in just about any other activity. The proliferation of security cameras and means of tagging and tracking people and objects increases the scope and nature of available data. Law-abiding citizens leave extensive digital tracks, and so do criminals and terrorists.

Gathering and analyzing electronic, behavioral, biological, and other information can play major roles in the prevention, detection, and mitigation of terrorist attacks, just as they do against other criminal threats. In fact the U.S. government has increased its investment in counterterrorism programs based on communications surveillance, data mining, and information fusion. Counterterrorism agencies are particularly interested in merging several different databases (information fusion) and then probing the combined data to understand transactions and interactions of specific persons or organizations of interest (data mining). They would also like to identify individuals (through data mining and behavioral surveillance) whose transactions and behavior might indicate possible terrorist links.

Such techniques often work well in commercial settings, for example for fraud detection, where they are applied to highly structured databases and are honed through constant use and learning. But the problems confronting counterterrorism analysts are vastly more difficult. Automated identification of terrorists through data mining (or any other known methodology) is neither feasible as an objective nor desirable as a goal of technology development efforts.

One reason is that collecting and examining information to inhibit terrorists inevitably conflicts with efforts to protect individual privacy. And when privacy is breached, the damage is real. The degree to which privacy is compromised is fundamentally related to the sciences of database technology and statistics as well as to policy and process. For example, there is no way to make personal information in databases fully anonymous. Technical, operational, legal, policy, and oversight processes to minimize privacy intrusion and the damage it causes must be established and uniformly applied. Even under the pressure of threats as serious as terrorism, the privacy rights and civil liberties that are the cherished core values of our nation must not be destroyed.

The quality of the data used in the difficult task of preempting terrorism is also a substantial issue. Data of high quality are correct, current, complete, and relevant, and so they can be used effectively, economically, and rapidly to inform and evaluate decisions. Data derived by linking high-quality data with data of lesser quality will tend to be low-quality data. Because data of questionable quality are likely to be the norm in counterterrorism, analysts must be cognizant of their effects, especially in fused or linked databases, and officials must carefully consider the consequent likelihood of false positives and privacy intrusions.

The preliminary nature of the scientific evidence, the risk of false positives, and operational vulnerability to countermeasures argue for behavioral observation and physiological monitoring being used at most as a preliminary screening method for identifying individuals who merit additional follow-up investigation. Although laboratory research and development of techniques for automated, remote detection and assessment of anomalous behavior, for example deceptive behavior, may be justified, there is not a consensus within the relevant scientific community nor on the committee regarding whether any behavioral surveillance or physiological monitoring

techniques are ready for use at all in the counterterrorist context given the present state of the science.

The committee has developed and provides in [Chapter 2](#) a specific framework for evaluation and operation of information-based counterterrorism programs to guide deployment decisions and facilitate continual improvement of the programs.

National security authorities of course should always adhere to the law, but the committee recognizes that laws will have to be reviewed and revised from time to time to ensure that they are appropriate, up to date, and responsive to real needs and contemporary technologies.

With these several concerns and issues in mind, the committee makes the following recommendations.

Recommendation 1. U.S. government agencies should be required to follow a systematic process (such as the one described in the framework proposed in [Chapter 2](#)) to evaluate the effectiveness, lawfulness, and consistency with U.S. values of every information-based program, whether classified or unclassified, for detecting and countering terrorists before it can be deployed, and periodically thereafter. Under most circumstances, this evaluation should be required as a condition for deployment of information-based counterterrorism programs, but periodic evaluation and continual improvement should *always* be required when such programs are in use. The committee believes that the framework presented in [Chapter 2](#) defines an appropriate process for this purpose.

Periodically after a program has been operationally deployed, and in particular before a program enters a new phase in its life cycle, policy makers should apply a framework such as the one proposed in [Chapter 2](#) to the program before allowing it to continue operations or to proceed to the next phase. Consistency with relevant laws and regulations, and impact on individual privacy and civil liberties—as well as validity, effectiveness, and technical performance—should be rigorously assessed. Such review is especially necessary given that the committee found little evidence of any effective evaluation performed for current programs intended to detect terrorist activity by automated analysis of databases. (If such evidence does exist, it should be presented in the appropriate oversight forums as part of such review.) Periodic review may result in significant modification of a program or even its cancellation.

Any information-based counterterrorism program of the U.S. government should be subjected to robust, independent oversight. All three branches of government have important roles to play to ensure that such programs adhere to relevant laws. **All such programs should provide meaningful redress to any individuals inappropriately harmed by their operation.**

To protect the privacy of innocent people, the research and development of any information-based counterterrorism program should be conducted with synthetic population data. If and when a program meets the criteria for deployment in the committee's illustrative framework described in [Chapter 2](#), it should be deployed only in a carefully phased manner, e.g., being field tested and evaluated at a modest number of sites before being scaled up for general use. At all stages of a phased deployment, data about individuals should be rigorously subjected to the full safeguards of the framework.

Recommendation 2. The U.S. government should periodically review the nation's laws, policies, and procedures that protect individuals' private information for relevance and effectiveness in light of changing technologies and circumstances. In particular, Congress should reexamine existing law to consider how privacy should be protected in the context of information-based programs (e.g., data mining) for counterterrorism. Such reviews should consider establishment of restrictions on how personal information can be used. Currently, legal restrictions are focused primarily on how records are collected and assessed, rather than on their use.