Creating a Trusted Information Network for Homeland Security

Second Report of the Markle Foundation Task Force

ZOË BAIRD, JAMES BARKSDALE
CHAIRMEN

MICHAEL A. VATIS
EXECUTIVE DIRECTOR

MARKLE FOUNDATION

MARKLE FOUNDATION

TASK FORCE ON NATIONAL SECURITY IN THE INFORMATION AGE

Chairmen

Zoë Baird
Markle Foundation

James L. Barksdale
The Barksdale Group

Executive Director

Michael A. Vatis
Markle Foundation

Members

T. Alexander Aleinikoff
Georgetown University Law Center

Robert D. Atkinson
Progressive Policy Institute

Stewart Baker
Steptoe & Johnson

Eric Benhamou
3Com Corporation

Jerry Berman
Center for Democracy and Technology

Robert M. Bryant
National Insurance Crime Bureau

Ashton B. Carter
Harvard University

Wesley K. Clark
Wesley K. Clark & Associates

G. Wayne Clough
Georgia Institute of Technology

William P. Crowell
Security and Intelligence Systems

Sidney D. Drell
Stanford University

Esther Dyson
EDventure Holdings

Amitai Etzioni
The George Washington University

David J. Farber
Carnegie Mellon University

John Gage
Sun Microsystems, Inc.

Slade Gorton
Preston, Gates & Ellis

Morton H. Halperin
Open Society Institute

Margaret A. Hamburg
Nuclear Threat Initiative

John J. Hamre
Center for Strategic and International Studies

Eric H. Holder, Jr.
Covington & Burling

Arnold Kanter
The Scowcroft Group

Michael O. Leavitt
Governor of Utah

Tara Lemmey
Project LENS

Gilman Louie
In-Q-Tel

Judith A. Miller
Williams & Connolly

James H. Morris
Carnegie Mellon University

Craig Mundie
Microsoft

Jeffrey H. Smith
Arnold & Porter

Abraham D. Sofaer
Stanford University

James B. Steinberg
The Brookings Institution

Paul Schott Stevens
Dechert LLP

Rick White
TechNet

Philip Zelikow
Miller Center of Public Affairs,
University of Virginia

Participating Experts (non-government)

Bruce Berkowitz
RAND

Robert Clerman
Mitretek Systems

James Dempsey
Center for Democracy and Technology

Mary DeRosa
Center for Strategic and International Studies

Lara Flint
Center for Democracy and Technology

Lauren Hall
Microsoft

Jeff Jonas
Systems Research & Development

James Lewis
Center for Strategic and International Studies

Terrill D. Maynard

Mary McCarthy
Center for Strategic and International Studies

Patrick J. Sullivan, Jr.
Cherry Creek Schools

Winston Wiley
Booz Allen Hamilton

Associate Director

Mary McKinley
Markle Foundation

Task Force Staff

Nancy Boursiquot
Administrative Assistant

Todd Glass
Director, Public Affairs

Caroline McCarus
Assistant to the President

Jennifer Obriski
Administrative Assistant

Stefaan Verhulst
Chief of Research

Creating a Trusted Information Network for Homeland Security

SECOND REPORT OF THE MARKLE FOUNDATION TASK FORCE

December 2003

A Project of

The Markle Foundation
New York City

In Alliance with

The Brookings Institution
Washington, DC

Center for Strategic and International Studies
Washington, DC

ISBN 0-9747809-0-1

Table of Contents

Part Two: Working Group Analyses 43

Part Three: Appendices 71

Overview

This is the second report of an extraordinary task force we have been privileged to co-chair. This remarkable and diverse group has come together to serve our nation by doing the hard work of considering how we can create an information network that prevents terrorism and protects the security of our homeland, while preserving the civil liberties that are a fundamental part of our national values.

In the Task Force's first report, we stressed the importance of creating a decentralized network of information-sharing and analysis to address the challenge of homeland security. We emphasized the need to form that network around presidential guidelines shaped by public debate on how to both achieve security and maintain liberty. We also set forth principles for capitalizing on our society's strengths in information technology. In this second report, we reaffirm those principles and provide greater detail on how to implement our approach.

The network we envision would be created with the following key elements, which reflect the character of the distributed, asymmetric threat we confront:

1. The handling of information should be decentralized, and should take place directly among users, according to a network model rather than a mainframe or hub-and-spoke model.

2. The network should be guided by policy principles that simultaneously empower and constrain government officials by making it clear what is permissible and what is prohibited.

3. Our government's strategy should focus on prevention.

4. The distinguishing line between domestic and foreign threats is increasingly difficult to sustain. Thus, in its approach, our government should avoid creating blind spots, or gaps between agencies, that arise from this distinction. At the same time, though, our government needs urgently to define new rules—rules to replace the old "line at the border" between domestic and foreign authorities for information-collection and use—to ensure that agencies do not infringe on our traditional civil liberties.

5. The network should reflect the fact that many key participants are not in the federal government, but rather in state or local government and the private sector.

6. The network should make it possible for the government to effectively utilize not only information gathered through clandestine intelligence activities and law enforcement investigations, but also appropriate information held by private companies. This should happen only after clear articulation by the government of the need for this information and the issuance of guidelines for its collection and use.

7. Combating terrorism is a long-term effort that is designed to protect our way of life and our values along with our security. Therefore, the policies and actions undertaken need to have the support—and trust—of the American people. Privacy and other civil liberties must be protected.

What do these principles mean in practice?

First, our government should give greater priority to sharing and analyzing information. In the Cold War intelligence architecture, the government placed a premium on the security of information. It developed a system that tightly controlled access to information by requiring that every individual have a demonstrable "need to know" certain information before he could see it and by allowing the agency that initially acquired the intelligence to restrict further dissemination of that intelligence. This system assumed that it was possible to determine *a priori* who needed to know particular information. And it reflected the judgment that the risk of inadvertent or malicious disclosure was greater than the benefit of wider information-sharing.

This architecture and the underlying assumptions are ill suited to today's challenges. The events of September 11, 2001 have starkly demonstrated the dangers associated with the failure to share information, not only within the federal government, but also between the federal government, on the one hand, and state and local governments and the private sector on the other. Therefore, the government should open up the system to state and local agencies and officials and, in some circumstances, to private sector actors, providing access not just to information but to technology and money as well. Our government should reengineer operational processes where needed and build the technology architecture and tools that will facilitate two-way sharing and interoperability. Our government should also take into account the needs of the users, as well as the agency that originally developed the information, in deciding whether or how to control where the information goes. This should take place in an environment in which the need to protect both the security of sensitive information and individual civil liberties is consistently addressed.

Furthermore, our government should effectively utilize the valuable information that is held in private hands, but only within a system of rules and guidelines designed to protect civil liberties. Our nation can never hope to harden all potential targets against terrorist attack. Therefore, we must rely on information to try to detect, prevent, and respond to attacks. The travel, hotel, financial, immigration, health, or educational records of a person suspected by our government of planning terrorism may hold information that is vital to unveiling both his activities and the identities and activities of other terrorists. But until the government devises consistent guidelines for controlling the justification for when and how such information is accessed and used—and until those guidelines are publicly debated—the public's concerns over potential privacy infringements will continue to hamper the necessary development of new technologies and new operational programs to use that information.

The need to create the network we envision is more urgent than ever. Terrorism remains a continuing threat around the world. And the potential for terrorists to use weapons of mass destruction raises the stakes considerably. Building the technical architecture, changing agency cultures, establishing new rules and procedures, and securing the necessary funding all take time. It is therefore imperative that the steps we recommend receive immediate attention. We urge the Executive Branch and Congress to implement the measures necessary to create the proposed Systemwide Homeland Analysis and Response Exchange (SHARE) Network—which would empower all participants to be full and active partners in protecting our security, and which would be governed by guidelines designed to protect our liberties.

Zoë Baird James Barksdale

Acknowledgments

The Task Force issued its first report in October 2002. This initial report was seen as a useful contribution by a broad cross section of those concerned with the new environment for America's homeland security. We decided to continue our work and developed an agenda for further action. The Task Force began its second year in March 2003. Renewing its commitment to providing the government with practical recommendations informed by diverse perspectives, the Task Force organized several areas of work, leading up to a summer plenary meeting. At that meeting, Task Force members agreed that we should issue a second report by the end of this year, given the urgent need for our government to improve its ability to use information to protect our nation.

We organized two Working Groups. Working Group I, ably led by Tara Lemmey and Bill Crowell, focused on how to construct a network for sharing and analyzing information among governmental entities at all levels and relevant private sector organizations. This Working Group's paper is included in Part Two of our report. Working Group II, equally ably led by Gilman Louie and Jim Steinberg, addressed the issue of how our government can more effectively utilize privately held data while protecting privacy and other civil liberties. This Working Group's paper, too, is in Part Two. We also convened a smaller subgroup, thoughtfully led by Amitai Etzioni, to consider the crosscutting issue of how to make forms of identification more reliable while protecting civil liberties. This paper is in Part Three.

The papers developed by the Working Groups and the subgroup, along with many other papers, including the selection of appendices found in Part Three, informed the Task Force's discussions, which took place in two plenary sessions, numerous meetings, and email and telephone exchanges. These discussions led to the report of the Task Force as a whole, found in Part One.

We thank the leaders of the Working Groups and the subgroup for their devotion to our work and their high standards regarding what we together could achieve. On technology issues, Tara Lemmey, Gilman Louie, and Task Force associate Jeff Jonas made our work a central part of their daily lives, tirelessly researching and vetting papers, drafting sections of our report, and informing our efforts with their knowledge and creativity. On policy and governmental issues, Jim Steinberg, Amitai Etzioni, Stewart Baker, and Task Force associates Mary DeRosa, Mary McCarthy, Winston Wiley, Terry Maynard, and Jim Dempsey developed innovative approaches, careful research, compelling ideas, and drafts of sections of the report. These people and the many other members and associates of the Task Force put in dozens of hours melding their expertise into a common understanding of the interplay between technology and policy. They attended meetings, reviewed documents, consulted one another via email or telephone, and sought advice and information from other professionals. We cannot adequately thank them here for their wisdom and dedication.

Eric Benhamou, among his other contributions, helped us chair our principal plenary meeting. Lauren Hall and a team at Microsoft—drawn together by Task Force member Craig Mundie—assisted Tara Lemmey in helping us think through how to create visual representations of our recommendations.

Stefaan Verhulst, Lara Flint, and Tanvi Madan provided valuable research assistance. Todd Glass handled our public education and outreach, and Karen Thomas provided an excellent platform on the Internet for the Task Force's work. Sharon Lucius and Caroline McCarus were important members of the team, as were Paulette Layton, Nancy Boursiquot, Jennifer Obriski, Brendan Lavy, and Linda Hutchins, who worked day to day to support the Task Force members and staff. Karen Byers provided sound financial management. Mila Drumke tirelessly edited our final product.

Finally, none of this would have been possible without the dedication and hard work of our Executive Director, Michael Vatis, and Associate Director, Mary McKinley. Michael thoughtfully managed the many different areas of activity that constituted this project and ultimately brought them together to form a valuable whole in this report. Mary again operated the daily activities, keeping them on course with professionalism and experience.

Zoë Baird James Barksdale

PART ONE
The Task Force Report

Achieving a networked community for homeland security

In October 2002, the Markle Foundation Task Force issued its first report, *Protecting America's Freedom in the Information Age*. In that report, we expressed our belief that the threats to America at home from terrorism and weapons of mass destruction could be met only if we developed first-rate information collection, analysis, communications, and sharing. The nation could never sufficiently harden all potential targets against attack, so the government should instead develop the means to obtain advance warning of terrorist intentions through better intelligence, and use that intelligence to interdict terrorist plans and focus our protection resources on the most likely terrorist targets. It should also use information to enhance our response capabilities. We proposed a national strategy for using information and information technology in a robust decentralized network and for strengthening the processes for collecting data and turning it into actionable information. These new capabilities, we stated, could be achieved in a manner that protects our rights to privacy and other traditional civil liberties. In fact, any government undertaking to build such an information network would not be sustainable if the government did not build public trust by embedding protection of well-established civil liberties throughout that system.

We expressed our belief that our nation must capitalize on its leadership in information technology and on our citizenry's commitment to have both security and civil liberties. Thus, the recommendations in our first report provided a road map for the development of new networks and relationships among government agencies and officials at all levels. We also provided a framework for considering how the government might make most effective use of data residing in the private sector, while preserving liberties and avoiding the imposition of undue costs on businesses.

In our first report, we emphasized the need for a next-generation homeland security information network that would empower local participants to contribute, access, use, and analyze data, while also allowing them to identify, access, communicate with, and assemble other participants in both the public and private sectors. We argued against a centralized mainframe system in Washington, DC, and stated that "most of the real frontlines of homeland security are outside of Washington, DC," and that "likely terrorists are often encountered, and the targets they might attack are protected, by local officials" (page 10). In addition, we said that "the government will need access to public and private sector data for national security" and called for the Department of Homeland Security (DHS) to "develop innovative service-delivery models for using information held within and outside government" (page 37).

In this report, we reaffirm the principles of our first report and offer greater detail on how we believe the government should create networks for information collection, sharing, analysis, and use across federal, state, and local agencies and the private sector, while preserving—and even enhancing—privacy and other civil liberties. The network we envision consists not just of the technological architecture, but also of the people, processes, and information that must go hand-in-hand with the technology, and the rules that should govern how all of these elements interact. We repeat our call for the President to issue guidelines for government collection and use of information. As we said in our first report, "Only the President can establish and be accountable for the proper balance between development of domestic intelligence and preservation of liberty" (page 2). And only with such guidelines and attendant public discussion can the government hope to engender and maintain the trust of the people in its efforts, which is vital to implementing the network we envision.

Assessment of government progress toward a trusted, decentralized network

Since we issued our first report, the federal government has made some progress in fostering the development of the network we envisioned. Both the Executive Branch and Congress appear to have a greater understanding now of the need for more information sharing and for networks that break down agency "stovepipes."[1]

[1] The new *Attorney General's Guidelines for FBI National Security Investigations and Foreign Intelligence Collection* (effective 31 Oct. 2003), for instance, stress that "information should be shared as consistently and fully as possible among agencies with relevant responsibilities to protect the United States and its people from terrorism and other threats to the national security, except as limited by specific constraints on such sharing" (page 24). Available at http://www.cdt.org/security/usapatriot/031031nsiguidelines.pdf (last visited 11 Nov. 2003). These guidelines move the government forward on considering how it will share information, and we intend to look more closely at this issue. However, the guidelines do not appear to reflect the comprehensive effort that we encourage to openly develop policy principles and tools that can help implement guidelines for use when both privacy and security interests are implicated by the collection of information domestically.

But our nation's efforts continue to suffer from the absence of the national vision and public discussion called for in our first report. Progress thus has been ad hoc and sporadic at best. The government also has not yet developed guidelines to govern the collection, use, and retention of information in conducting the war on terrorism and issued them as a presidential directive. As a result, each agency is making its own decisions, and this is undermining public confidence—which in the long run limits the prospects for successful implementation of the necessary information-gathering and analysis efforts. It is critically important that guidelines be established before another major terrorist incident occurs. If public debate were to take place in the shadow of another major national tragedy, it could lead to rushed and poorly conceived initiatives that not only fail to solve the underlying problems, but also have a detrimental impact on civil liberties.

