Working Group II: Building an Effective, Sustainable Partnership Between the Government and the Private Sector

Co-chairs of Working Group II are Gilman Louie and James Steinberg. Members are Zoë Baird, Stewart Baker, Jim Barksdale, Jerry Berman, Wesley Clark, James Dempsey, Esther Dyson, Amitai Etzioni, David Farber, John Gage, Margaret Hamburg, John Hamre, Danny Hillis, Jeff Jonas, Arnold Kanter, James Lewis, Jeffrey Smith, Abraham Sofaer, Paul Schott Stevens, Michael Vatis, Philip Zelikow, and Jim Zimbardi. This paper was written by James Steinberg.

Introduction

The challenge of preventing and responding to the new security threats is very different from the one we, as a nation, faced in the Cold War. Today, the private sector is on the frontline of the homeland security effort: Its members are holders of information that may prove crucial to thwarting terrorist attacks; stewards of critical infrastructure that must be protected and dangerous materials that could be used to do harm; and important actors in responding to attacks. As we said in our first Task Force report, private sector information is essential to counterterrorism, and government agencies should have timely, needed access to that information, pursuant to guidelines that give confidence that the information will be used in a responsible way.

Government agencies already have access to certain kinds of privately held information. However, the rules governing access to it have evolved haphazardly and are confusing and sometimes contradictory. Moreover, the rules and practices fail to take into account the dramatic evolution of information technologies that can substantially increase the value of such data in helping to prevent acts of terror. The time has come for a fresh look at how the government can make the most effective use of the information that it truly needs to meet emerging security challenges.

At the same time, if our government is to sustain public support for its efforts, it must demonstrate that the information it seeks to acquire is genuinely important to the security mission, and that it is obtained and used in a way that minimizes any negative impact on privacy and civil liberties. Current privacy protection laws and procedures are not in synch with the challenges and possibilities that rapidly advancing technologies are bringing; there are few reliable processes to ensure that information is accurate and up-to-date; and some of the proposed information-related programs seem to offer little added value and may impose substantial costs on industry. Plus, there are inadequate mechanisms of oversight and accountability to prevent unauthorized access to, and use of, information. The reason we seek to strengthen our homeland security effort is to protect our safety and our way of life. Therefore, our approach must give the public confidence that the information collected by the government has significant value in relation to the potential negative impact on civil liberties and other important interests.

In our initial report, we stated, "The government will need access to public and private sector data for national security. The Department of Homeland Security (DHS) should develop innovative service delivery models for using information held within and outside government (on trade or specific cargo, for example) and guidelines on the circumstances and procedures for purchasing or requesting access to such data" (p. 37). We also outlined some general principles that should guide government access to, and use of, information from the private sector (pp. 32 to 33).

Working Group II was charged with going beyond the basic principles in our initial report to consider in depth the issue of access to, and use of, private data to meet new security threats and to develop recommendations for the public and private sector. Our goal is to identify the kinds of information that exist in the private sector that are valuable to homeland security and counterterrorism efforts, and to develop a strategy that will allow government the ability to access and use them effectively, but in a way that is most consistent with our national interest in privacy and civil liberties. In our discussions, we specifically addressed six key questions:

1. What information exists in the private sector? Who holds it, and under what strictures?

2. What information does our government need to acquire, retain, and disseminate in order to carry out the homeland security mission?

3. What civil liberty interests are at stake?

4. What rules and oversight mechanisms should govern the acquisition, retention, and dissemination of the information identified?

5. How can technology help with both tasks: assuring that we can use the information effectively and protecting civil liberties?

6. How can we assure that the data collection is cost-effective and that the burden on the private sector is proportionate to the value of the information acquired?

Our report is organized in five interrelated sections. We begin, in Section 1, with a description of the kinds of information held by the private sector, who holds them, and in what form and under what conditions. In Section 2, we look at the kinds of information the government has a legitimate interest in acquiring, and include the relevant time frames for access and use. In Section 3, we discuss the guidelines that should cover access, use, and dissemination of information that we have determined is both available and valuable. In Section 4, we consider how technology can help assure access and use in conformity with the guidelines. And finally, in Section 5, we consider measures to assure the cost-effectiveness of the recommended approach.

The premise of Working Group II is that the government must have access to the information it needs to protect the U.S., and that with well-crafted guidelines, backed up by effective oversight using modern information technology, it will be possible to assure that the government gets that information in a way that protects basic liberties and other important national interests. The objectives of this report are twofold. Our first goal is to provide concrete recommendations concerning the capabilities the government should possess in terms of access to and use of data, which will allow policymakers to develop a goal-oriented plan (including principles that will govern procurement of relevant information technology) to achieve these capabilities. Our second goal is to provide concrete recommendations concerning the policies that should govern the access to, and use and dissemination of, private sector data.

Section 1: The complex world of private sector data

The past decade has seen a truly extraordinary explosion in the quantity of personal information held by the private sector. The exponential increases in both computing and storage capability—at exponentially diminishing costs—have made it both possible and valuable to collect and exploit petabytes of data on virtually every aspect of our lives. Transactional data, such as point-of-sale data, credit card records, travel arrangements, and cell phone call logs, increasingly make it possible to track, in minute detail, the activities of individual citizens. Internet technologies such as the use of cookies allow, at least in principle, access to some of the most private indicators of personal behavior and interest.