Information sharing and analysis

Steps have been taken at the federal, state, and local levels to broaden the sharing of terrorist-threat data among government agencies at all levels and to improve analysis of terrorism-related information.[2] To date, however, the government is still a long way from the creation of the dynamic, distributed network for sharing and analysis that we envision. The sharing of terrorist-related information between relevant agencies at different levels of government has been only marginally improved in the last year, and remains haphazard and still overly dependent on the ad hoc "sneaker net" of personal relations among known colleagues. It is not the result of a carefully considered network architecture that optimizes the abilities of all of the players.

At the federal level, the President announced, in January 2003, the creation of the Terrorist Threat Integration Center (TTIC), an interagency center created by the CIA and the FBI, with participation of the DHS, the Department of State, the Department of Defense (DoD), and the intelligence community. The TTIC reports to the Director of Central Intelligence. The Executive Branch established the TTIC to perform many of the analytical functions that Congress had assigned to (and that our initial report recommended be performed by) the DHS and its Intelligence Analysis and Infrastructure Protection Directorate. Thus, the White House announced that the TTIC would close the gap between analysis of foreign and domestic intelligence on terrorism. According to the White House, the center will do the following:

- *Optimize use of terrorist threat-related information, expertise, and capabilities to conduct threat analysis and inform collection strategies,*
- *Create a structure that ensures information sharing across agency lines,*
- *Integrate terrorist-related information collected domestically and abroad in order to form the most comprehensive possible threat picture,*
- *Be responsible and accountable for providing terrorist threat assessments for our national leadership.*[3]

While the TTIC's personnel are working hard to build an integrated analytical capability and are apparently considering innovative ways to analyze and disseminate information,[4] the very fact of the TTIC's creation has caused confusion within the federal government and

[2] One example is the Antiterrorism Information Exchange (ATIX), a network being developed by the Justice Department and the FBI to provide law enforcement agencies and public safety, infrastructure, and homeland security groups access to "Sensitive But Unclassified" homeland security information. ATIX is also intended to serve as a means to deliver security alerts to public officials and private sector groups and to allow users to create collaborative bulletin boards where they can exchange information. See Wilson P. Dizard III, "First Responders Get Homeland Security Network," 22 *Government Computer News 9* (28 Apr. 2003), at http://www.gcn.com/22_9/news/21878-1.html (last visited 11 Nov. 2003).

[3] White House Fact Sheet: Strengthening Intelligence to Protect America (Jan. 2003), available at http://www.whitehouse.gov/news/releases/2003/01/20030128-12.html. For more information on how the TTIC was formed and is intended to operate, see testimony (as prepared for delivery) by Winston P. Wiley, Chair, Senior Steering Group, Terrorist Threat Integration Center, and Associate Director of Central Intelligence for Homeland Security, before the Senate Governmental Affairs Committee (26 February 2003), available at http://www.cia.gov/cia/public_affairs/speeches/2003/wiley_speech_02262003.html (last visited 8 Nov. 2003).

[4] For instance, over 2,000 government users have access to "a TTIC-sponsored classified website providing terrorism-related information. This website…is currently being updated to include expanded need-to-know access with rich content available at varying classification levels, from "Top Secret" to "Sensitive But Unclassified"…[and to] enable users to search across disparate data sets in many different ways. The website will increasingly include products tailored for the needs of state and local officials, as well as private industry" for dissemination by the DHS and the FBI. Statement for the Record of John O. Brennan, Director, TTIC, on "The Terrorist Threat Integration Center and its Relationship with the Departments of Justice and Homeland Security," before the House Judiciary Comm. and the House Select Comm. on Homeland Security (22 July 2003), available at http://hsc.house.gov/files/Testimony%20Brennan.doc (last visited 12 Nov. 2003).

among state and local governments about the respective roles of the TTIC and the DHS.

Moreover, we understand that the TTIC is presently focused mainly on only one part of its assigned mission—"providing terrorist threat assessments [such as the President's Terrorism Threat Report (PTTR)] for our national leadership," including the President, the National Security Council, and other senior officials in Washington, DC. While this is obviously an important function, the TTIC's almost single-minded focus on this one aspect of its mission has prevented it from addressing the urgent intelligence needs of operational entities throughout the government and serving as a key locus for intelligence fusion and sharing. As an intelligence community official told one Task Force member, "As good as the PTTR is, it won't save anyone's life."

Meanwhile, the DHS also does not appear to have taken the necessary steps to build the communications and sharing network required to deal with the threat, or to begin producing regular, actionable intelligence products for other agencies. Indeed, the DHS has yet to articulate a vision of how it will link federal, state, and local agencies in a communications and sharing network, or what its role will be with respect to the TTIC and other federal agencies. The DHS instead seems to be focused on building a new information-technology infrastructure to support and unify its 22 components.[5] This is an important step, but one that should be grounded in a plan for the whole system.

Moreover, neither the TTIC nor the DHS has gotten very far in putting in place the necessary staff or framework for analyzing information and sharing it broadly among the relevant federal, state, and local agencies. Government at the federal level thus remains very much in need of an overarching decentralized framework for building an information-sharing and analysis network.

Many state and local governments have grown increasingly frustrated at the perceived lack of progress at the federal level in sharing information, at the dearth of actionable intelligence coming from federal sources, and at the lack of transparency and feedback regarding how the information they provide is being utilized. Some have responded by developing their own ideas for information-sharing and analysis networks.[6] However, without an overall framework that links regional or local networks with one another and with federal entities, the full potential of state and local governments will never be realized. Moreover, without broad national agreement on how state and local government programs should function, and when and how they should access and use private sector data, they run the risk of being shut down in the same way federal programs that would have used private sector data have been.[7]

In August 2003, the General Accounting Office (GAO) issued a study that found that the poor coordination of information-sharing efforts might cause critical clues of impending terrorist attacks to go unnoticed.[8] Although the Homeland Security Act of 2002 requires the DHS to share information with state and local authorities, representatives from states and cities told the GAO that the current system is close to failing. One of the major obstacles cited in the GAO report is the federal government's belief that the fight against terrorism remains its responsibility alone. In addition, GAO investigators said many federal officials expressed concern about sharing national-level intelligence information with state and local agencies.

[5] See, for example, Dibya Sarkar, "DHS Still Working on Info-Sharing Plans," FCW.com (7 Nov. 2003), available at http://www.fcw.com/geb/articles/2003/1103/web-dhs-11-07-03.asp (last visited 11 Nov. 2003).

[6] For example, the 10 northeastern states, from Delaware to Maine, have formed a consortium to combine their homeland security efforts and develop information-sharing strategies. See Testimony of James Kallstrom, senior advisor to New York Governor George Pataki for Counter Terrorism, before the House Select Comm. on Homeland Security, Subcomm. on Intelligence and Counterterrorism (24 July 2003) at pages 4-5, available at http://hsc.house.gov/files/Testimony%20Kallstrom.doc (last visited 11 Nov. 2003). In addition, Pennsylvania, New York City, and Washington, DC, have formed a model project—primarily with local funding—that would link existing law enforcement, public safety, and justice systems across jurisdictions to provide real-time data sharing over the Internet. See "The Shield Pilot," available at http://www.search.org/integration/pdf/ShieldPilot.pdf (last visited 11 Nov. 2003).

[7] For example, several states now participate in the Multistate Anti-Terrorism Information Exchange (MATRIX) project, a data-mining effort run by a private company for the participating states and aided by the federal government. MATRIX, started by police in Florida, combines law enforcement and court records with commercially available information about individuals, and purportedly allows officials to look for patterns and linkages among people. See Robert O'Harrow, Jr., "U.S. Backs Florida's New Counterterrorism Database," Washington Post, page A1 (6 Aug. 2003, available at http://www.washingtonpost.com/ac2/wp-dyn/A21872-2003Aug5?language=printer (last visited 4 Nov. 2003). Privacy concerns have reportedly caused several states to reconsider their initial decision to participate and have prompted criticism and questions from civil liberties groups. See, for example, Kristen Wyatt, "Georgia Decides Against Crime Database," Associated Press (21 Oct. 2003) available at http://www.bayarea.com/mld/mercurynews/business/7068004.htm (last visited 8 Nov. 2003); American Civil Liberties Union, "What is the Matrix? ACLU Seeks Answers on New State-Run Surveillance Program," (30 Oct. 2003), available at http://www.aclu.org/Privacy/Privacy.cfm?ID=14257&c=130 (last visited 4 Nov. 2003).

[8] See GAO, Homeland Security Highlights, "Efforts to Improve Information Sharing Need to Be Strengthened" (Aug. 2003), available at http://www.gao.gov/highlights/d03760high.pdf (last visited 21 Nov. 2003).

Furthermore, since many states and localities are presently enduring serious financial difficulties, most of these efforts are in need of federal funding to make any significant advances. A 50-state study released by the U.S. Conference of Mayors in September concluded, for instance, that 90 percent of cities have not received funds from the country's largest federal homeland security program, designed to assist local officials, police, fire chiefs, and other first responders to prepare for terrorist attacks.[9] Though we do not believe that all cities need significant federal funding—because all are not equally likely to be terrorist targets—this is nevertheless a telling statistic. We must ensure that local officials are neither overwhelmed by, nor without adequate resources to deal with, what is expected of them, because they are important players in the overall system. Moreover, their participation is important for creating deep and lasting public trust.

Another recommendation of our first report was that the government create virtual consolidated watch lists to allow agencies to check individual names against the many different lists maintained by various parts of the federal government. We also called for guidelines and procedures that would determine how individuals get put on a list and how they can be removed from it. Some progress is apparently being made here. In particular, the government is creating the Terrorist Screening Center (due to become operational on December 1, 2003), which is supposed to consolidate the many existing watch lists.[10] But it remains to be seen how successful this center will be in practice. Also, to date, no government-wide guidelines have been issued concerning how individuals get placed on—and removed from—a watch list; how accuracy is maintained and errors are corrected across lists; or on how information on the lists is shared among agencies and with private companies.[11]

Utilizing privately held data

Government access to, and use of, privately held data remains a vexing problem. On the one hand, as we pointed out in our first report, there is a great deal of readily available private sector data that can expose patterns, identify terrorists, and save lives. In our initial report, for example, we showed how the September 11 terrorists could have been identified from airline reservation systems and searches of public-record data starting with the information that two individuals on terrorist watch lists had bought airline tickets using their real names (page 28). These individuals were linked by common past addresses, common phone numbers, and frequent-flyer numbers. On the other hand, government efforts to collect information on Americans without a demonstrated, compelling government need have been met with outcries of invasion of privacy and repeatedly have been shut down.[12]

One major development since our first report has been the controversy over the Defense Department's Terrorist Information Awareness (formerly known as Total Information Awareness) program (TIA). The aim of this initiative, created by the Defense Advanced Research Projects Agency (DARPA), was to develop information-analysis and collaboration tools to enhance the government's ability to detect terrorist activity. Another aim was to develop software that would allow government officials to search for patterns across databases of transactional records (medical, financial, educational, travel, immigration, communications, etc.) in order to detect potential terrorist activity. The TIA faced vocal opposition from the public and Congress, in part because of shifting explanations of how the TIA's proposed technology (described as, among other things, "data-mining" or "knowledge discovery" tools) would operate. For instance, the TIA left ambiguous

[9] See "The United States Conference of Mayors, U.S. Conference of Mayors Announces: 90 Percent of Cities Left Empty-Handed Without Funds from Largest Federal Homeland Security Program" (17 Sept. 2003), available at http://www.usmayors.org/uscm/news/press_releases/documents/homelandfunding_091703.pdf (last visited 21 Nov. 2003).

[10] See Homeland Security Presidential Directive/Hspd 6 (16 Sept. 2003), available at http://www.whitehouse.gov/news/releases/2003/09/20030916-5.html (last visited 7 Nov. 2003).

[11] In April 2003, the GAO issued a report on federal watch lists, finding that 9 agencies maintained 12 different watch lists, that the lists contained overlapping but different information, and that the agencies had different policies governing when and how information on the lists was shared with others. It also found that sharing was constrained by the watch lists' differing technological architectures. GAO, "Information Technology: Terrorist Watch Lists Should Be Consolidated To Promote Better Integration and Sharing" (Apr. 2003), available at http://www.gao.gov/new.items/d03322.pdf (last visited 9 Nov. 2003).

[12] In 2002, for example, Congress prohibited implementation of the Justice Department's "Operation TIPS" (Terrorism Information and Prevention System) after the media, public, and members of Congress expressed concern about infringements of privacy. Homeland Security Act of 2002, 6 USC 142, § 880. TIPS would have given workers in the transportation, trucking, shipping, maritime, and mass-transit industries (and, originally, mail carriers and utility workers as well) a formal mechanism for reporting and sharing information about suspicious, publicly observable activity possibly related to terrorism.

In 1999, the Federal Deposit Insurance Corporation withdrew a proposed "Know Your Customer" regulation that would have required certain state banks to develop programs to determine the identity of their customers, the customers' sources of funds, and their "normal and expected transactions"; monitor their account activity to identify transactions inconsistent with those normal and expected transactions; and report any suspicious activities to the government. The regulation, which was intended to "protect the integrity and reputation of the financial services industry" and assist in combating money-laundering and other illegal activities, was withdrawn in the face of widespread opposition from industry and the public. "The overwhelming majority of commenters were individual, private citizens who voiced very strong opposition to the proposal as an invasion of personal privacy." FDIC, "Minimum Security Devices and Procedures and Bank Secrecy Act Compliance," 12 C.F.R. Part 326, Fed. Reg. vol. 64 no. 59, p. 14845 (29 Mar. 1999), available at http://www.fdic.gov/news/news/financial/1999/FIL9934b.pdf (last visited 11 Nov. 2003).

whether its technology would be used to search transactional data only for information about specific subjects of terrorism investigations or to find suspicious patterns that matched analysts' hypotheses of how potential terrorists might launch an attack, and thereby identify individuals requiring further scrutiny. Similarly, it was never clear whether the TIA envisioned technology that would allow the government to aggregate private sector data into one centralized government database, or technology that would allow the government to search across private sector databases while leaving the data in private hands.[13] Moreover, the TIA's defenders never adequately explained the extent to which transactional data of U.S. citizens would be searched using the agency's technology. In January 2003, Congress barred funding for domestic deployment of the TIA but allowed research to go forward.[14] Ultimately, in September 2003, Congress eliminated all funding for the TIA program and "any successor program."[15]

> Innovation in technology is an important part of our nation's competitive edge against terrorist organizations and the states that back them.

We are disappointed that Congress found it necessary to ban research and development of technologies that would make use of privately held data. Innovation in technology is an important part of our nation's competitive edge against terrorist organizations and the states that back them. Had the government, in developing the TIA, formulated policy principles and guidelines on the research and use of such technologies to access privately held data—and engaged in a public discussion of those policies—it would not have become so mired in the controversy that resulted in the banning of research by Congress. Policy guidelines like these are meant to empower government officials as well as limit them, and Congress and the Executive Branch should share a common commitment to both objectives.