All of this data is collected not as a result of government order, but as a consequence of the more or less voluntary decision of citizens to avail themselves of services in return for allowing the provider to collect information on their activities. For the most part, companies collect this data to improve their ability to market their goods and services to their current and future clients. Thus the customer gives up a certain amount of privacy for a benefit. For example, Amazon.com uses customers' profile of past purchases to suggest new titles that may be of interest, and Visa alerts customers to unusual purchasing patterns that may signify a stolen credit card or identity theft.

In recent years, the scale of information collection has been dramatically augmented by the rise of data aggregation companies (companies such as ChoicePoint and Acxiom that acquire data from individual collectors in order to create vast databases that allow users to cross-reference data from diverse sources, including, in some circumstances, public sector information such as driver's licenses and property deed transfers). Data from data aggregation companies has been used for activities ranging from marketing to risk assessment, and even by the government for law enforcement and to track missing children. The wider the range of data, the more favorable the potential cost-benefit for users, who are spared the difficulty of having to acquire and correlate a large number of databases themselves. This is a benefit not only for private sector users but also for the government, including in the homeland security effort.

But there is a flip side to this benefit: From a civil liberties perspective, the implications of data aggregation may be far more significant than the sum of individual data points. This concern exists whether the aggregation is being done by the government (as in the case of the Department of Defense's Terrorism Information Awareness program, formerly the Total Information Awareness program) or by the private sector in support of the government—as can be seen in the controversy over the use of airline passenger data from JetBlue for data aggregation by Army contractor Torch Concepts. The fact that individual pieces of personally identifiable data are freely available does not mean that we can ignore the broader impact of the ability to compile a comprehensive personal dossier. Aggregated data in the hands of the government poses potential risks that are far more consequential than those raised by private sector aggregators.

In principle, individuals can choose to avoid this data collection, either by refusing to transact business with those who use objectionable data practices or by "opting out" (removing one's information from a program that assumes inclusion unless stated otherwise) of specific uses of the data (such as sharing the data with third parties) under companies' privacy policies. Other strategies available to individuals include using anonymizing technologies and providing false personal information.

As a result, data collectors and data aggregators face enormous challenges in assuring and maintaining the value of information they collect. In particular, it is impossible to assure the accuracy and reliability of information, particularly when it is collected from diverse sources under diverse collection protocols. And, of course, keeping the data up-to-date is a particularly important challenge in maintaining the data's value. False or incomplete data will accentuate the problem of both false positives and false negatives. There are even broader implications if the government can access this faulty data and attach consequences to it (for example, restricting the right of an individual to board an airplane).

A variety of rules govern who can acquire information from private citizens and how and when that information can be shared. Under some circumstances, especially where the information is considered to be highly personal and sensitive, the rules are dictated by the government (as, for example, under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which gives patients considerable control over the dissemination of their health information, and in the financial sector, under Gramm-Bliley-Leach and Sarbanes/Oxley). (For a matrix showing the laws governing commercial entities seeking the use of personally identifiable information for risk assessment and other commercial applications, see www.markletaskforce.org.) In other circumstances, the limits are contractual. For example, when using an Internet-based service, users are given the opportunity to opt out of having information shared with third parties by clicking a box on the website. These rules are usually incorporated into each company's privacy policies. However, these rules often do not cover the third-party transfer of non–personally identifying information, such as statistical data on demographics and usage. In addition, as the JetBlue case illustrates, companies' compliance with these privacy policies remains an issue.

Separate rules often govern how and when government agencies can acquire privately held information. In many cases, private sector entities voluntarily share information with the government. Even for strictly regulated areas such as health care, the basic laws governing access to information for law enforcement purposes override legislative or any contractual limits on third-party information-sharing. (For a matrix showing the laws governing government acquisition, see www.markletaskforce.org.) By contrast, without a warrant or similar legal instrument, such as a National Security Letter, the federal government may not collect information that is generally available to the public (such as membership lists of religious organizations). At the other end of the spectrum, some laws (such as those governing suspicious financial transactions) create an affirmative obligation for private entities to collect and share private data with the government.

In Appendix H, we present the landscape of available data, organized by category of information, form, terms under which it is available, whether the information is personally identifiable, and what entities, if any, currently aggregate the data. The appendix helps to illustrate both the extraordinary range of types of data available, and the often bewildering complexity of the rules and procedures governing its acquisition. Our challenge (discussed in Section 3) is to help develop the basic principles and procedures that should govern how and when the government accesses this information.

Section 2: What information does the government need to have? 12 illustrative challenges

Protecting our citizens is the first responsibility of our government. Yet we recognize that part of what we are protecting is the freedom that defines our country's strength. While at times we may face difficult choices concerning freedom and security, we need to be sure that any potential infringement on important liberties is based on the potential for actual security gains. In our first report, we warned about the danger of the vast explosion of available data—that the government would face the temptation to collect it not because it is particularly valuable but because, like Mount Everest, it is there. As we said:

Data mining can be a useful tool. But it is also a tool that invites concern about invasion of privacy. Extravagant claims have been made about the potential uses of data mining, matched by similarly extravagant notions of the vast private or public databases that should be opened to such journeys of exploration. Neither the real needs nor the real capabilities are so exotic.... Data mining, like any other government data analysis, should occur where there is a focused and demonstrable need to know, balanced against the dangers to civil liberties. It should be purposeful and responsible (p. 27).

In this section, we explore the kinds of situations in which there may be a focused and demonstrable need to know certain information. In the next sections, we examine how we can make sure that the government has access to that information in a way that is consistent with our civil liberties.

The debate about government access to private data is too often mired in abstractions, pitting those who cite the theoretical value of certain kinds of information against critics who warn of hypothetical intrusions on liberty. To understand better the kinds of information that are needed to meet real security challenges, our Working Group decided to look at a number of concrete, plausible scenarios that our government might face (see Appendix F). Of course, these examples are only illustrative. But as a heuristic device, they help to answer the following questions:

1. What information is truly necessary?

2. What technological capabilities does the government need to acquire in order to gain access to the information in a timely, useful way?

3. What potential civil liberties violations and other concerns must be addressed by policies governing the circumstances under which the information is acquired and used?

As we examined each of the scenarios in detail, it became clear that information needs revolve around four basic questions: "Who?" "How?" "Where?" and "When?" These four questions are the key variables in trying to thwart an attack on our country. (To see how these questions can help us to develop information strategies to meet the security challenge, see Appendix E.)

Who?

In our first six challenges, we present data issues that arise when something is known about the identity of a potential terrorist—by far the most productive approach to preventing terrorism, and the most common focus of counterterrorism investigations. At the same time, the search for information related to "who?" frequently leads to requests for personally identifiable information. Therefore it is particularly important to be clear about what information is truly valuable enough to justify the potential intrusion on civil liberties.

Challenge 1 focuses on tracking a known suspect and his or her confederates. In it, we outline data that would be useful and the time frame in which it is reasonable to expect that the data be accessible. In Challenge 1 there is particularized, evidence-based suspicion about the individuals. Thus there is a high value associated with gaining access to such information as phone listings, DMV records, basic financial data, INS visitor and immigration information, academic enrollment, special licenses, and travel records. With appropriate safeguards in place (discussed in Section 3), these agencies must then have the technological capability to identify the suspects' associates, in a very short time, through shared addresses, phone and email records, financial transactions, travel records, and common memberships in organizations.

Challenge 2 focuses on the question of whether, under some circumstances, the government needs to take steps to improve the private collection of data—in this case, on foreign students in the U.S. The argument for greater scrutiny of foreign students is based on two factors: the fact that some terrorists in the past have used student visas to enter the U.S.; and that there is an associated legitimate purpose to the data collection, which is to assure that students comply with their visa conditions. The scenario illustrates the kinds of information that would be of value in determining whether a student is in status. While the information is personally identifying, it is limited to information relevant to a legitimate government purpose (in this respect, the scenario is analogous to government data requirements associated with regulatory functions, such as the anti–money-laundering laws). More troubling questions would be raised if the desired data included, for example, information on the student's religious practices. At the same time, even if legitimate, requiring private sector entities to collect information they would not otherwise collect has a cost, which places an additional burden on the government to demonstrate that the value of the information outweighs the cost.

Challenges 4 and 5 concern sharing information on identity: Who should be able to access information on identity and in what form, both to protect privacy and to assure security? As our first report demonstrated, timely, effective information-sharing—including sharing with state and local government and the private sector—is at the heart of a successful approach to meeting the new security challenges. At the same time, the wider the dissemination of the information, the greater the risk that the information could be used for improper purposes, particularly if the information is personally identifiable. Challenge 3 involves integrating local law enforcement agencies into federal counterterrorism efforts to prevent suspects from slipping through the cracks. This might entail having a system of automatic tailored alerts in place, which get triggered when local agencies run the documentation of a terrorist suspect to determine if the suspect is on a federal watch list. Challenge 4 involves information requirements associated with developing a consolidated watch list from those of different agencies. These two scenarios demonstrate how technology can be used to mitigate a number of the problems associated with widespread data sharing, including improper use and protecting the security of sensitive information. Tools for these purposes include the following: (1.) anonymous identity resolution (a privacy-enforcing method in which analysis is performed only on anonymized data, thus eliminating the need for organizations to share personally identifying data); (2.) "one-way hash" (a mathematical technique that changes a piece of data into an abstract number that cannot then be reversed to its original value); (3.) advanced user authentication; (4.) use of identity metadata; and (5.) anonymization and audit practices.

Challenges 5 and 6 focus on two key accuracy issues concerning data on identity: false positives from inaccurate or ambiguous data (the David Nelson problem[1]) and false negatives from false identities, etc. Accuracy is vital not only to protect the privacy and civil liberties of individuals who would be harmed by the use of inaccurate data, but also to assure that information has real value to the counterterrorism effort. In Challenge 5, we identify some of the technologies that can help assure that information is up-to-date; in Challenge 6, we address the critical question of how to deal with the problem of false and stolen identities. Technology of course, is only one part of the solution; we also need policies that make it possible for individuals to have an opportunity to correct errors while preserving the necessary security of the data.

The accuracy problem is one that deserves considerable attention in assessing what data is useful to the government. According to industry experts, most data integration today is based on only name and address (although in some circumstances additional information, such as social security numbers, dates of birth, or driver's license data, is available). Name and address information is captured in a multitude of formats that allow errors to be introduced. In addition, this information is frequently out of date: 20 percent of the population moves every year; 5 percent has second homes; 5 million marriages and 2 million divorces occur annually, many resulting in name changes; and 8.7 percent of the population dies every year. Data integrators have developed sophisticated techniques to help deal with some of these problems (for example, algorithms that recognize that Bob equals Robert or that more data is needed to match a common name than a rare one). But, at best, these techniques have reduced the error rate to 1 to 2 percent.[2] Whether this level of accuracy is useful will depend to a considerable degree on how the information is used. If it yields a false positive that imposes only a minor inconvenience (for example, by subjecting an individual to a more intensive airport screening process) but demonstrates high value in identifying potential suspects, the benefit may justify the cost. Conversely, if a false positive imposes significant consequences, the requirement for data accuracy should be more stringent.