Yet, even though the TIA has been shut down, other still-extant governmental efforts—both research and operational activities—raise many of the same issues. For instance, the National Security Agency's (NSA) Advanced Research and Development Activity (ARDA) is pursuing research programs in "Novel Intelligence from Massive Data" (aimed at "focusing analytic attention on the most critical information found within massive data—information that indicates the potential for strategic surprise"[16]) and "Information Exploitation" ("the process of extracting, synthesizing, and/or presenting relevant information from vast repositories of raw and structured data"[17]). Meanwhile, the CIA reportedly is implementing a data-mining program called Quantum Leap that "enables an analyst to get quick access to all the information available—classified and unclassified—about virtually anyone."[18]

Similarly, if implemented, the Transportation Security Administration's (TSA) Computer Assisted Passenger Prescreening System (CAPPS II) would check passenger-provided data against commercial databases, government databases, and a watch list of suspected terrorists and people wanted for violent crimes to determine if specific passengers should receive further checkpoint scrutiny or be barred from boarding planes altogether.[19] However, following the revelation that JetBlue Airways turned over to a DoD

[13] Compare, for instance, "Overview of the Information Awareness Office," remarks as prepared for delivery by Dr. John Poindexter, Director, DARPA Information Awareness Office, at DARPATech 2002 Conference, Anaheim, CA. (2 Aug. 2002) ("One of the significant new data sources that needs to be mined to discover and track terrorists is the transaction space. If terrorist organizations are going to plan and execute attacks against the United States, their people must engage in transactions and they will leave signatures in this information space…*The relevant information extracted from this data must be made available in large-scale repositories with enhanced semantic content for easy analysis to accomplish this task.*" [emphasis added]) with Terrorist Information Awareness Program, Guide to the Report to Congress (20 May 2003) ("the TIA Program is *not attempting to create or access a centralized database that will store information gathered from public or privately held databases*" [emphasis added]).

[14] Omnibus Appropriations Act for Fiscal Year 2003, H.R.J. Res. 2, Amend. 59, 108th Cong. (2003).

[15] H.R. Conf. Rep. No. 108-283 (2003) (Conference report on H.R. 2658, Department of Defense Appropriations Act, 2004), (24 Sept. 2003), available at http://thomas.loc.gov/cgi-bin/query/R?r108:FLD001:H08501 (last visited 21 Nov. 2003).

[16] Advanced Research and Development Activity, Novel Intelligence from Massive Data," at http://www.ic-arda.org/Novel_Intelligence/ (last visited 31 Oct. 2003).

[17] "Advanced Research and Development Activity, Information Exploitation Thrust," at http://www.ic-arda.org/InfoExploit/index.html (last visited 31 Oct. 2003).

[18] See Bill Powell, "Inside the CIA," *Fortune* (29 Sept. 2003), available at http://www.fortune.com/fortune/articles/0,15114,490641,00.html (last visited 7 Nov. 2003). The CIA's deputy chief information officer reportedly stated that the program's technology is "so powerful it's scary," and that in the wrong hands, it "could be Big Brother." The MATRIX program, discussed in footnote 7, is another example of an ongoing data-mining effort.

[19] According to testimony from DHS officials: "CAPPS II is intended to identify terrorists and other high-risk individuals before they board commercial airplanes. CAPPS II will conduct a risk assessment of each passenger using national security information and information provided by passengers during the reservation process—including name, date of birth, home address, and home phone number, and provide a risk score to the TSA. The risk score includes an authentication score provided by running passenger name record (PNR) data against commercial databases to indicate a confidence level in each passenger's identity. CAPPS II will be a threat-based system under the direct control of the federal government and will represent a major improvement over the decentralized, airline-controlled system currently in place." Testimony of Stephen J. McHale, Deputy Administrator of the TSA, et al., before the Senate Comm. on Commerce, Science, and Transportation (5 Nov. 2003), available at http://www.senate.gov/~commerce/hearings/testimony.cfm?id=985&wit_id=2785 (last visited 8 Nov. 2003).

contractor the names and addresses of 5 million passengers, which were then used for a data-mining study on airline-passenger risk assessment, Congress has stalled the implementation of CAPPS II pending a GAO review of the program's effectiveness and potential effects on privacy.[20]

Thus, while various government agencies pursue efforts to utilize privately held data, those efforts continue to provoke controversy because of the lack of a systematic effort to consider the privacy implications of the proposed programs or to develop an overall policy framework that would govern the deployment of new technologies. Here, too, then, the principle of establishing clear policies and guidelines for the acquisition, use, and retention of private data laid out in our first report has not been adequately implemented.

Building a networked community for homeland security

We developed this second report through a rich, multilayered process that involved a diverse group of Task Force members with vast experience in federal, state, and local government; intelligence; law enforcement; defense; the technology industry; computer science; sociology; law; medicine; and privacy protection. We consulted with government officials in Washington, DC, and across the country so that we could better understand the current state of governmental activity—the successes achieved and barriers encountered to date. We also consulted broadly with technology experts and businesses to learn the state of technology development.

We approached our task by considering concrete situations in which the government would need to obtain, analyze, share, and act on certain information in order to learn about and prevent terrorist activity. This process of scenario-based envisioning allowed us to evaluate the following: (1.) how such information would be analyzed and shared today; (2.) where current roadblocks exist that prevent or impede necessary information sharing; (3.) what additional players should be getting the information in order to activate all the sensors in the system and increase the intake of relevant information; (4.) what procedures should be developed so that sensitive information (such as intelligence sources and methods) can be protected from unauthorized disclosure but actionable information can be routinely and quickly disseminated throughout the network; (5.) what process changes and new technologies can enhance analysis and information sharing; and (6.) how the government can avoid flooding the system with noise while ensuring that potential signals of terrorist activity are distinguished from the noise and shared widely.

To this end, we developed a set of information vignettes (see Appendix D) that served to inform the Task Force's discussions and our ultimate recommendations. Our goal was to discover where the present homeland security initiatives are optimized to achieve the dynamic and decentralized network required to take on the challenge of distributed and complex threats, and where more work is needed. We looked at whether existing networks were created to maximize the potential contribution of all of the participants, including those at the state and local levels. And we sought to identify how the government could utilize the enormous volumes of potentially significant data in private hands while protecting precious liberty interests. This process has grounded our recommendations in reality, thus allowing us to see more clearly the very real impediments to change and to make recommendations that can help the government work through those obstacles more effectively.

Vision and objectives

The networked community that we envision would protect national security by drawing on the best talent and technology available and by fostering a robust sharing of information and ideas. Collectively, this community could provide the public with confidence that the government was doing everything reasonably possible to prevent and respond to terrorist attacks on the homeland. The network we recommend would be guided by a practical set of policy guidelines that would simultaneously empower and constrain government officials by making clear what collection, analysis, sharing, and uses of information were permissible and what were not. And it would focus on eliminating the gaps between government agencies. All players in this network—including those at the edges—would be able to create and share actionable and relevant information. The focus of the network itself would be to get information into the hands of people who could analyze and act on it, and to leverage information from private data holders within a system of rules and guidelines. The objective of this network would be to enhance the government's "sensemaking" ability—that is, its ability to discern indicators of terrorist activity amid overwhelming

[20] See Department of Homeland Security Appropriations Act, 2004, 108 Pub. 90 (1 Oct. 2003), Title VI, § 519; Judi Hasson, "Congress Demands Study of CAPPS II," *Federal Computer Week* (26 Sept. 2003), available at http://www.fcw.com/fcw/articles/2003/0922/web-capps-09-26-03.asp (last visited 21 Nov. 2003).

amounts of information, and to create more time for all of the actors to make decisions and to prevent or respond to terrorist acts more effectively.

This is government acting in new ways, to face new threats. And while such change is necessary, it must be accomplished while engendering the people's trust that privacy and other civil liberties are being protected, that businesses are not being unduly burdened with requests for extraneous or useless information, that taxpayer money is being well spent, and that, ultimately, the network will be effective in protecting our security.

Building the networked community presents enormous challenges. It requires changes in policies, processes, and the use of technology. And it requires fundamental changes in the harder intangibles of cultures and attitudes, which have impeded the creation of the sort of network we envision. Leadership is emerging from all levels of government and from many places in the private sector. What is needed now is a plan to accelerate these efforts, and public debate and consensus on the goals. This report attempts to contribute to paving that path.

Using the principles outlined in this and our previous report, and building on the information-sharing initiatives around the country, the federal government should create an interagency, public-private group, led by the DHS and comprising representatives of all the relevant network players, to develop a national strategy and architecture for the homeland security network. Because of the daunting political, organizational, and technical challenges, it is impossible to conceive of designing and building this network all at once. It will be necessary to grow capabilities in pieces, building on existing systems and incorporating new systems and technologies over time. To do so will require an architecture that is flexible and adaptable. Such an effort could render a working plan within a year, one that could guide investment and network development for both the short and long term.

Throughout this report, we recommend actions the government should take to begin creating the System-wide Homeland Analysis and Resource Exchange (SHARE) Network we envision. In Exhibit A, we summarize the principal steps that should be taken by the federal government in the near term to begin this urgent undertaking.

Exhibit A

ACTION PLAN FOR FEDERAL GOVERNMENT DEVELOPMENT OF THE SHARE NETWORK

The President should issue an Executive Order that does the following:

1. Sets the goal of creating a decentralized network along the lines set out in this report

2. Sets forth specific and clear objectives for improved analysis and information sharing, which each federal agency should be required to meet by December 31, 2004

3. Establishes guidelines for agencies' collection, use, and dissemination of information, including private sector information

4. Establishes a process for Executive Branch review of agencies' performance in improving analysis, information sharing, and utilization of private sector information, to take place after December 31, 2004

5. Designates the DHS as the lead agency of an interagency, public-private process to establish the concept of operations for the network, directs other agencies to offer their full assistance and cooperation, and establishes a time frame for implementation

6. Clarifies the respective roles of the DHS, the TTIC, and other federal agencies in information sharing and analysis

The President should also issue a second Executive Order or other directive that does the following:

1. Establishes guidelines governing the authority of the TTIC and other intelligence, defense, and security agencies to receive, retain, and disseminate information gathered in the U.S. about U.S. persons

2. Establishes guidelines governing intelligence agencies' ability to set requirements for (or "task") domestic collection of information

3. Creates within the TTIC appropriate institutional mechanisms to safeguard privacy and other civil liberties

   The contents of the Executive Order should be unclassified to the maximum extent possible and put out for notice and comment. In addition, the President should consider introducing legislation to codify the appropriate scope of the TTIC's use and dissemination of information about U.S. persons.

The DHS should do the following:

1. Convene an interagency, public-private group to design a strategy and concept of operations for the decentralized network we describe, which should render a working plan within a year

2. Work with state, local, and private sector entities to create decentralized analytical centers, foster their ability to communicate with other players in the network, and establish standards for digitization, retention, and communication of information

3. Establish clear mechanisms for responding to requests for threat and vulnerability information from local officials

4. Establish a process for ensuring that as much information as possible is being shared among network entities, including a dispute-resolution mechanism to resolve disagreements among agencies about how much information can be shared

5. Establish a process for overseeing federal agency development and implementation of guidelines governing the acquisition, use, retention, and dissemination of private sector information and the creation of methods for ensuring oversight and accountability

6. Work with state, local, and private sector entities to institute information-sharing and analysis objectives for these entities, and establish a process with them for jointly evaluating their performance after December 31, 2004, and thereafter on an ongoing basis

The FBI should do the following:

1. Establish mechanisms for sharing information with state and local law enforcement agencies, and for encouraging those agencies to share directly with other players in the network

2. Establish clear mechanisms for responding to requests for threat and vulnerability information from local officials

All government agencies with homeland security intelligence responsibilities should do the following:

1. Set up mechanisms to produce more information that can be readily disseminated to other players in the network, including unclassified information

2. Identify specific categories of private sector information they need, using a scenario-driven process that considers the types of situations they might confront in investigating or seeking to uncover terrorist activity

3. Institute guidelines governing the acquisition, use, retention, and dissemination of private sector information, and establish methods to ensure oversight and accountability

Congress should do the following:

Undertake to review the performance of federal agencies in improving analysis and information sharing along the lines set out in this report, and in utilizing private sector information while protecting civil liberties; this review should take place after December 31, 2004

Closing the gaps between agencies

One of the biggest challenges we face is the reduction of information gaps that exist between our various federal and state agencies, between intelligence and law enforcement, and between government in general and the private sector. In 1947, President Truman created the Central Intelligence Agency (CIA) to help eliminate the intelligence gaps that existed between government agencies before World War II. The National Security Act of 1947 charged the CIA with coordinating our nation's intelligence activities and correlating, evaluating, and disseminating intelligence.[21]

Every day our intelligence and law enforcement agencies, health care providers, private companies, and numerous other players receive information that might be relevant to uncovering a terrorist plot and preventing an attack.

Today, if anything, the gaps between different agencies are even broader and more numerous than in the Cold War years. This is particularly true in the context of counterterrorism, where important information or analytical ability resides not just in the 14 intelligence components of the federal government and federal law enforcement and security agencies, but also with the 17,784 state and local law enforcement agencies,[22] 30,020 fire departments,[23] 5,801 hospitals[24] and the millions of first responders who are on the frontlines of the homeland security effort. Add to this the thousands of private owners and operators of critical infrastructures, who are responsible for protecting potential targets of terrorist attacks, and the many more private companies that may have information in their databases that could lead to the prevention of terrorist activity. Communication, collaboration, and sharing across the gaps between and among these actors are critical to countering terrorism because we cannot predict where the first sign of a potential terrorist threat will come from—a communications intercept from the National Security Agency (NSA), a human source of the CIA or the FBI, an investigation by a local police department, or an observation by an alert private security guard or emergency room nurse.

The decentralized nature of the terrorist threat thus leads to exponentially more—and widely scattered—information to process and share. The reality is that every hour of every day, our intelligence and law enforcement agencies, health care providers, private companies, and numerous other players receive information that might be relevant to uncovering a terrorist plot and preventing an attack.

Attempting to centralize this information is not the answer because it would not link the information to the dispersed analytical capabilities of the network. Centralization could also lead to information becoming obsolete, since a centralized analytical entity would not have the ability to keep up-to-date much of the information collected from dispersed sources. But making all the information available to everyone in the network is not the answer either, because this could increase the threat to civil liberties, heighten the risk of a leak of sensitive information, cause uncoordinated action by different agencies, and simply overwhelm the recipients. Indeed, the sheer volume of data would create such a high degree of noise that it would be extremely difficult for analysts to make useful correlations or for local agencies to take meaningful protective action.

The network we envision therefore would enable participants to distinguish useful signals of potential terrorist activity from useless noise.[25] It would utilize the expertise of all the participants in the network and address their need to collect, update, and understand the information that is important to their primary functions, without flooding them with extraneous information they cannot use.

Moreover, the threat today requires unprecedented speed in the way we collect, share, and act on information. Unlike in the Cold War, we are not trying to discern the size or movements of distant armies or the goings-on

[21] See *Central Intelligence Agency Factbook on Intelligence 2002,* "The Genesis of the CIA," at http://www.cia.gov/cia/publications/facttell/genisis.html (last visited 31 Oct. 2003).

[22] See DOJ—Office of Justice Programs, Bureau of Justice Statistics, Law Enforcement Statistics Summary Findings (2000), available at http://www.ojp.usdoj.gov/bjs/lawenf.htm (last visited 12 Nov. 2003). There are approximately 800,000 full-time sworn law enforcement officers in the United States, including federal, state, and local agencies.

[23] See the DHS, FEMA, U.S. Fire Administration, Fire Data (2001), at http://www.usfa.fema.gov/inside-usfa/nfdc/nfdc-data9.shtm (last visited 12 Nov. 2003). There are 1,078,300 firefighters in the U.S.