[1] The reference is to a real-world experience in which the relatively common name *David Nelson* was placed on a "do not board" aviation security watch list. Innocuous David Nelsons found it very difficult to establish that they posed no danger and should be permitted to fly.

[2] This data was presented to Working Group II by Jennifer Barrett of Acxiom.

How? Where? and When?

The shadowy nature of terrorist networks means that in some circumstances we will know little, if anything, about the identity of potential adversaries. But there are circumstances that may suggest a potential target (for example, the receipt of reports about possible attacks on the Golden Gate or Brooklyn Bridges); a potential means of attack, such as chemical agents spread through crop dusters; or a time of attack, such as an anniversary associated with past attacks. These pointers may arise either through specific intelligence (suspicious activity, intercepts, etc.) or through contextual analysis of the threat (targets that are of high symbolic or economic value, or past threats or attacks). It is far more difficult to formulate meaningful, focused data requirements under these circumstances than with cases in which there is information pointing to a specific individual. Therefore, in such cases it is important to develop tools and methodologies that assure data requests are more than fishing expeditions—not only to prevent unwarranted intrusions on privacy but also to conserve valuable investigative resources. In our initial report, we outlined a number of analytic approaches to this challenge.[3]

Challenges 7, 8, and 9 illustrate potential data needs when something is known about the mode of attack (for example, information on specific individuals who have access and capability to employ that mode of attack, and information on facilities where the means are stored, sold, or transported).

In Challenge 7 we consider an example in which the government knows the mode of attack (a scuba diver attack on a hazmat tanker). In principle, it might seem desirable to run a background check on all 1 million certified scuba divers in the country.[4] Fortunately, that is not quite as daunting as it appears: Two national certification agencies —the National Association of Diving Instructors and the Professional Association of Diving Instructors—hold information on more than 80 percent of all U.S.-certified scuba divers. But even with that information, there may be serious false positives. Just what background data would constitute a hit? Certainly, past travel to Afghanistan might be a worry, but what about a long record of traffic violations? And, of course, there is also a risk of false negatives—the terrorists might have hired an unlicensed diver, or one deliberately chosen because he or she has a clean record.

The value of this kind of information can be enhanced by the development of "training sets" (rich sets of transactional history used to "train" software, especially to detect normal versus abnormal behaviors) that build on experience to allow refinement of the search and increase utility. (The link with "travel to Afghanistan" is an example of a training set.) These models might initially be developed through "red-team" exercises (simulations that provoke thinking like an adversary in order to better identify vulnerabilities), and then validated through experience. To protect civil liberties in a case such as this where there is not a particularized suspicion of an individual, anonymizing techniques should be used until the point at which the virtual background investigation raises an articulable and concrete suspicion.

Challenge 8 (What?), Challenge 11 (Where?), and Challenge 13 (When?) focus on cases in which we know something about the target or timing of an attack and want to acquire information concerning both vulnerabilities of the target and those who might have access to it. Vulnerability issues rarely pose civil liberties concerns; rather, the data issues involved more typically concern the willingness of the private sector to share the information in ways that do not jeopardize competitive advantage or trade secrets or expose the vulnerabilities to those who seek to do harm. Thus, in these cases, data security and limits on third-party sharing must be developed through a combination of technologies and policies (such as the recently enacted, and still controversial, exemption of critical-infrastructure data from the Freedom of Information Act (FOIA). Of course, data on those with access to potential targets does raise questions about personally identifiable information. But in the case of especially sensitive sites, requirements of preemployment clearance may be appropriate and may help avoid the problems of unfairness and violation of privacy that are associated with ad hoc data collection.

Challenge 9 deals specifically with the vulnerabilities posed by the vast number of cargo containers entering our ports, and indicates that the government should be able to determine the past history of inbound containers and be able to identify suspicious patterns before any container reaches U.S. waters. In this case, the challenge is largely a technological one: to develop the sensors, networks, and associated protocols that allow for tracking and monitoring a complex system.

[3] See *Protecting America's Freedom in the Information Age: A Report of the Markle Foundation Task Force*, pp.46-47.

[4] There is some discrepancy regarding the number of certified scuba divers in the U.S. Most estimates are between 1.5 and 3.5 million. However, the Professional Association of Diving Instructors estimate that there are 8.5 million.

Challenge 10 focuses on a unique dimension of "how?": the availability of financial resources to support terrorist operations. To restrict this availability, financial institutions conducting reviews should be able to identify account holders whose finances reflect indicia of concern, such as irregular deposits from overseas. Further, it should be possible to review the background of such account holders for other indicia of concern on a rapid basis. At the same time, these requirements pose serious issues concerning privacy (as well as efficacy of the associated data searches). There is considerable uncertainty about what patterns or practices of financial activity are associated with terrorism, which leads to considerable problems of both false positives and false negatives, with considerable intrusion upon an area of great sensitivity. Therefore, as with other cases of sensitive personal information, absent an articulable suspicion—such as a cross-match between a suspicious financial activity report and a terrorist watch list—anonymization techniques and restrictions on data use seem appropriate.

Challenge 12 concerns data requirements associated with responding to attacks. As the September 11 example shows, the ability to mobilize and interconnect resources is a critical component of attack response, with demanding requirements for data collection and sharing. Many of the technologies associated with meeting the other data challenges can also be applied to meeting this requirement.

These illustrative scenarios are designed to help stimulate thinking about the kinds of information the government needs to carry out its homeland security responsibilities. While it is inherently impossible to specify in advance all the kinds of information that may be relevant to this mission, it is a well-established practice (as part of the process of intelligence-collection prioritization) for intelligence consumers to identify for intelligence collectors the information they believe they need to carry out their responsibilities.

For that reason, Working Group II recommends that the U.S. government, under the leadership of the Director of Central Intelligence and the Secretary of Homeland Security, should conduct a government-wide review of its information-collection requirements and develop a plan (to be periodically updated) for meeting the information-collection and information-analysis needs outlined in this section. This effort should be integrated in the overall intelligence community prioritization and tasking process, and should be subject to appropriate oversight and review by Congress.

As the discussion of the individual challenges makes clear, developing a strategy for identifying the information the government needs to meet its national security challenges must go hand in hand with the development of appropriate policies and technologies associated with the acquisition and use of private data. We turn to these issues in Sections 3 and 4.

Section 3: Guidelines for government acquisition, storage, and retention of private data

In our initial report, we offered 12 principles that we believed should govern the acquisition, retention, and dissemination of information from the private sector. Working Group II endorses those principles (listed again here) and, in this report, offers an additional five.

17 PRINCIPLES THAT SHOULD GOVERN THE ACQUISITION, RETENTION, AND DISSEMINATION OF INFORMATION FROM THE PRIVATE SECTOR

1. **Importance of access to information in public and private hands**
   Access to information in the hands of public and private entities is an essential tool in the fight against terrorism. Government agencies responsible for combating terrorism—including state and local as well as federal authorities—should have timely and effective access to needed information, pursuant to appropriate legal standards. The legal constraints and exceptions provided by current law are generally sufficient to allow a homeland security agency to gain necessary access to information held by other government agencies. These new guidelines offer a framework and procedures to allow that information to be effectively used, analyzed, and disseminated. At the same time, these guidelines are intended to ensure that information about people in the U.S. is used in a responsible manner that respects reasonable claims to individual privacy.

2. **Purpose and interpretation**
   These guidelines should be interpreted and applied in a fashion that encourages rapid, effective, and responsible access to data that can assist in the task of identifying, thwarting, or punishing terrorists. These guidelines

should also be interpreted and applied in a manner that encourages respect for fundamental liberties, creativity, innovation, and initiative in the use of data for the purpose of fighting terrorism.

In addition, these guidelines should be used only for the gathering and analysis of information for intelligence in the war against terrorism. The procedures and authorities for using the legal process to obtain information for law enforcement purposes should remain unchanged.

3. **Coordination and authorization**
   An intergovernmental body, chaired by the Secretary of Homeland Security and composed of representatives of the relevant federal, state, and local agencies, should be formed to coordinate the procurement and use of private, state, and local databases containing information about U.S. citizens. Because databases have varying degrees of utility, privacy interest, and reliability, our Task Force believes that a single point of coordination would provide accountability for privacy concerns and would allow for the effective and efficient use of information. In addition, that intergovernmental body would provide a focal point for private companies' and state and local administrators' concerns about burdensome, duplicative, and inconsistent requests for information.

   Similarly, the authorization for procuring or requesting access to databases should not be burdensome on investigators and analysts. With regard to these guidelines, we envision a process in which a single authorization for the procurement of the database will be sufficient for all necessary and continuing access by agency personnel, if it is for the authorized use.

4. **Relevance**
   Agency personnel should have access to, and use of, information available under these principles only for purposes relevant to preventing, remedying, or punishing acts of terrorism.

5. **Accountability**
   Agencies and their employees should be accountable for the ways in which they access and use information available under these guidelines. An agency should be able to identify how its uses of databases are relevant to preventing, remedying, or punishing acts of terrorism. While it would be plainly inconsistent with the purposes of these guidelines to require that an agency or employee explain the relevance of every query before gaining access to data, mechanisms such as database-access records, audits, and spot checks should be used to ensure that agencies move toward demonstrable compliance with this principle.

6. **Dissemination and retention**
   Information about U.S. citizens should not be disseminated or retained by the collecting agency unless doing so is demonstrably relevant to the prevention of, or response to, an act of terrorism. Administrative rules, training procedures, and technology should be implemented to prevent the unauthorized disclosure of private personal information. An electronic audit trail of how information is used—and penalties for misuse—can reinforce these guidelines.

7. **Reliability of information**
   Agencies should strive to use the most accurate and reliable information available. Nevertheless, data used under these guidelines may include information of questionable or varying reliability. Where feasible, and to promote effective antiterrorist action, limitations on the reliability or accuracy of data should be made known to those using the data. In the event that an agency determines that information is materially inaccurate and that an individual is likely to be harmed by future use of that inaccurate information, reasonable efforts should be made, and a process put in place, to correct the inaccuracy or otherwise avoid harm to the individual concerned.