[24] See Hospitalconnect.com, Advancing Health in America Resource Center, "Fast Facts on U.S. Hospitals from 'Hospital Statistics'" (10 Dec. 2002), available at http://www.hospitalconnect.com/aha/resource_center/fastfacts/fast_facts_US_hospitals.html (last visited 12 Nov. 2003).

[25] One potential model for thinking about this problem stems from an analogy to the functioning of our bodies' immune system. See Appendix C.

within a foreign government. Rather, we are trying to detect and thwart potentially imminent attacks that could take place at any time against what are often soft, civilian targets. What's more, the potential modes of attack—sniper attacks, suicide bombers, truck and car bombs, airline hijackings, weapons of mass destruction (chemical, biological, nuclear, and radiological) as well as mass disruption (cyber attacks)—are as varied as the imaginations of those who wish to do us harm. To detect, thwart, and respond to these types of threats, time is of the essence. And information needs to be tailored to facilitate decision-making and action at all levels—not only by the President, but also by police officers on the street.

Our Task Force's fundamental objective, then, is to identify the technological tools and infrastructure, the policies, and the processes necessary to link these different communities so that important information can be shared among the people who need it, and as rapidly as possible. Information sharing itself is not the goal; rather, it is the means by which we can maximize our ability to make sense of the information available. And it is also the means by which we can give all participants more time to make the right decisions and take more effective actions to prevent terrorist attacks.

For our envisioned network to work, rules are needed to define the following: (1.) how decisions are made about what information might be useful; (2.) who has what responsibilities for creating potentially useful information for the system; and (3.) who is authorized to have access to information and what uses of the information are permissible. There must also be rules to ensure oversight and public accountability.

Finally, guidelines covering how information is collected, used, and shared among the relevant actors are critical for several different but complementary reasons. First, they are vital to preventing the misuse of information that is gathered and shared in the network. A robust sharing of information must only be pursued consistent with civil liberties interests. Second, they are needed to empower government officials who, not knowing what the rules are, or fearful of public criticism, may refrain from taking legal and necessary action that might uncover a terrorist plot or thwart an attack. Third, guidelines are needed to ensure coordination by the participants in the network; if participants feel that they do not know what will happen with information they share with others, they will simply refrain from sharing regardless of how many directives are issued to mandate it. Finally, guidelines are needed to engender the public's trust in what the government is doing when it acts across the specifically defined boundaries of agencies in the network. That is, the public must understand, to the fullest extent possible, why the government needs information and what the government will do with it. The public must also have confidence that the information—and individuals' rights—will not be abused. It is not enough to write the code that operates the network; we must also write the code that governs the network.[26]

Scenario-based concept of operations

To build the network, we must start with a concept of operations that is based on a realistic understanding of the ways in which information comes into the system and the means employed for turning it into useful knowledge. The concept of operations must be scenario-based, and derived from the needs of the users across the network rather than from central authorities in Washington, DC. Thus, the government should generate realistic terrorist-threat scenarios that agencies might confront and then conduct exercises against them in order to understand the required information and communication flow, collection requirements, data sources, analytical requirements, decision processes, responder needs, and response timetables. The concept of operations that the government derives from these scenarios should be a living framework that is regularly updated based on new threat assessments and evolving user needs.

The network we envision would enable participants to distinguish useful signals of potential terrorist activity from useless noise.

The creation of a concept of operations would provide a better understanding of the gaps, single-point dependencies, and bottlenecks in the network architecture that exists today, and of what we need to do to move toward the network we envision. The concept of operations should define the most efficient and effective information workflow as well as the minimum acceptable bandwidth, connectivity, storage, and sharing requirements for every required connection path on the network. It should also allow individual agencies to better plan their internal information-technology acquisition plans and workflow-improvement programs. And it should help to establish

[26] See Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books, 1999), for a discussion of how code embedded in software and hardware and code found in laws and regulations enacted by the government both serve to regulate behavior on the Internet.

benchmarks for response time, data-sharing requirements, information-quality standards, responsibility, and authority for each node on the network. The information vignettes that our Task Force has developed (see Appendix D) are helpful for understanding how information flows today and how it needs to flow to optimize the capabilities of the players at the edges of the network.

The network should be decentralized, of course, but someone needs to be responsible for designing, building, and maintaining it. Consistent with the recommendations in our first report, we believe the DHS should take the lead by convening the relevant players, collaboratively designing the network, and securing the funding necessary to build it. We firmly believe that an agency with domestic jurisdiction, rather than a foreign intelligence agency, would be the most trusted federal entity to lead the creation of a network that is sustainable when privacy concerns are raised. Furthermore, Congress has established internal oversight mechanisms within the DHS, including a privacy officer and a civil rights and civil liberties officer, which would help give the DHS credibility.[27]

Designing a robust architecture for the future

Many agencies are enhancing their internal information and communications infrastructures in an effort to improve their ability to perform analytical tasks and respond to customers. Various agencies are adopting data and communication standards for sharing information. We see this as a positive first step, but more work needs to be done.

Agencies' current information-sharing efforts tend to focus on sharing data laterally and narrowly—federal agency to federal agency, law enforcement to law enforcement, and state government to state government. Work needs to be done to enable a network in which data moves across all the gaps and analysis occurs at multiple nodes rather than only in a few centralized locations. This section outlines some of the technical design elements of such a network, which we call the Systemwide Homeland Analysis and Resource Exchange (SHARE) Network.

Exhibit B

WEAKNESSES IN THE CURRENT INFORMATION-SHARING SYSTEM

There are multiple weaknesses in the current system that need to be fixed:

1. The system is susceptible to single points of failure for both analysis and communication of information.

2. The system is designed mainly to flow information up, to senior officials, and not down, to operational entities, and out, to the edges of the network.

3. The system does not adequately support real-time operations.

4. Many critical information repositories are not compatible with the analytic tools, and many still are air-gapped and not accessible online to analysts.

5. There is a lack of trust between federal, state, and local agencies.

6. It is difficult to sort the important signals of potential terrorist activity from the noise. Analytic tools are outdated and incapable of dealing with the current volume of data.

7. State, local, and commercial information is not well leveraged.

8. Many people are concerned about potential misuses of private information.

9. Information that is disseminated to first responders typically is not actionable. That is, relevant information does not enter into the everyday workflow of first responders.

10. Clear lines of authority and responsibilities for information sharing and analysis have not been established.

11. The system has not been well tested to see how it meets potential terrorist threats.

[27] See Homeland Security Act of 2002, 6 USC 142, §§ 221, 705 (2002).

Essential to an effective network that links disparate players is a set of directory services that helps each actor to find what he or she is looking for. We need directory services for information about locations (such as critical infrastructure assets, landmarks, and geographical references), people, organizations, terrorist methods, and other topics; and pointers to experts on various subjects. Directories would also help people find others working on similar problems. Fortunately, this is one of the areas in which existing technology can make a significant contribution. Automated directories with appropriate security and access controls can be deployed to solve these problems. These directories can be structured to give originators of information control over what can be shared and where it can be routed. Directories also can be updated automatically through real-time monitoring, synchronization, and profiling of the skills and interests of the network users.

Another important element of the network is the separation of data from data applications in order to foster interoperability. Data sets are often not directly interoperable because they are constructed for different purposes, use different standards, contain different terminology, and were not intended for integration with other data sets. But through the combination of data directories, metadata standards,[28] and commercially available exchange standards such as Extensible Markup Language (XML), a user can identify what data exists in other agencies and then contact those agencies to obtain the underlying data (assuming the user has the requisite authority). XML organizes information by allowing categories of data to be tagged with agreed upon names for each field, and thus can enable different organizations to share information more efficiently. To an extent, this process can also be automated: Software tools ("agent technologies"[29]) can be used to search for and identify data at the edges of the network, collecting only directory-level information without actually moving and consolidating the underlying data into a centralized database.

Directories can also enable ad hoc collaboration and sharing so that groups of players across levels of government can come together on matters of mutual interest and, by doing so, not only inform one another, but also collectively enhance the network analysts' ability to make sense of the huge volumes of data flowing through the system. To achieve this, the network needs to be a part of every user's workflow. For the reasons discussed above, a centralized, consolidated repository of information in Washington, DC, is impractical and vulnerable. We must therefore operate from a distributed model of interconnected databases that are made available to users through the directory services described above. The network, therefore, needs to take advantage of tools that can federate the data for analysis (that is, draw on appropriate information from various sources in the network).

Moreover, the network must allow users to move large amounts of data easily and in any form (such as written reports, photographs, video, and biometric data). Participants in the network must also be able to share across all levels of security, from "Top Secret/Code Word" to "Sensitive But Unclassified" and vice versa.

If we expect various agencies to share, then the network also needs to have strong data protection, including the ability to restrict access privileges so that data can be used only for a particular purpose, for a finite period of time, and by people with the necessary permissions.[30] Thus, the network also needs access control, authentication, and full auditing capability. Data protection is critical to preventing unauthorized disclosures and to preserving traditional civil liberties. A variety of new technologies has increased the capacity for online identification and authentication, which are prerequisites for providing permission to the right people to use the network for the right reasons. These technologies can enhance the security of the network, permit multiple users to interact and trade information in a trusting environment, and allow effective oversight of systems to prevent or detect misuse. These technologies include smart cards with embedded chips, tokens, biometrics, and security circuits. Many identification systems are being developed in conjunction with new data-anonymization technologies and strategies that can ensure that privacy objectives are achieved. Having these protections in place would not restrict information sharing. The protections would actually encourage sharing by engendering trust in the network and in the rules by which information is shared.

[28] "Metadata" is essentially data about data. A common example of metadata is a library catalog, which contains information (metadata) about publications (data).

[29] An agent is "a program that performs some information-gathering or processing task in the background," thus allowing the technology user to multitask. "Typically, an agent is given a very small and well-defined task." Webopedia, "Agent," available at http://www.webopedia.com/TERM/a/agent.html (last visited 3 Nov. 2003).

[30] In technical terms, a permission is "[a]n access privilege (for example, read, write, execute) associated with a file or directory. Depending on the operating system, each file may have different permissions for different kinds of access and different users or groups of users." InstantWeb Online Computing Dictionary, "Permission," available at http://www.instantweb.com/foldoc/foldoc.cgi?permission (last visited 4 Nov. 2003).

Exhibit C

AUTHENTICATION TECHNOLOGY

While authentication technologies are improving, no single approach can provide high assurance on its own. There are no smart cards or tokens that cannot be cracked, biometrics are not 100 percent reliable, and high-quality passwords are difficult to remember, manage, and enforce. With all of these technologies there are also people and process issues (such as enrollment procedures and audit trails) that can undermine their integrity. Therefore, a multifactored system is a preferable approach. Multifactor authentication typically combines a password with a token or smart card and can include other forms of authentication, such as biometrics, challenge codes and questions, and profile access matching. Authentication is strongest when a part of the information resides with the user, a part with the token or smart card, and a part in the network. Credit card companies, good users of multifactor authentication, rely on tokens (credit cards), passwords (PINs), challenge questions ("What's your mother's maiden name?"), and profile matching ("Is this a typical charge for this individual?").

Information rights management technologies (such as those being developed for the next generation of personal computers, operating systems, and document applications) can also protect data at the document level and may be key enablers for policy management systems.[31] Rules can be created about who can have access to particular documents and when documents expire. High-speed encrypted storage systems are also being developed to protect the data at rest.

Immutable audit (the ability to maintain tamper-resistant logs of user activity on the network) and tracking are also important capabilities for building trust. The ability to trace the origin of a piece of information, who has accessed it, and how it has been used facilitates accountability.[32] Audit technology also facilitates monitoring to improve security and to prevent inappropriate access and use. Security watch centers can employ tools that constantly monitor data use and notify watch officers of potential violations of policy and out-of-profile usage that might warrant a call to the user. These tools also allow the originator of information to track where the information is flowing, employing technologies similar to those used for tracking express mail.

Another benefit of having strong audit and tracking is the improved ability to understand the factual dependency of information. For example, if several pieces of analysis are dependent on a single data point, and that data point is later found to be wrong, the government can trace the noted dependencies on that data point in various analytical products and then notify analysts and other users that the data is inaccurate. An example of a wrong data point is the report that a white panel truck was associated with the 2002 sniper attacks in the Washington, DC area. This wrong data point diverted the police's attention from other, truly relevant data points. As a result, the public was on the lookout for the wrong type of vehicle.

Since the network itself will be a target for both inside and outside threats—and because the information on it could also be misused—the security of it (from both physical and cyber attack) and of the information within it must also be a priority. This requires not only new technology, but also rules and procedures for building an environment of trust. For without trust, no one will share. The system must constantly be screening for potential insider threats and misuses of the information, and it should have access controls and multifactor authentication built in. In short, security and information assurance must be designed directly into every element of the network. They cannot be grafted on.

The network, too, must not only enable users to push information to others, it must also enable users to pull it on demand, or at least give each user pointers to a person who can determine whether the user is authorized to

[31] We use the term "information rights management" rather than "digital rights management" to refer not just to the specific technologies used by the music and movie industries to protect their products against piracy, but to all technologies that protect and control access to and use of information. Information rights management allows individuals or organizations to specify who can access and use documents or portions of documents, and helps prevent sensitive information from being printed, forwarded, or copied by unauthorized people.

[32] A system with immutable audit capabilities would, for instance, immediately and permanently record who authored, changed, or accessed information; who posed queries to the system, what the queries were, and what the responses were; and who shared information with whom, and when. This means no individual could inappropriately access information or query the system and then hide the fact from an after-the-fact audit. Inspection of the audit logs can also be controlled in a way that would require multiple parties to unlock the logs, so as to make those logs tamper-resistant as well.

access the information. The network must support the bidirectional sharing of information between public agencies and between agencies and private data holders and transaction processors. Users should be able to make data requests, publish and subscribe information, perform directory searches and federated queries across databases, and examine and integrate information from other portals, all from their own work environment.

In addition, the network itself must be aware. Too many of our existing analytical tools are based on the query model, and assume that analysts can ask every smart question every day. This is especially difficult because most of the information that analysts obtain is processed sequentially, in the order received. The network should constantly screen, without the need for human direction, for information that matches government watch list data and for new patterns that indicate potential terrorist activity.

Because we cannot realistically expect a new architecture to be built overnight, or ask that the federal government require all players to upgrade their operational systems to comply with sharing requirements, legacy systems will inevitably be part of the new architectures. Moreover, legacy data pertinent to the mission will still have to be accessible and shareable in the new architecture. The government should therefore have the capability to take information in any form, transform it into a useable format, perform quality assurance, and publish it on the network for access and use by the appropriate players.

With all of this information being shared, however, comes the risk of flooding the system with too much data, thereby causing the meaningful signals to be lost amid the noise. Some existing tools can help ameliorate this problem. For example, automatic personalization, extraction, and categorization tools can allow users to select what sorts of information they want from the network according to their individual needs. The government also could place sensors throughout the network, which would look for information about specific threats or individuals and only report matches. These sensors could be updated electronically so that they are always current and reporting relevant information.