8. **Information-technology tools**
   To the extent consistent with the purpose of these guidelines, information-technology tools should be developed and deployed to allow fast, easy, and effective implementation of the relevance, accountability, and reliability principles of these guidelines. Consistent with a vigorous defense against terrorism, we envision tools that create audit trails of parties who carry out searches; that anonymize and minimize information to the greatest extent possible; and that prevent both the intentional and unintentional dissemination of irrelevant information to unauthorized persons or entities.

9. **Information in the hands of intermediaries**
   Much of the information relevant to the fight against terrorism will be in private hands. As a general principle, and where consistent with the purposes of these guidelines, it is preferable to leave information in the hands of private intermediaries, rather than consolidating it in agency databases. In many cases, government agencies are forced to transfer information into an already-existing government database because the agencies do not have the tools needed to search the data while keeping the

information separate from their own. Agencies are encouraged to develop and deploy tools that would allow these separate searches of privately held data, thereby allowing information to remain exclusively in private hands.

Private databases are not created for the government. Private parties create them for their own commercial purposes. Because of this, private databases are subject to the constraints of the marketplace. An agency seeking access to such databases should treat these intermediaries fairly. In particular, the agency should do the following: (1.) preserve necessary confidentiality, and protect intermediaries from liability for any assistance they may provide to the agency in good faith; and (2.) use commercial contracts or similar arrangements to compensate intermediaries for any assistance provided to the agency.

Agencies should initiate and maintain a cooperative dialogue with the private sector to develop voluntary data-retention policies that maintain information necessary for the war on terrorism. Agencies should endeavor to identify critical information and advise private firms of the importance of their voluntary efforts to retain such data. If necessary, the government may even encourage the formation of self-policing groups within the private sector to help achieve the data-retention objectives. In other words, the more the government does to articulate specifically what information should be retained and why, the greater the obligation the private sector should feel to cooperate with these agency requests. In a narrowly defined set of circumstances, such as with airline passenger manifests and sales of certain biological pathogens, data retention may, appropriately, be required.

10. **Revisions and public comment**
These principles are preliminary steps toward establishing the fundamental authorities and protections for the use of information in thwarting terrorism. They should be reviewed, revised, and made more specific in the light of actual experience. These guidelines, and any future revisions and specific rules that are established based on them, should be available to the public and subject to public comment—unless the President finds that disclosure will endanger classified intelligence collection or analytic methods and threaten national security.

11. **Agency implementation**
Compliance with these guidelines should be achieved to the greatest extent possible through training, advice, and quick correction of problems, rather than through after-the-fact punitive measures that may lead antiterrorism agencies or employees into risk-averse behavior. In addition, investigations of suspected violations should be performed by a single office and should focus principally on systemic measures to avoid future violations.

12. **Congressional oversight**
Nothing in these guidelines restricts review of the guidelines by Congress. Members of Congress or congressional staff conducting reviews of the guidelines or their implementation should expressly agree to protect the privacy of individuals, classified information, and confidential sources and methods used to combat terrorism.

13. **Early implementation**
The guidelines should be implemented from the beginning of the government's efforts to integrate its data collection from now-disparate public and private sources.

14. **Ease of use**
Guidelines for governmental collection, use, and dissemination of data should be clear and easy to follow. There should be a relatively small number of different standards and procedures for the government and the private sector to observe.

15. **Transparency**
Because it is imperative that the public—both individuals and private companies possessing databases—feels it can place its confidence in the government's actions as being in accordance with the rule of law, guidelines for data collection should, on the whole, be publicly available. Some guidelines may need to be classified for security purposes, but in general, the public should be granted access to the standards by which the government is acting in its efforts to collect and analyze data for counterterrorism purposes. (This principle augments principle 10, above.)

16. **Different standards for different applications**
Guidelines should include different standards for different activities related to the use of public and private databases. The need for such variance derives

from the fact that different privacy concerns are implicated by both the nature of the information acquired and the use to which it will be put. In addition, the standards need to take into account both the specificity and the urgency of the need. Specifically, guidelines should differentiate among the following: (1.) the acquisition of data; (2.) the implementation and oversight of the use of data; (3.) the retention of data; and (4.) the dissemination of collected data.

17. **Avoidance of premature stovepiping**
The government ought to have the capacity to quickly—ideally in real-time—collect information related to counterterrorism efforts. When that data is first collected, the government ought not to be constrained to identify whether the data will be used for intelligence or law enforcement purposes. Rather, identification of eventual use should be delayed until after the data has been collected and subjected to initial review. This way, the nature of the data will influence its eventual use, instead of having its use determined before the relevant agency has had an opportunity to discover its characteristics and value. In addition, characteristics of the processes of data acquisition and dissemination should be recorded so that collected data may be used as evidence in legal proceedings.

Of particular note is principle 16: different standards for different applications. As we discussed above, different types of data, the way in which the data is collected, and the use to which it is put all affect privacy and other civil liberties concerns. Therefore, policies need to be tailored to take these factors into account, while keeping in mind the admonition, in principle 14, that the guidelines be easy to use. This means the development of a reasonably small number of standards (and associated procedures for applying those standards) that treat reasonably similar data in the same way, while recognizing that each phase of the operation—collection, retention, dissemination—raises unique issues.