Exhibit D

ATTRIBUTES OF THE SHARE NETWORK

The Systemwide Homeland Analysis and Resource Exchange (SHARE) Network we envision would have the following attributes:

1. **No single points of failure**
   - a. Support for redundant or complementary analyses in numerous locations
   - b. Multiple and redundant communication pathways

2. **Loosely coupled architecture**
   - a. Implemented in a decentralized, peer-to-peer environment in which information flows without dependence on a central information broker
   - b. Data repositories should be accessed through a common data layer and kept independent from applications to allow for easier interoperability
   - c. Adherence to industry-standard data-exchange practices
   - d. Ability to support on-demand as well as ad hoc information sharing

3. **Directory-based services**
   - a. Ability to find pointers to all information relating to persons, organizations, locations, time, and methods
   - b. Ability to support publish and subscribe models for information dissemination and to permit remote queries

4. **Support for real-time operations**
   - a. Real-time dissemination, collaboration, and communication
   - b. Leverages the edges of the network
   - c. Gets information to and from users at all levels, and provides feedback

5. **Security and accountability to prevent abuse**
   - a. Multifactor authentication and access control
   - b. Strong encryption and data protection
   - c. Immutable audit capabilities
   - d. Automated policy enforcement
   - e. Perpetual, automated screening for abuses of network and intrusions

Participants in the network: organizational structure

The technical architecture described above is only part of what is required to develop a network that brings many disparate participants together. In order to create the sort of decentralized, coordinated network we envision, the government must also address the organizational structure of the players in the network.

The federal participants

We need to begin with a structure at the federal level that makes the sharing of information among relevant federal agencies, and with state, local, and private sector entities, a central part of its mission. Although some agency officials have become convinced that this is the right direction for government investments, the federal government has not designated an organization to lead the creation of a decentralized network. We believe that the President should take steps within the Executive Branch to clarify this leadership role. We have expressed our preference for the DHS to be assigned the lead in designing the architecture of the network and overseeing its implementation with representatives of the other participants in the network. For this effort to succeed, the cooperation of all agencies must be ensured through a combination of Executive Order, organizational structure (for example, an interagency, public-private group), and incentives that ensure a clear assignment of responsibility, adequate resources, and accountability for outcomes. In addition, the President needs to set forth, in an Executive Order, guidelines that establish the principles for using this network to improve information collection, analysis, and sharing while protecting civil liberties.

The President needs to set forth, in an Executive Order, guidelines that establish the principles for using this network to improve information collection, analysis, and sharing while protecting civil liberties.

It is also important that the President clarify, in an Executive Order, the roles of the TTIC and the DHS, and clearly delineate their respective responsibilities.[33] If the TTIC is to be a crucial, though not exclusive, locus for fusing and sharing information within the federal government and for eliminating the gap between foreign and domestic intelligence on terrorism, then it must begin placing more emphasis on producing analyses for intelligence consumers throughout the government, rather than devoting its attention almost entirely to serving the needs of the President and other senior officials as it currently does. Any agency will naturally adjust its activities to respond to requests for information from the Executive Office of the President, even if that necessarily means devoting less attention to other parts of its mission. Therefore, the President himself must call for this change. He should make clear that serving the needs of operational components of the network is a major priority for the TTIC. Failing the necessary presidential action, Congress should consider stepping in with legislation.

In addressing the needs of the other participants in the network, the TTIC should not serve as the centralized hub of a hub-and-spoke model for information sharing. That is, information should not have to pass through the TTIC in order to be shared with other agencies. Rather, the TTIC should be one important analytical node in a decentralized system in which the participants share directly with one another.

Moreover, the creation of the TTIC as an all-source intelligence fusion and analysis center—with access to both foreign intelligence and domestic intelligence and law enforcement information concerning U.S. persons—confronts us with the question of what will replace the previous "line at the border" that largely defined the distinctive rules for foreign and domestic intelligence. There has been no significant public debate on this fundamental question, and it is a critical area for presidential guidance. It is possible that the Executive Branch has radically changed the balance of liberties with this organizational move.

Foreign intelligence agencies have traditionally operated abroad with relatively few constraints on their collection activities. Domestic law enforcement and counterintelligence agencies, on the other hand, traditionally have operated under much stricter rules designed to safeguard the rights and liberties of U.S. citizens and residents. Since at least the mid-1980s, with the growth of international terrorism and international narcotics trafficking, the activities of foreign intelligence agencies and domestic law enforcement and counterintelligence agencies have increasingly overlapped. As a result, the two communities have had to work more closely and share more information than ever before.

[33] Indeed, while the President announced the concept of the TTIC in January 2003 in his State of the Union speech, and a subsequent presidential directive (Homeland Security Presidential Directive/Hspd-6) refers to the TTIC, there is, to our knowledge, no presidential order that actually created the TTIC. Rather, the TTIC's roles and responsibilities are set out in Director of Central Intelligence Directive (DCID) 2/4 (effective May 1, 2003), which is classified. This exemplifies the lack of adequate public discussion attending the creation, mission, and authorities of this important new organization.

The creation of the TTIC, however, takes this coordination and sharing to a new level. It is therefore imperative that we have an open, public debate about what new rules are needed to replace the "line at the border." At the very least, the President should set out in an Executive Order clear guidelines governing the authority of the TTIC—and any other agencies that have access to both foreign and domestic intelligence and law enforcement information—to receive, retain, and disseminate to U.S. and foreign intelligence agencies information gathered in the U.S. about U.S. persons.[34] The Order should also contain guidelines to govern the intelligence agencies' ability to set requirements for (to "task") domestic collection of information. These guidelines should, to the extent possible, be unclassified and put out for notice and comment so that the American public can have insight and confidence in the way domestic information is collected and used by the government. It may even be appropriate for the President to initiate this important public debate by introducing legislation to codify the appropriate scope of the TTIC's use and dissemination of information about U.S. persons. In short, guidance is needed to empower the TTIC and other agencies' analysts, as well as to constrain improprieties. Without it, agency personnel may be reluctant to share information that could prevent a terrorist incident.

Moreover, the Executive Branch should create within the TTIC the appropriate institutional mechanisms to safeguard privacy rights. When Congress passed legislation to establish the DHS, it was careful to include a privacy officer and a civil rights and civil liberties officer. If the TTIC is going to perform much of the analysis and information-sharing mission Congress had intended for the DHS, then it should have commensurate privacy-protection measures.

Even if the TTIC plays the role described above, we continue to believe that the DHS has a vital role to play as well. First, as noted above, the DHS should have the lead responsibility for developing the architecture for the SHARE Network we envision. Second, while we firmly believe that all federal agencies have a responsibility for sharing relevant information across all levels of government, the DHS should have the principal responsibility for facilitating and ensuring the sharing of information with state and local governments and the private sector. Third, the DHS should focus its own analytical resources on the nation's vulnerabilities to terrorist attack and on matching those vulnerabilities with threat information from the TTIC and others to determine which targets are at greatest risk and what protective measures are needed.

The creation of the TTIC as an all-source intelligence fusion and analysis center confronts us with the question of what will replace the previous "line at the border" that largely defined the distinctive rules for foreign and domestic intelligence.

The DHS's role in ensuring that information is shared with state and local governments should not preempt the FBI's unique role in sharing investigative information with state and local law enforcement agencies, since those agencies must often work jointly with the FBI on investigations. But the FBI should be more willing to share directly with state and local law enforcement agencies, and not just with the state and local representatives on the FBI-led Joint Terrorism Task Forces (JTTFs), who are precluded from sharing with their home agencies without the FBI's approval.[35] (See Exhibit E for more information about the JTTFs.) Indeed, the FBI should see itself as part of the whole network, sharing appropriate information with all of the other relevant players, rather than viewing itself as the top entity in its own law enforcement stovepipe. Similarly, it should encourage its state and local law enforcement partners to share information directly with other players in the network, rather than actively discouraging such broad sharing.

[34] Analogous issues are raised by the creation of Northern Command (NORTHCOM), the military's unified command responsible for the defense of the U.S. and for support to civil authorities engaged in homeland security. In addition to concerns unique to the military (such as the restrictions on military involvement in law enforcement activity under the Posse Comitatus Act and DoD regulations), NORTHCOM's mission raises question about what guidelines are necessary to govern the military's access to information about domestic activities. As a Congressional Research Service report put it, "In order to defend the U.S. from attack, NORTHCOM has a strong rationale for access to information collected by various intelligence and law enforcement agencies. However, at a certain point, such access could create the perception—or the reality—that the military is spying on U.S. citizens. What type of access should NORTHCOM be given to various types of sensitive data? What types of safeguards need to be established to ensure that this data is used properly?" See CRS Report for Congress, *Homeland Security: Establishment and Implementation of Northern Command* (14 May 2003), at 5, available at http://www.fas.org/man/crs/RS21322.pdf (last visited 12 Nov. 2003).

[35] See, for example, Testimony of James Kallstrom (footnote 6), at page 4 (stating that "important information [from the JTTFs] does not reach the officers responsible for patrolling the cities, towns, highways, villages, and neighborhoods of our country," and that the JTTFs have not sufficiently empowered state and local officers to act as "eyes and ears" by providing them with necessary information).

Exhibit E

JOINT TERRORISM TASK FORCES (JTTFs)

JTTFs are groups of state and local law enforcement officers, FBI agents, and other federal personnel who work jointly under the leadership of the FBI to investigate and prevent acts of terrorism. These task forces are designed to tap the expertise of different agencies and facilitate the collection and sharing of intelligence.[36] The first JTTF was created in 1980, and the total number of task forces has more than doubled since September 11, 2001. Today, there are 84 JTTFs.[37] In addition, in 2002, the FBI created a National Joint Terrorism Task Force at FBI headquarters in Washington, DC. The National JTTF comprises nearly 30 intelligence, public safety, and federal, state, and local law enforcement agencies. It collects terrorism information and distributes it to the 84 JTTFs as well as to terrorism units within the FBI and partner agencies.[38]

Moreover, beyond pushing information to other players, both the DHS and the FBI should build the capability, and instill a culture of willingness, to respond to requests for information from state and local entities. Those entities have knowledge of their communities and of vulnerabilities within and potential threats to their jurisdictions, and they need to be able to tap into the information held by the federal government in order to be effective. Accordingly, the DHS and FBI should establish clear mechanisms for responding to requests from state and local officials for threat and vulnerability information, and these agencies should establish a culture that makes responding to such requests a priority.

Ultimately, these sharing mechanisms should be automated, allowing state and local officials to pull relevant information from federal databases. Automation would require the use of directories, data-transformation capabilities, and technology that identifies users who have permission to access certain information, as discussed above. It would also require the requisite security and auditing procedures and technology. Agencies should make such technology a procurement priority. In the short term, until the requisite technology is introduced, however, federal agencies should at least make clear whom state and local agencies can call to obtain information. This can be done by establishing online directories, which can then be built into automated systems over the long run.

Decentralized analytic nodes

There has been much debate about how best to achieve intelligence fusion and analysis. The discussion is often cast as a choice between centralizing this function in one agency or within several agencies in Washington, DC, and decentralizing analysis among all relevant players. In fact, this is a false choice. As our vision of the SHARE Network indicates, we need both centralized and decentralized analysis. Redundancy, or complementarity, of analysis is beneficial. We need, for example, an agency like the DHS or the TTIC that is capable of pulling together relevant intelligence and law enforcement information so that the government can put together as many pieces of the puzzle as possible and gain a full view of terrorist threats. But we also need other entities at the edges of the network that are capable of gathering pieces. Intelligence analysis is largely a matter of trying to assess the probabilities of connections among people or events from uncertain facts that are susceptible to different interpretations, and making predictive judgments about the future. Therefore, a system in which multiple analysts look at information from different points of view is more likely to reveal signs of potential terrorist activity. In addition, the reality is that the TTIC and a local or state agency might be working on different puzzles or different parts of the same large puzzle. The TTIC might be looking at the activities of foreign terrorist groups and their plots against U.S. interests in general, while a local police agency might be looking at a specific criminal group that is only one small part of a terrorist group. We would not want, nor can we reasonably expect, a single entity to be responsible for performing both sorts of analyses. Moreover, different analytical entities produce different sorts of products for different audiences, ranging from a strategic intelligence analysis for the President or Cabinet officials to tactical leads for local police departments.

[36] See FBI, "War on Terrorism, Counterterrorism Partnerships," available at http://www.fbi.gov/terrorinfo/counterrorism/partnership.htm (last visited 12 Nov. 2003).

[37] See FBI, speech prepared for delivery by Director Robert S. Mueller, III, at 110th Annual Conference of the Int'l Ass'n of Chiefs of Police (24 Oct. 2003), available at http://www.fbi.gov/pressrel/speeches/iacp102403.htm (last visited 24 Oct. 2003).

[38] See FBI, "War on Terrorism, Counterterrorism Partnerships," available at http://www.fbi.gov/terrorinfo/counterrorism/partnership.htm (last visited 12 Nov. 2003).

Currently, the FBI's JTTFs constitute one form of decentralized analytic node. Other, interdisciplinary analytical groups should also be encouraged, and these groups should be tied into the network and encouraged to communicate directly with one another as well as with the DHS and the FBI. In order for these decentralized entities to be a true part of a network rather than becoming their own stovepipes of information, it is critical that they adopt common (or at least interoperable) standards and formats for communicating and that they publish metadata about their information in integrated directories so that their information may be easily located and shared quickly with others in the network. In addition, guidelines are needed that address not only how information should be shared, but also when it should be shared, and with whom. The DHS should work with state and local government entities to create additional decentralized analytical centers, and should foster their ability to communicate not only with the DHS and the FBI, but also directly with one another.

Beyond state and local government, private sector entities must also be brought into the network. To date, some industries have formed Information Sharing and Analysis Centers (ISACs) for the purpose of analyzing and sharing information among companies and between that industry and the federal government. These ISACs were originally formed to deal with cyber-security information. Since September 11, however, many have broadened their scope to deal with terrorism-threat information as well. But ISACs have a mixed record when it comes to the amount of information actually shared among companies or with the government. Moreover, existing ISACs are generally limited to critical infrastructure sectors (such as electrical energy, information technology and telecommunications, and financial services). As terrorists increasingly seek soft targets where they can take innocent lives without confronting tight security, it is important that the federal government have the ability to communicate quickly and broadly with non-infrastructure companies.

Thus, we believe the DHS should work with private companies to improve the two-way flow of terrorism-related information between government and industry. The DHS should help to expand the scope of all existing ISACs beyond cyber threats to include focus on terrorism-threat information, and it should encourage the ISACs to share more information with the government and with other industry ISACs. The DHS should also foster the creation of new ISACs or other mechanisms to bring together non-infrastructure companies that might be the target of attack or that might, in the course of their business, collect information related to terrorist activity. The DHS should also work with ISACs to establish information-sharing standards and, where necessary, provide seed funding.

The creation of such new analytical centers, of course, will only exacerbate the current shortage of qualified, trained analysts. It is therefore imperative that the government make training of new and existing counterterrorism analysts a priority—not only at federal agencies, but at the state and local level and in the private sector as well. Such training would make these decentralized nodes more attuned to the kinds of information they should be looking for and enable them to be more valuable participants in the network.

The road to a culture of distribution

The biggest obstacle to implementing the best-designed systems in the world is often culture. Organizations, processes, and technologies can be changed, but unless fundamental changes occur in the culture of the participants in an existing system, progress is stymied. We have identified some critical vehicles for changing culture, which are discussed in this section along with the necessary processes and procedures to cement the change. We emphasize, however, that no vehicle will lead to change unless the leader at the top is completely clear about the objectives he or she seeks. Thus to implement our model, the President has to make absolutely clear that his objective is to create a decentralized network for robust information sharing and analysis that produces actionable intelligence.

Decisions about sharing intelligence in the government are still made largely in the context of a system of classification that was developed during the Cold War. Our collection efforts then were focused on maximizing collection, by human or technical means, against targets overseas. Agencies were organized around collection and had analytical units to help sort, analyze, and reduce the data to semifinished intelligence reports. Analysis was designed to first serve a small number of senior policymakers (the President, Vice President, Secretary of Defense, Secretary of State, DCI, etc.) and second, to serve a larger but still small number of high-level decision-makers. In this context, classification was seen as an important tool to protect the sources and methods through which intelligence was collected, because access to information was limited to a small group of individuals. The government further limited access to information by imposing a requirement that any individual who wanted to see the information had to have a demonstrable "need to know," and by establishing procedures that allowed the originating agency (the agency which first obtained the information) to strictly control the dissemination of that information within the government. This system assumed that it was possible to determine in advance who needed to know particular information, and that the risks associated with disclosure were greater than the potential benefits of wider information sharing. The formal limits on sharing imposed by this system were exacerbated by the widely acknowledged problem of overclassification. That is, far more information was classified initially—and remained classified—than was necessary or appropriate.

This mind-set of classification and tight limits on sharing information is ill suited to today's homeland security challenge. While certain information must be protected against unauthorized disclosure, the general mind-set must be one that strives for broad sharing of information with all of the relevant players in the network. The system must be designed to address the needs of the potential users of information, not just the security concerns of the collectors.

Instead of a culture of classification and occasional, post-facto sanitization of classified documents, we need a culture of distribution, in which the rewards go to those whose information has been found most valuable by people across the network.

One of the principal reasons that federal agencies do not widely share information with one another and, especially, with state and local governments and with private sector entities is fear that the information would be leaked to the media and the public—and thus to our nation's adversaries as well—thereby putting lives at risk, jeopardizing intelligence sources and methods, compromising law enforcement investigations or prosecutions, or violating individual privacy rights.[39] These are legitimate concerns. But these concerns can be ameliorated if federal agencies put in place regular processes for producing information in a way that allows it to be shared even if it comes from sensitive law enforcement or intelligence sources.

Government agencies currently rely on processes for "sanitizing" classified information so that it can be shared with other agencies. Sanitization involves removing from a report any sensitive information that the originating agency believes cannot be shared widely with other agencies without undue risk to sources and methods or some other legitimate interest, while still providing the gist of the information so that recipient agencies can take appropriate investigative or protective actions or utilize the information in their analyses.

Currently, some federal agencies sanitize some reports to remove source and method information. But the sanitized version is often still classified, and is usually designed for dissemination only to other federal agencies. Sanitization does not generally occur as a matter of course for many agencies, and no agency, to our knowledge, regularly produces a sanitized version of information that is unclassified and appropriate for wide-scale dissemination to state, local, and private sector entities. The sanitization process is also often slow and cumbersome.

The process needs to be reversed so that distributable products are created at the outset.[40] That is, instead of a culture of classification and occasional, post-facto sanitization of classified documents, we need a culture of distribution, in which the rewards go to those whose information has been found most valuable by people across the network. We need to reward those who figure out exactly what information others in the distributed system need to see, and who make sure the other players get that information in a form they can use.

All federal agencies responsible for collecting terrorism-threat information also should see state and local government agencies and, in some instances, private sector entities, as regular consumers of their information. Thus, these agencies should produce unclassified reports of relevant information that may be disseminated to state, local and, in some instances, private sector entities. But one agency, the DHS, should serve as the backstop, the guarantor that as much information as possible is being shared. To do this, the DHS should establish a process for resolving disputes between originating agencies that want to prevent further dissemination and those agencies that need more information.

[39] This is also true for state and local agencies. Private companies also often decline to share information because of their concerns about disclosing information that is proprietary or that might cause public embarrassment or a loss of shareholder confidence.

[40] To some extent, this can be seen as an expansion of the current approach of some agencies to producing "tear-line" reports, in which an agency produces a classified version of information with a less classified, or unclassified, version below a tear-line. In our approach, the production of such alternate versions would be commonplace and automatic. And it would be a top priority. For example, an agency would create a "Top Secret/Code Word" report that reveals the source of the information; a "Secret" version that would not reveal the source but might give explicit detail on the threat; and a "Sensitive But Unclassified" version that might only contain the necessary action the recipient agencies should take given their specific roles in the network (for example, to be on the lookout for certain individuals or indicators of specific terrorist activity).

Technologies exist that can facilitate the sharing of sensitive information. For example, screening tools could be used to assist in the redaction process when moving information across security levels. Screening tools can automatically alert disseminators when potentially sensitive information is about to be transmitted, or when information may be about to be sent to parties that lack the requisite permission to receive it. Semiautomated systems could also suggest special-handling guidelines as well as who should be included on dissemination lists.

While such measures would foster the dissemination of actionable information to other players in the network, they would not entirely eliminate the risk of unauthorized disclosure and the harm that such disclosure can cause to both government counterterrorism operations and to citizens' rights. Even when sources and methods or personally identifiable information is removed from a disseminated report, that report still could, if made public, reveal important clues about the government's knowledge of a terrorist group or plot, or infringe on a citizen's privacy if the missing pieces of data can be discerned from other sources. Moreover, as information is shared among agencies with overlapping jurisdictions, there is a risk that uncoordinated action by one agency in response to that information could impede or disrupt a sensitive counterterrorism operation by another agency. If one federal agency, for example, shares information about a terrorist group that it has been investigating clandestinely for a long period, and another agency then undertakes its own investigation of that group, the second agency's actions could disrupt the first agency's investigation and cause the loss of vital intelligence. Finally, a recipient of information that is not suitable for public disclosure (for example, information of uncertain credibility about a potential terrorist threat to an infrastructure asset) could take action or make public statements that cause undue public alarm if the threat turns out to be unfounded. Additional measures must therefore be taken to minimize the risk of unauthorized disclosure of information and ensure coordination by recipient agencies before information is acted on.

While there is no easy solution to this problem, improvements can be made. Auditing technology, for example, could be deployed to track the flow of information to different players and to record how the information is used (whether, for example, it is printed, forwarded, or edited). This could help deter leaks. The auditing tools should use strong means of authentication that have forensic value (that is, they should be permissible in court to prove access). Information rights management technologies, when combined with digital certificates, can also help by allowing agencies to create self-enforcing rules about who can have access to particular documents, how they can be used, and how long the documents can be viewed before access expires. Another possibility would be to make federal funding for information-sharing purposes contingent on the adherence to certain rules prohibiting unauthorized disclosure. Another improvement would be the establishment of "deconfliction" centers populated by representatives of relevant agencies, which would ensure the coordination and deconfliction of investigations and operations by multiple agencies. Finally, information could be accompanied by clearer, more specific handling requirements and dissemination limitations. While none of these measures is perfect, a combination of such efforts might reduce the chance of unauthorized disclosure or uncoordinated action, and thereby foster a healthy environment for the sort of broad communication that we envision.

Another issue that must be dealt with to foster more sharing among government agencies is digitization of data, both active data sets as well as certain important legacy data sets. As discussed above, because it is not always possible to distinguish signal from noise when information is first collected, we must ensure that even when information is not actively disseminated, or pushed, to other entities, it is registered in a directory so that it can be easily located later and pulled by analysts with the appropriate permissions. Given the vast amounts of data that are already in the system—and the vast amounts of additional data that will be collected—we cannot rely on analysts to remember information that seemed unimportant at the time it was collected, but that may be of use later. Thus, to make the system work, information must be stored digitally and retained long enough for it to be useful when other information comes to light. Standards should therefore be developed—under the leadership of the DHS, but with participation from experts in government, industry, nonprofit organizations, and academia—to ensure that information in the network is digitized, stored, and retained, and that it is searchable at a later date.

Measuring performance

Instituting these new processes and, more fundamentally, instilling a culture of sharing will not happen overnight. As we have said, active engagement from the President himself, the National Security Council and Homeland Security Council, and the heads of agencies, as well as continual oversight from Congress will be required to ensure follow-through. As part of the process, then, we believe agencies' performance in meeting the information-sharing and analysis objectives should be evaluated after a reasonable implementation time. We therefore recommend that the President set forth specific and clear objectives for improved analysis and information sharing, based on the recommendations above, which each federal agency should be required to meet by December 31, 2004. At the conclusion of this period, the Executive Branch and Congress should evaluate how agencies have performed in meeting those objectives. If an agency has not performed adequately, the President and Congress should consider making any necessary changes. The government could also evaluate agencies' performance by assessing how well they would do in meeting the information-sharing challenges set out in some of our information vignettes (see Appendix D).

We also think the DHS should include state and local government and private sector entities in a regular process for assessing how well information is being shared with them, akin to the process the intelligence community currently uses for having customers of intelligence evaluate collectors. Concomitantly, the DHS should work collaboratively with state and local governments and private sector entities to set analysis and information-sharing objectives for them to meet as well, and jointly evaluate their performance after December 31, 2004, and thereafter on an ongoing basis. See Exhibit F for a set of metrics that we believe should be the basis for the Executive Branch's and Congress' evaluations.

Exhibit F

EVALUATING IMPROVEMENTS IN INFORMATION SHARING AND ANALYSIS

We have recommended that after December 31, 2004, the Executive Branch and Congress evaluate the progress of federal, state, local, and private sector entities in improving information sharing and analysis, consistent with the recommendations in our report. We set forth here some questions that Congress or others may ask to determine whether adequate progress has been made toward the goals set forth in this report. The questions reflect an ambitious but realistic set of expectations. With issues as important as these, progress must be rapid. On matters that require significant organizational changes or new funding, however, it is not realistic to expect that the job will be completed in a single year. Therefore, some of the objectives embodied in the questions are interim steps that would represent reasonable progress toward satisfying the goals.

Clarifying roles, responsibilities, and authorities

For effective information sharing, the Executive Branch must clarify the respective roles, responsibilities, and authorities of the players responsible for homeland security information. The respective roles of the TTIC, the DCI's Counterterrorist Center (CTC), the DHS's Directorate of Information Assurance and Infrastructure Protection (IA&IP), the FBI and its JTTFs, and the Defense Department's Northern Command (NORTHCOM) are not clearly defined. As long as this remains true, there will be turf battles among agencies and, most significantly, gaps in information sharing and analysis. Moreover, regardless of how these entities' roles are defined, foreign intelligence agencies will continue to have greater access to information about U.S. persons (citizens and legal resident aliens) than in the past. This increased access blurs a line that has long been in place to reduce the risk of government abuse of privacy and other civil liberties of U.S. persons. Although increased information sharing among law enforcement and intelligence entities is critical to the counterterrorism mission, no clear government-wide direction has been established for appropriate handling of domestic information while protecting civil liberties.

Question set 1

Are the federal government's homeland security agencies and players acting with clear guidance about their respective roles and responsibilities for information sharing, collection, and analysis? Which agency or agencies are responsible for communicating with state, local, and private sector players about homeland security intelligence, threats, and warnings? What are the respective analytic responsibilities of the players? Which intelligence entities have tasking authority over domestic collection, and how can that authority be exercised and coordinated?

Question set 2

Are homeland security agencies and players acting with clear guidance for the collection, handling, and dissemination of information about U.S. persons? Does this guidance permit the flow of information necessary to fight terrorism, but maintain the protection traditionally afforded U.S.–person information? Is the guidance, to the maximum extent possible, available to the public?

Information sharing within the federal government

Although there have been significant advances since September 11 in the ability and willingness of intelligence, law enforcement, and other agencies to share information relevant to countering terrorism, significant roadblocks remain. Thus, the Executive Branch must make greater and more rapid progress toward removing them.

Question set 3

Have the federal homeland security agencies taken significant and measurable steps toward adopting an information-technology architecture with the basic characteristics that the Task Force has described? Does each agency have sufficient guidance for procuring new technology so that it does not buy products that are incompatible with this architecture?

Question set 4

Are all terrorism-related watch lists in the federal government available for combined searching in real time, or at least for the matching of names and related information? Are there consistent standards regarding how individuals are placed on watch lists, how information about such individuals is managed, what types of data should be kept to enhance the government's ability to confirm identities of individuals, and the process for correcting errors and allowing innocent people to be removed from such lists?

Question set 5

Are terrorism intelligence and threat and warning information flowing efficiently and effectively through clear channels and with regular auditable procedures rather than through informal channels that are based on personal relationships and ad hoc judgments about who should receive information?

Question set 6

Have bureaucratic or other institutional roadblocks to sharing information—such as requirements for originator approval, inadequacy of facilities for storing classified information, and "paper only" intelligence products—been eliminated or minimized? Have positive incentives been developed to foster more information sharing, such as rewarding analysts who produce disseminable products that are of great value to others in the network?

Question set 7

Are FBI field offices producing intelligence reports—even from ongoing cases? And are they immediately and automatically sharing these reports with FBI headquarters and other appropriate recipients?

Producing intelligence for a new customer

Intelligence agencies often see their job as sending information up to the President and other senior officials. They do not always view operational entities—particularly those outside of the federal government—as their customers. Therefore, they are not accustomed to creating reports that are available or useful for these other entities. One of the principal reasons that homeland security threat information and other intelligence reports are not shared widely is that they are classified. An important step in creating the culture of distribution that the Task Force recommends is to increase the information that is available for distribution—that is, unclassified information.

Question set 8

Are intelligence agencies responding to the intelligence needs of their new customers? Is there regular, formal interaction between those responsible for preparing intelligence in the federal government and the state, local, and private sector players who need information? Does this interaction result in specific, substantive requirements for intelligence producers?

Question set 9

Has it become part of the culture of intelligence agencies to create unclassified versions of intelligence reports on terrorism? Are there reward and audit mechanisms to encourage this culture? Do automated report formats have required fields for an unclassified version? Are unclassified versions prepared for at least

80 percent of these reports? Is there a mechanism for a non-originating agency to seek further declassification? When used, does this mechanism result in further declassification a significant percentage of the time?

Communicating with state and local governments and the private sector

In addition to producing intelligence for these new customers, there are a number of steps the Executive Branch should take to promote a networked information architecture and improve communication with players outside of the federal government about intelligence, threats, and other information relevant to countering terrorism.

Question set 10

Has the federal government convened state, local, and private sector players to develop common standards for information sharing? Have the parties developed common or interoperable metadata formats and definitions, directory formats, and communications methods and protocols to facilitate information sharing by all players across the network? Have they developed common or consistent policies on retention, dissemination, and sharing of data? Has the government leveraged appropriate technologies (such as anonymization, information rights management, automatic policy enforcement, and immutable audit) that may enhance information sharing, protect sensitive information, and foster auditing and accountability?

Question set 11

Has the federal government coordinated and provided incentives for the digitization of data in state, local, and private sector systems? Is this data accessible online in a secure manner and in near real time? If not, have these parties taken significant and measurable steps toward adopting these system changes?

Question set 12

Has the DHS established up-to-date and updateable contact and profile directories that are available to all players in the homeland security network? Do those directories include contact and profile information for businesses and other entities that are potential targets or that might require threat information for other reasons; for experts in government, the private sector, and academia, who can be called upon for guidance or insight with regard to a particular threat; and for other specialists and entities that might either contribute or require information about homeland security threats or warnings? Is the DHS utilizing existing technologies to create and manage these profiles and ensuring that the profiles are current and relevant?

Question set 13

Are state and local law enforcement personnel able—quickly and automatically—to do a name match between federal terrorism watch lists and individuals they have in custody or under surveillance and obtain more information about those individuals from federal sources? Is that additional information sufficient to provide useful guidance to state and local authorities on how to proceed with that individual and to reduce false positives (that is, to determine whether the person in custody or under surveillance is actually the same person as the one on the federal watch list)?