Guidelines concerning acquisition

There are a number of factors that affect the degree of sensitivity of information in the private sector, which we identify in Section 1, above. These include the technique by which the data was acquired; the subject matter of the information; whether it is personally identifying; and whether it was collected with a promise of confidentiality vis-à-vis third persons. Different levels of sensitivity warrant greater degrees of scrutiny before acquisition of private data should be allowed. In addition, the sensitivity of the information must be measured against the urgency of the need and the relevance of the information to a specific need. That is to say, just because information is not sensitive, or because it is broadly available to private citizens or entities, does not automatically mean that the government should have access to it. A higher bar of relevance to a legitimate purpose applies to government acquisition, and that bar should be even higher with greater degrees of sensitivity. In addition, as noted above, aggregation of data, even data that individually might seem inoffensive, poses distinct issues, that must be taken into account in establishing what kind of need the government must demonstrate before acquiring the data.

Under existing law, there is a patchwork quilt of standards, with different standards for information with similar sensitivity (such as wire, cable, and Internet communications) and inappropriate or nonexistent standards for others. To help think about the policy choices that should govern acquisition of private sector data, we believe it is useful to have three broad levels of required scrutiny for data acquisition—low, medium, and high—and that data should be classified accordingly. For each level, there is an associated standard that the government must meet to justify the acquisition of the information and a companion process to assure that the standard is met. Even for information that has little or no sensitivity (such as non–personally identifying information), we believe that the decision by government to acquire it must be based on more than a whim. That is, there must be some connection to the underlying mission, and there must be some procedures to assure that such information is not acquired for an impermissible purpose. For nonsensitive information, after-the-fact audit and review should be adequate. With increasing levels of sensitivity, the bar should be correspondingly higher, and procedural protections should increase.

Proposed data-classification structure and acquisition requirements

LOW	MEDIUM	HIGH
Types of information Non–personally identifiable data; information concerning non–U.S. persons	**Types of information** Personally identifiable information that would be available without restriction to private citizens	**Types of information** Private, personally identifiable information not generally available to private citizens and entities; all personally identifiable information on sensitive topics (health, financial, and First Amendment activity, such as communications content), whether or not it is available to private citizens and entities
Standard Request for access to information is reasonably related to a homeland security mission	**Standard** Specifically identifiable facts suggest that the information is relevant to a counterterrorism mission	**Standard** Request for data is necessary to obtain valuable intelligence information related to a threat to the U.S.
Process Training and post-facto periodic review; no a priori approval required	**Process** Senior official signoff prior to acquisition	**Process** Foreign Intelligence Surveillance Act–type process (involving a judge or other third party, such as a magistrate)

IMPLEMENTATION AND OVERSIGHT

Policies have value to the extent that there is confidence that the policies are followed in practice. Working Group II therefore places particular importance on mechanisms to ensure compliance. These mechanisms include training personnel, rigorous record-keeping, technological tools that embody the policies, maximum possible transparency (consistent with the mission), periodic review, and enforcement mechanisms. In particular, we advocate the following six practices.

1. **Organizational oversight**
   There must be organizational oversight of the data-collection and use process. The integrity of the government's efforts to collect and analyze data from disparate databases is essential both for efficiency and for privacy protection purposes. Accountability and access control are necessary elements of an efficient, sustainable process. As such, guidelines need to be enforced through auditing and permissioning systems that are integrated from the beginning. (This assertion supplements our fifth guideline, below.)

2. **Dispute resolution**
   There must be a dispute-resolution mechanism in place to ensure that disputes between the public and private sector, or between individuals and data collectors and users, can be resolved. This is important with respect both to the process by which information is acquired and the accuracy of the information.

3. **Dialogue with industry**
   To make the process more efficient for both businesses and the government, there should be a forum for dialogue between the two, in which matters of concern can be discussed.

4. **Training**
   To ensure effective compliance with the guidelines and systems, there must be appropriate training of government personnel throughout their government service.

5. **Technology**
   Technology that would facilitate proper use of data and compliance with the guidelines must be utilized.

6. **Consequences for violations**
   If data is misused or there is noncompliance, there must be penalties that are imposed on the violators appropriate to the nature of the violation.

Guidelines concerning retention

Principle 9 indicates our strong preference for leaving private data in private hands, rather than having the government retain it in its own databases. The rationale for this principle is largely prophylactic—it makes it harder for the government to acquire information for one purpose and then use it for another. This is particularly important if we want to facilitate access to information for counter-terrorism purposes but insure that the accessed information is not then used for purposes that would otherwise require stricter procedures or additional protections. Therefore, the guidelines should provide that, if the government wants to retain data gathered from the private sector, it must show that its inability to retain the information would, for example, substantially impede the counter-terrorism mission. Wherever possible, the government should seek to rely on pointers or directories that identify where data can be located in the private sector rather than retaining the underlying data. When the government does retain data, that data should not be commingled with nonrelated databases, absent reliable procedures to assure that commingling would not allow the data to be used for impermissible purposes (see below).

In some circumstances, the government may need to retain information that is broadly related to the counter-terrorism mission, though not necessarily related to a specific case. For example, basic information needed to conduct entity checking (the ability to differentiate among the David Nelsons) is a legitimate basis for retaining information in government databases. For this kind of information, appropriate restrictions on use will provide needed protection.