Question set 14

Does the federal government—in particular the DHS and the FBI—have clear, workable procedures for sharing intelligence and threat information with state, local, and private sector homeland security players? Do these nonfederal players understand these procedures, and if so, do they take advantage of them?

Question set 15

Is there regular and substantive communication between the federal government—particularly the DHS and the FBI—and nonfederal homeland security players about intelligence and threat information? Does the communication flow in both directions? Have federal entities established mechanisms to encourage and reward their employees' responsiveness to nonfederal players? Are the nonfederal players satisfied with this interaction? Can the federal entities effectively ingest and utilize information from state and local actors?

Question set 16

Have the parties developed mechanisms for communicating targeted requests for information from the federal government to nonfederal homeland security players, and if so, are these mechanisms in use? Can the DHS or FBI point to several specific examples of targeted requests for information that have generated thorough and useful responses from state, local, or private sector entities?

Improving analysis

It is critical to a homeland security information-sharing network that the information being shared is accurate and that its significance is understood. This requires analysis that combines substantive expertise and first-rate analytic tradecraft. In addition, analysis of threats to homeland security cannot all occur at the top—that is, in Washington, DC. To be effective, analysis must occur at all levels of the network.

Question set 17
Is there a federal government entity responsible specifically for producing long-term, strategic analysis of terrorist threats? Does that entity have the number and quality of analysts necessary to carry out that function? Is it actually producing a steady and useful stream of such intelligence?

Question set 18
Are federal government intelligence agencies producing analysis for the entire range of customers who need it, including operational entities, the DHS and nonfederal actors? Are agency analysts according equal priority to analyses directed primarily at customers other than the President and senior policy officials, such as the President's Terrorism Threat Report and the President's Daily Brief? Are federal agencies providing access to useful data sets to foster decentralized analysis at the nonfederal levels?

Question set 19
Is the DHS producing—as its authorizing statute requires—analysis of the nature and scope of threats and potential vulnerabilities? Do its analysts have sufficient training and expertise in important areas, such as target industries (the airline industry, for example) and threat categories (the energy sector, for example), to produce quality analysis? Is the DHS producing actionable intelligence?

Question set 20
Is analysis occurring in the field, including at JTTFs and FBI field offices and in state, local, and regional analysis centers and organizations? Do these field analytical units have an understanding of broader analytical needs, and do they communicate regularly and effectively with other homeland security players? Are these field units receiving adequate training?

Improving the capabilities of state, local, and private sector entities

To be effective participants in the network, state, local, and private sector entities also need to take steps to increase their capacity to share and analyze information.

Question set 21
Have regional or state groups formed to promote information sharing and prepare for homeland security threats? Do these groups include representation from federal, state, local, and key private sector players in law enforcement, public health, and emergency preparedness? Are all 50 states, the District of Columbia, and U.S. territories participating in this type of group? Are state, local, and regional entities effectively sharing their information with each other and with the federal government?

Question set 22
Have law enforcement organizations in key localities that are the most likely targets of terrorist attack, or that have been the locus of terrorist planning and other activity (such as the New York metropolitan area; the Washington, DC, metropolitan area; Los Angeles, Chicago; Detroit; Phoenix; southern Florida; the Bay Area; Las Vegas; and Seattle) implemented information-system upgrades and digitization of metadata using common standards? Have additional jurisdictions taken steps toward implementation?

Question set 23
Have local or regional analytic centers been formed in key cities or regions such as the ones listed above? Have additional jurisdictions taken steps toward the formation of such centers?

Question set 24
Have ISACs or other government-industry groups been formed by key industries (including those not considered critical-infrastructure industries) for which no such groups currently exist? Are these groups effective at sharing threat and vulnerability information quickly (and preferably automatically), conducting analysis relevant to their industries, and communicating with federal, state, and local agencies?

Accessing private sector data

Today, the private sector is on the frontline of the homeland security effort. Its members are holders of data that may prove crucial to identifying and locating terrorists or thwarting terrorist attacks, and stewards of critical infrastructure and dangerous materials that must be protected. Thus, the private sector is a source of information that is essential to counterterrorism. We therefore start from the premise that the government must have access to that information, which is needed to protect our country, and that through a combination of well-crafted guidelines, careful articulation of the types of information needed for identified purposes, and effective oversight using modern information technology, it will be possible to assure that the government gets that information in a way that protects our essential liberties.

Information available in the private sector

In the past decade, we have seen an explosion in the quantity of personal information held by the private sector. Transactional data—such as point-of-sale data, credit card records, travel records, and cell phone call logs—increasingly makes it possible to track in minute detail, and sometimes in real time, the activities of individuals. (See Appendix H for a description of the many types of data available.) Access to this sort of data can be critical to a government agency's ability to investigate and understand the intentions of a suspected terrorist.

The challenging policy issue comes when the government tries to use private sector data to detect signs of potential terrorist activity by people it does not already have reason to suspect. Although using data in this way can be beneficial (see Appendices D and E for illustrations of how privately held data might help agencies understand the significance of suspicious activity by previously unknown entities), it raises serious civil liberties concerns.

Internet technologies, such as cookies, potentially allow (if linked with other information) access to some of the most private indicators of personal behavior and interest. And the exponential increases in both computing and storage capability at exponentially diminishing costs have made it possible—and inexpensive—to collect and exploit petabytes of data on virtually every aspect of our lives. For example, supercomputer performance can now be obtained at desktop prices by clustering 64-bit processors with terabytes of storage. This can be both a benefit and a threat. It can, for example, allow government authorities to examine transactional information that a terrorist believes is effectively beyond government scrutiny, and thus help those authorities to uncover a plot in progress. But it can also allow for intrusion into the personal lives of individuals whom the government has no cause to suspect of criminal activity.

All of this data is collected not under government mandate, but as a consequence of the more or less voluntary decision of citizens to avail themselves of services that require (or allow) private companies to collect information on their activities. In others words, customers appear willing to give up a certain amount of privacy in exchange for better service. For example, an online bookseller uses a customer's profile of past purchases to suggest new titles that may be of interest; a credit card company alerts a customer to unusual purchasing patterns that may indicate a stolen credit card or identity theft.

Often, however, information collected by private sector entities is used for purposes other than those for which the customer provided it. In fact, a great deal of information sharing takes place for commercial purposes without the knowledge or express consent of the consumer. This seems to be tolerated by consumers in part because it has led to an expansion of the commercial services available to them. Moreover, the standards for technologies like cookies were established in technical organizations before many of the public policy issues involved had surfaced, and many consumers appear to accept them for the limited purposes to which they have been put to date.

Moreover, in recent years, the scale of information collection has been dramatically augmented by the rise of data aggregation companies 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 records such as driver's licenses and property deed transfers). Collection sources as well as the algorithms used to create these data sets generally are proprietary. Data from aggregators has been used by companies for activities ranging from marketing to risk assessment, and by the government for law enforcement and to locate missing children.

Government agencies can readily buy these data sets from data aggregators, who can deliver the data to government users in any format necessary for immediate analysis. In addition, the aggregator can perform a certain amount of initial analysis, breaking down larger data sets into more focused collections of data called "data marts."

Government agencies can then merge these data sets into other data sets that they routinely maintain or collect (for example, criminal records or intelligence information). Moreover, sophisticated data-mining or “knowledge management” software, as well as technologies for profiling, pattern analysis, link analysis, and transactional fingerprinting, are available either as procurable hardware or software or as a service from the commercial sector. These technologies can allow agencies to analyze both structured (organized in a predefined, meaningful way) and unstructured data, allowing them to find patterns of activity or links among individuals and derive value from the sea of largely disorganized data available in various sources of transactional information. The result is a vastly richer data context, and thus more wide-ranging and, ultimately, effective analysis.

Much data is also available from open sources, such as the Internet. While much of that information can be valuable, it can also be of poor quality, come from questionable sources, and be easily manipulated. The government therefore needs to take special care when it integrates open source information into its data analyses.

Government agencies have always had access to certain kinds of privately held information. But historically, information requests to commercial organizations were made by government agencies on a case-by-case basis. Companies would either volunteer the information or fulfill a specific subpoena request from law enforcement. In some cases, the law might impose specific collection and reporting requirements (such as with financial services firms, which must submit Suspicious Activity Reports to regulatory agencies).

With the advent of data-mining and analysis tools and the increasing computational capability of computers and decreasing costs of storage, agencies at all levels of government are now interested in collecting large amounts of data from commercial sources. Such data might be used not only for investigations of specific people (for example, to help find associates of a suspected terrorist) but also to perform large-scale data analysis and pattern discovery in order to discern potential terrorist activity by unknown individuals. Both uses of private-sector data, but particularly the latter, have raised a number of concerns from industry and privacy advocates as well as from the broader public and Congress. Companies are concerned about both the cost of supporting ongoing data flows to the government and the potential damage to their reputations and businesses if their data is misused. Privacy advocates and many in the public are concerned that the government will have access to large repositories of personal identifiable information, which are often of questionable quality and which could be used inappropriately to profile U.S. citizens or legal residents and possibly result in the denial of services or infringement of civil liberties based on people’s race, country of origin, political views, or personal habits. Additionally, there is concern about potential mission creep: While the government may collect data today for counterterrorism purposes, once the government has the data, there are no guarantees that it won’t use the data for other purposes in the future. In fact, many responsible legislators advocate that collected data should, in fact, be used for other purposes. Contributing to all of the above concerns is the lack of transparency and accountability of the algorithms used by the government against large data sets to rate (or “score”) potential threats.

Another, more specific use of private sector data is to resolve, or confirm, identities. When a government agent is using a watch list, investigating a person, or tracing a phone number, he or she needs to be able to determine whether the person or thing being examined is, in fact, the intended object of inquiry. Without identity resolution, watch lists can be cumbersome and ineffective; use of them can generate many false positives and false negatives, and investigations can be led down blind alleys. The government needs access to appropriate identity-resolution data and services, and private companies are the best source of those services. Moreover, with appropriate safeguards, effective identity resolution can also have important civil liberties benefits: It can help distinguish between those who should legitimately be the subject of scrutiny and those who should not.

Identifying the private sector information the government needs

So the question is, how can the government best make use of the vast volumes of private sector data while protecting civil liberties and avoiding the imposition of undue costs on industry? As we said in our first report, “Data mining can be a useful tool. But it is also a tool that invites concern about invasion of privacy.... 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” (page 31).

We believe the place to start is identifying concretely the information the government needs to carry out its homeland security responsibilities. The best way to do this is to consider realistic situations that the government might confront. To this end, the Task Force developed several scenarios in order to identify some kinds of privately held data that the government might need when confronted with certain situations (see Appendices D and F). In some of our scenarios, the government would start with limited information about the mode of a planned attack (for example, a scuba diver attack on a hazmat tanker ship) and might need information on the identities of people who have the capability or access to the means to carry out that mode of attack (such as certified scuba divers) and information on facilities where the means can be obtained (such as dive shops). In other scenarios, the government would start with particularized suspicion about specific individuals, and then might seek to identify the suspects' associates by looking at records of common or related addresses, telephone and email accounts, financial transactions, and travel.

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 its impact on privacy and civil liberties.

But our scenarios, and the resulting information categories, are purely illustrative. The crucial point is that this is the sort of approach that the government should take as a preliminary step, so that it can concretely identify its true information needs before launching controversial efforts to accumulate or mine large volumes of privately held data.

Guidelines for government use of private sector information

The next critical step is for the government to establish guidelines to regulate access to, use, and sharing of private sector data among agencies. These guidelines would help the government to ensure the following: (1.) that information is used in ways that are consistent with core national values, including privacy, other civil liberties, and the functioning of an accountable democratic political system; (2.) that investigatory resources are deployed in a cost-effective manner to achieve priority goals, without wasting government resources or imposing undue costs on industry; and (3.) that government personnel have clarity about what is permissible so that they are not overly reluctant to engage in perfectly legitimate activity for fear of public or congressional backlash. For many, these goals are seen as contradictory, requiring trade-offs between security and liberty, or between government empowerment and individual liberties. But we believe these goals are, in fact, complementary. By focusing information strategies on high-priority uses and making the rules clear, the government could reduce the impact on privacy and related concerns, empower agency personnel to take the necessary steps to collect and analyze information, and enhance public support for those aspects of the information strategy that are truly essential.

Given the nature of the security problem, it is inevitable that some of the details of the guidelines will need to be classified. But that should not stand as a barrier to public discussion of the core issues discussed in this report. If the 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 the information is obtained and used in a way that minimizes the impact on privacy and civil liberties. The reason we seek to strengthen our homeland security effort is to protect our safety and our way of life. So the government's approach must give the public confidence that the value of collected information is significant in relation to the potential impact that collection will have on civil liberties and other important interests.

In thinking about guidelines, the government should start with the basic architecture—what is the appropriate level of protection for different types of information, and what kinds of standards and procedures might provide that protection. The current legal framework governing access to and use of privately held data is a patchwork quilt of different standards for information with similar sensitivity (such as wire, cable, and Internet communications) and inappropriate or nonexistent standards for

other kinds of information.[41] The complexity of these rules, and the confusion they engender, may cause government officials to be reluctant to take lawful and necessary action to gather important counterterrorism information for fear of crossing a vague line. At the same time, these rules offer little assurance to the public that their rights are adequately protected. We do not think it would be realistic or desirable to replace this set of rules overnight. But we do think that greater clarity is needed, and that over time the government should seek greater consistency in the rules governing various forms of privately held information, and that it should develop guidelines that bear a closer relation to the fundamental interests at stake. New guidelines should, at a minimum, address the following: (1.) government acquisition and use of private sector data; (2.) government retention of the data; (3.) sharing of the data by the acquiring agency with other agencies for purposes other than counterterrorism; and (4.) accountability and oversight. Following are principles that the government should consider in developing guidelines to address these issues.

Acquisition and use of private sector data

Rules governing access to and use of private sector data should be based primarily on two dominant considerations: the value of the information to the government, and the sensitivity of the information from the perspective of individual privacy and other civil liberties. Thus, for example, large data sets with information on individuals with no known connection to terrorism are of relatively low value to the government, while information on the whereabouts and activities of an individual who is credibly believed to pose a threat is of high value. With regard to the sensitivity of the information, nonpersonally identifiable information is the least sensitive; personally identifiable information that is generally available to the public (such as through a Google search) is more sensitive; personally identifiable information not generally available to the public (such as information provided to a vendor on the condition that there be no third-party dissemination) is still more sensitive; and certain personal information (such as financial or health records) is the most sensitive.

One particularly contentious issue is under what circumstances the government should have access to information that is widely available to the public. As discussed above, the explosion of information technology has meant that vast quantities of information are now generally available to the public, including personally identifiable information about relatively sensitive matters. A whole industry has sprung up consisting of firms that collect, aggregate, and mine that data for a variety of tasks, including employment screening, marketing, and risk assessment. In most cases these firms neither seek nor require the approval of the subject of the information; and the legal constraints are few except with regard to a small number of sensitive areas, such as health records under the Health Insurance Portability and Accountability Act (HIPAA).