Guidelines concerning the dissemination of data from the private sector to other government users and the private sector

Consistent with our overall network approach to the desirable information-sharing architecture, information legally acquired for counterterrorism purposes should flow as freely as possible within the community necessary to conduct the mission at all levels. Wherever possible, an effort should be made to use anonymized information. But in many cases the personally identifiable information will be indispensable. To prevent abuse of information for unrelated purposes, procedures should be established that would tag information in a way that would block its use for other purposes or, alternatively, would alert other potential users that use of the information was restricted. At the same time, the government should not be forced to face artificial hurdles to using information for legitimate purposes. If an agency other than the acquiring one has the legal right to acquire the information directly from the private sector, it should be able to acquire it from the original acquiring agency so long as the standards by which the second agency could acquire the information from the private sector are similar to, or lower than, those governing the acquisition by the initial agency. Procedures should be put in place to assure compliance with this principle, but the acquiring agency should not be required to police how the second agency actually uses the data. The burden of compliance should rest on the agency that actually uses the information.

These principles need to be applied to all government information-gathering, retention, and dissemination. Therefore, Working Group II proposes the following: The President should issue an Executive Order—after public notice and comment and consultation with Congress—embodying these principles and the applicable standards. Although portions of the Executive Order may need to be classified, the President should make the maximum effort to issue unclassified guidelines. The DHS should be given the lead on implementation and oversight, to ensure that all agencies implement the guidelines, and should have in place procedures to assure that they are complied with.

Section 4: The role of technology

Information technology both creates and helps solve many of the issues involved in the interaction between government and the private sector. Information technology has made it possible to collect, store, and collate vast quantities of information, thus assuring its potential availability and utility in counterterrorism and homeland security measures. Equally important, these technologies can aid in the implementation and enforcement of safeguards that will help ensure that the information is put to proper use. In this section, we discuss the technologies that are needed to meet the 12 challenges identified in Sec-tion 2 and what it will take to make sure they are deployed. We will then discuss how technology can support the application of the safeguards proposed in Section 3. (For a list of the necessary technologies for each of the scenarios, see Appendix G.)

The capabilities identified are those that the federal government can and should develop in the near term (less than five years) to bring our data-processing capabilities to bear on the problem of terrorism. These capabilities focus principally on the federal watch lists and the use of data currently in private hands to allow civil authorities to locate and pursue suspected terrorists within our borders. All of these capabilities are achievable with resources and technology now available or in development. Indeed, many are currently in use by private industries.

Taking the list of key technologies as a whole (see Appendix G), we can see that they fall into several categories. A number of them concern enhancing the value of the data, including assuring data quality, while others focus on cross-correlation of diverse data sources. Some are concerned with the effective communication of the data. Some are related to ensuring data security. Some contribute to the implementation of policy guidelines and oversight. The list provides a highly focused, concrete checklist for policymakers and information-technology managers to guide procurement planning and research support over the coming years. This technology checklist should be subject to ongoing review and updates, through a collaborative process involving both the government and the private sector, to identify needs and emerging technologies that can meet those needs.

Working Group II recommends that the Office of Management and Budget, in conjunction with the DHS, conduct a government-wide review of the information-technology acquisition and implementation plans of all relevant agencies, and that it issue a comprehensive plan to assure that the technologies are procured and implemented within the time frames identified in this report.

Section 5: Cost-effectiveness and market dynamics—focusing investigatory resources

As we have stressed throughout this paper, access to private sector information is essential to the homeland security mission. But indiscriminate, ill-thought-out requests for information not only pose risks to civil liberties, they potentially place a serious burden on the private sector holders of the information. Equally important, a vacuum cleaner approach may actually impede homeland security efforts by inundating the government with information of little or no value, thus complicating the agents' ability to distinguish signal from noise and wasting valuable investigatory resources.

For all those reasons, it is important that requests for data from the private sector be focused on information that adds value. Market mechanisms can help ensure that government officials take into account the costs and benefits of data requests (for example, by requiring the government to compensate private holders for the costs of furnishing data, including data aggregation, as well as the actual costs of sending the information to the government). This requirement should apply, in particular, to cases in which the requests are ongoing; costs are high; the cost of complying might put the holder at a competitive disadvantage vis-à-vis those who are not asked to furnish information; and, especially, in which the holder is in the business of data aggregation. The government should enter into an ongoing dialogue with members of the private sector who are likely to be the subject of repeated requests, in order to formulate procedures that would minimize the impact on the private sector while assuring that the government is able to access the information it needs.

The market already prices much of the data that the government is likely to request. For that which is not priced, cost equations can be developed by a consortium of members of the private and public sector on the basis of the scope of information being requested and the timing and complexity of the request. In the absence of an agreed price list reflecting the range of costs and circumstances of purchase, fair-price mechanisms can be used for estimating costs, with some kind of accounting or arbitration system in place to oversee the process.

At the same time, private sector holders of information, be they individuals or corporations, also have some responsibility as citizens to assist in carrying out this vital national mission. Thus, in cases where the requests are infrequent and the costs are low, Working Group II believes that requiring compensation would be inappropriate. In these cases, appropriate employee training—supplemented by periodic, post hoc agency reviews—should be conducted to assure that government officials are sensitive to cost-benefit considerations in formulating data requests.

In many cases, the same policies and technologies that are designed to help safeguard privacy and civil liberties can also help assure that the value of the information sought is proportionate to the burden. Focused searches, based wherever possible on clear and articulable suspicion, with strong oversight to assure that standards are met, are likely to provide the highest yield at the lowest cost to important national values.