Under current law, there are few restrictions on the government's ability to gain access to this kind of information. And many argue that this is appropriate, that it should be no more difficult for the government to gather information than it is for a commercial company or private citizen. We believe, however, that different considerations apply to government acquisition and use of personally identifiable data, even when it is widely available to the public. Although there are consequences associated with the data's being available in the private sector (such as loss of job opportunities, credit worthiness, or public embarrassment), the consequences of government access to and use of the data can be more far-reaching and can include loss of liberty and encroachment on the constitutionally rooted right of privacy (both in the Fourth Amendment and more generally), which is designed to protect citizens from intrusions by government, not neighbors or credit bureaus. Therefore, we believe that the government should not have routine access to personally identifiable information even if that information is widely available to the public.[42] At a minimum, there should be

We believe that the government should not have routine access to personally identifiable information even if that information is freely available in the public.

[41] The Task Force has prepared two matrices that set out the diverse array of laws and regulations covering governmental and commercial access to privately held data. See www.markletaskforce.org.

[42] Another example of government forbearance regarding information that it might be legally entitled to collect involves "cookies." The OMB has issued a rule prohibiting federal agencies from using persistent cookies (for example, cookies that lasted longer than a single session) to track visits to their websites absent demonstration of a compelling need and clear notice to the public. See Joshua B. Bolton, U.S. OMB, OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002 (26 Sept. 2003), available at http://www.whitehouse.gov/omb/memoranda/m03-22.html (last visited 12 Nov. 2003).

a requirement that the information be relevant to the counterterrorism mission and that this showing be documented and subject to periodic audit.

Technology can assist in enforcing such access and use guidelines. For instance, anonymizing technologies could be employed to allow analysts to perform link analysis among data sets without disclosing personally identifiable information. By employing techniques such as one-way hashing,[43] masking, and blind matching, analysts can perform their jobs and search for suspicious patterns without the need to gain access to personal data until they make the requisite showing for disclosure.

Government retention of private sector data

In our first report, we expressed our strong preference for keeping data in the private sector whenever possible, rather than having the government retain it. This would be a prophylactic measure to help ensure that data gathered for one purpose was not impermissibly used for another purpose and to promote public confidence that the government is not inquiring into the activities of innocent people. Leaving data in private sector hands has another advantage: The data can be searched as part of a broad inquiry without creating any stigma that would be associated with the government's holding that data itself. When the government conducts a search of privately held data, it is easier to maintain the sense that the searched data is simply part of an overall information landscape, and that the fact that particular data was searched does not connote anything about the individual who is the subject of the data. The government should strive for an approach to data mining that allows it to find correlations but without suggesting anything about the meaning of the data until after the data is analyzed.

In addition to the policy objections, there are technical and security reasons not to create large government databases. Quality management of such centralized government databases would be very difficult; when obsolete or inaccurate data is updated or corrected by the private sector source, the data would not necessarily—or easily—be updated or corrected in the central government repository. In addition, a centralized repository would become a target for cyber attack and espionage, a problem that can be mitigated by leaving the data in decentralized private databases.

There is some concern that this approach could result in costly delays in government access when urgency is vital. However, we believe that such delays can be ameliorated if directories and pointers to the private holders of information are used, and if information is accessible electronically and in a useable format once permission is obtained. Virtual aggregation and networking can also be part of the solution to this problem. Similarly, anonymizing technologies can be used to permit enterprises that screen for specific patterns and watch list matches to report to the appropriate agency only the information necessary to indicate when there are specific matches. The agency could then obtain the underlying information only after it made the requisite showing under the applicable guidelines. This would help prevent the government from amassing large databases of private transactional information and provide a more robust real-time solution than classical data-mining approaches.

In areas where the government has a compelling need to retain information, a solution might be to create trusted data banks within the government with strict limitations on who has access to the underlying data and for what specific purpose. Another way to help limit the retention of data to that which is essential to the mission would be to require formal, written justifications for the creation and retention of data sets that contain personally identifiable information. These justifications would be subject to review at the time the data set was created and also periodically thereafter to ensure that there was an ongoing need to retain the data. Justifications should be subject to fairly rigorous standards, such as "inability of the government to retain the data would significantly impede the counterterrorism mission."

Sharing private sector data with agencies not involved in counterterrorism

A key reason for leaving information in the hands of the private sector originator or aggregator is to avoid the risk that, once acquired by the government for a legitimate counterterrorism purpose, the data will be used for a different purpose without authorization by policymakers. This poses problems not only when the subsequent purpose is illegitimate, but also when the purpose is legitimate but unrelated to counterterrorism (such as the use of counterterrorism data to enforce child-support obligations). That said, an absolute ban on using counterterrorism

[43] A one-way hash is "[a]n algorithm that turns messages or text into a fixed string of digits, usually for security or data-management purposes. The 'one way' means that it is nearly impossible to derive the original text from the string. A one-way–hash function is used to create digital signatures, which in turn identify and authenticate the sender and message of a digitally distributed message." Webopedia, "One-way–hash function," at http://www.webopedia.com/TERM/O/one-way_hash_function.html (last visited 4 Nov. 2003).

information for any other purpose achieves the prophylactic goal at a potentially high cost—why should the government be barred from sharing information between agencies if the second agency could acquire the information directly from the private sector on its own? There will be considerable resistance to requiring agencies to acquire costly duplicate data sets simply to guard against the theoretical possibility that the data might be misused.

To avoid this problem, clear guidelines and procedures are necessary to permit legitimate sharing, and also to establish accountability for improper use. An agency wishing to acquire data that was first obtained for counterterrorism purposes should have to demonstrate that it was entitled to get the information directly under equally or less stringent substantive standards than those applicable to counterterrorism. Technology can assist in enforcing these guidelines. For instance, role-based access-control technology could restrict access to certain information to only those who had the appropriate permissions. Permission levels would be determined according to policy determinations but could be enforced through the use of technological markers assigned to specific users. Thus, if a police investigator working on a white-collar crime case attempted to access a database or record that was restricted to counterterrorism uses, for which he did not have the requisite permission, he would automatically be denied access. Encryption and key management could also be used to control access by making data available for only a specified period of time and to a specified set of users. If someone without the appropriate decryption key attempted to access the data, he would not be able to access it (or view it in plaintext). And when the decryption key expired, users who had been authorized would no longer be able to view the data either.

Accountability and oversight

Guidelines must also address the question of how we assure compliance with the required policies and procedures and foster accountability. In the highly decentralized system that we envision, there will be no single agency or entity with overall responsibility for the day-to-day decisions to acquire, retain, or disseminate private sector information. At the same time, relatively uniform standards and compliance are difficult to achieve if each agency is separately interpreting, applying, and auditing compliance with the guidelines.

We believe a blended system is necessary. Government-wide rules will be necessary, and some agency must have overall supervisory responsibility to oversee the application of the guidelines, including the training of personnel, the implementation of auditing procedures, and the imposition of consequences for failure to comply. We think the DHS should play this central role, particularly in light of Congress' decision to create strong privacy oversight as an element of the DHS structure. At the same time, each agency has a responsibility to develop its own procedures to assure compliance. This will be the most effective day-to-day guarantee that the guidelines are, in fact, respected.

Technology can play a key role in assuring accountability and transparency. For example, personally identifiable data can be anonymized so that personal data is not seen unless and until the requisite showing (specified in guidelines) is made. Selective revelation, another technique that permits a user to see only that data for which he or she has the appropriate permissions, can also be used. Auditing technology, too, can provide built-in recording and documentation capabilities to track how information is used, retained, and shared. Strong auditing capabilities could also allow individuals to make Privacy Act or Freedom of Information Act requests to see what was done with the data about them.[44] These technologies would also permit independent, third-party auditing of the data-mining and scoring algorithms used in pattern-analysis systems such as those that might be used in CAPPS II. This would help ensure adherence to guidelines regarding permissible data sources and profiling. Information rights management technologies could also be employed by commercial enterprises to restrict the use of supplied data to a particular purpose and for a particular period of time. The potential utility of such technologies underscores the need to develop technology architecture in parallel with the development of the substantive policies embedded in the guidelines.

Another aspect of oversight is ensuring the accuracy of the data that is brought into the network. Accuracy is vital not only to protect the privacy and civil liberties of individuals who can be harmed by the use of inaccurate data, but also to assure that information has real value to the counterterrorism effort. Data anomalies or false positives that mistakenly suggest that an innocent person is somehow tied to terrorist activity can, if uncorrected, have significant adverse effects on the individual. They can also waste scarce investigative resources. Fortunately, technologies exist that can help assure that information is up-to-date. For instance, agencies could use directories, pointers, and Web services so that there is only one data source

[44] See Freedom of Information Act, 5 U.S.C. § 552 (2003), and Privacy Act, 5 U.S.C. § 552a (2003).

(preferably in the private sector, as discussed above), which is always kept current. Version control and update software can also ensure that information is updated according to a regular schedule. Expiration-enforcement software can ensure that data is unusable after a certain date. And data pedigree technology can permit users to track the information that has been used in an analytical product and visualize information dependencies. Technology can also enable systems to alert the holders of derivative documents if the original underlying data has been changed, or even to change the derivative document if the underlying data is replaced in full rather than merely modified.

Technology, of course, is only one part of the solution. We also need, as part of the guidelines, policies that make it possible for individuals to have an opportunity to correct errors in information about themselves.

It is also important to make forms of identification in the physical world more reliable, since the reliability of the identities of people who are the subject of government scrutiny—via an investigation, analysis, or a security checkpoint—is a crucial precondition to the successful implementation of the Task Force's main recommendations. We have identified some problems that currently render the most common forms of identification distinctly unreliable, and recommend both near-term measures and a longer-term research agenda to increase the reliability of identification while protecting privacy (see Appendix A).

Finally, indiscriminate requests for information not only pose risks to civil liberties but also potentially place a serious burden on private sector holders of the information. To the extent that data from the private sector is a "free good" for government, there will be an inherent tendency to overconsume it on the grounds that any information might eventually prove useful. Equally important, a vacuum cleaner approach could actually impede homeland security efforts by inundating the government with information of little or no value, thus complicating analysts' ability to distinguish signal from noise and wasting valuable investigative resources.

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. This requirement should apply in particular where the requests are ongoing, costs are high, and where the cost of complying might put the holder at a competitive disadvantage. The government should enter into an ongoing dialogue with companies that are likely to be the subject of repeated requests and formulate procedures that would minimize the impact on the private sector while assuring that the government is able to access and use 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 sectors on the basis of the scope of the information being requested and the timing and complexity of the request.

At the same time, private sector holders of information 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, we believe that requiring compensation would be inappropriate. In such cases, 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.

Congress plays a critical role in this system of oversight and accountability, and we encourage the development of informal and formal means of congressional oversight of the government's access to, use, retention, and dissemination of private sector data. In addition, we recommend that both the Executive Branch and Congress review agencies' performance in this area, from the perspective of both efficacy and protection of civil liberties. Some proposed metrics to evaluate the government's performance are set forth in Exhibit G. The government could also measure agencies' performance by assessing how well those agencies would do in meeting the challenges set forth in our technology challenge scenarios (see Appendix F) and in our information vignette concerning access to and use of privately held data (see Appendix D).

Exhibit G

EVALUATING IMPROVEMENTS IN THE GOVERNMENT'S USE OF PRIVATE SECTOR DATA WHILE PROTECTING CIVIL LIBERTIES

As with the issue of information sharing among government agencies, we believe the Executive Branch and Congress should evaluate the progress of federal agencies in improving the way they collect, use, and disseminate private sector data while protecting core national values such as privacy and civil liberties. We set forth here some questions that Congress or others may ask after December 31, 2004, to determine whether adequate progress has been made toward the report's objectives.

Question set 1

Has the President issued guidelines for the collection and use of private sector information on U.S. persons? Were these guidelines put out in draft form for public notice and comment?

Question set 2

Has the Executive Branch created a directory that includes all relevant information from both governmental and appropriate private sector databases, and has it made this directory available to all appropriate homeland security players? Has the government made these databases accessible for appropriate rapid, federated searches?

Question set 3

Has the intelligence community implemented an ongoing process for determining intelligence requirements for private sector data? Are the results of the process subject to adequately high-level review and approval? Are intelligence collection priorities adjusted periodically so that they remain in line with these requirements?

Question set 4

Have the DHS and law enforcement agencies developed policies and provided guidance to investigators on when to conduct searches of private sector databases? Do these policies and guidelines address the use of commercial data aggregation services? Do they promote consistency in the use of these searches but remain flexible enough to allow investigators to adjust to the unique circumstances of individual investigations? Do the policies reflect a balancing of investigatory benefits of these searches against the potential negative impact on the privacy of U.S. persons and the private sector's conduct of business? Do the policies include a requirement that the government compensate private sector data holders for the conduct of these searches under some circumstances and provide guidance on those circumstances?

Question set 5

Do government employees who have access to private sector data on U.S. persons for counterterrorism purposes have clear guidelines—that are broadly consistent throughout the government—on the reasons for which they may access this data? Do the guidelines make clear when approval is necessary before accessing data and at what level, and when post hoc reporting and review are sufficient? Do the standards and procedures in the guidelines reflect a balancing of the value of the information sought and the sensitivity of the information? Do the guidelines preclude completely unfettered access by government employees to personally identifiable information on U.S. persons—even if that information is available to the public?

Question set 6

Do government agencies that access private sector data on U.S. persons for counterterrorism purposes have clear guidelines—that are broadly consistent throughout the government—on when and for how long to retain that data? Do the guidelines reflect a preference for keeping data in private sector hands? Do the guidelines contain standards and procedures for when this preference is not followed?

Question set 7

Do government agencies that access private sector data on U.S. persons have clear guidelines—that are broadly consistent throughout the government—on when data collected for counterterrorism purposes may be used to carry out other missions? Do the guidelines disfavor dissemination for non-counterterrorism purposes, except when the agency or unit requesting such data would have been entitled to access the data directly with the same or fewer constraints?

Question set 8
Are there effective mechanisms in place to ensure compliance with the guidelines? Do these mechanisms include rigorous and consistent training on the guidelines, regular auditing and periodic review of compliance, and accountability for failure to comply? Is the approach to oversight consistent—although not necessarily uniform—throughout the government? Is a single agency or entity within the government responsible for ensuring this consistency?

Question set 9
Is the government deploying technology (such as anonymization, access control, and audit technologies) to further the goals of the guidelines? Is the government identifying gaps in existing technologies and those in development? And is it investing in research and development of needed new technologies to protect private sector data from misuse?

Question set 10
Is one government entity (such as the DHS or the OMB) auditing the use of private sector data across all agencies and providing an annual report to the appropriate committee(s) of Congress?

Future work of the Task Force

The Task Force plans to continue its work on the challenges addressed in this report. The current term of the Task Force extends to the summer of 2004, but, given the urgency of the questions we are addressing, we chose to publish an interim report. We will continue to focus on areas that supplement the good work being done by many in the government and the private sector.

We plan to deepen our research on best practices in the government and on how existing technologies and those in development can be deployed to greatest effect. To that end, we hope to develop collaborations in which we pilot the use of technologies (such as information rights management technology, publish and subscribe software, and anonymization tools) to achieve distribution of information with strong civil liberties protections. We also plan to pursue additional work on guidelines regarding the use of private sector data and on new rules for the collection and use of information on U.S. persons to replace the old “line at the border” between domestic and foreign intelligence. New rules and new dynamics between our nation’s security and our civil liberties need a great deal of additional work.

Conclusion

Since September 11, many people in the government and the private sector have given considerable thought and effort to solving the problem of how our nation can use information and information technology more effectively to protect our nation while preserving civil liberties. As sources of relevant information continue to proliferate and technology continues to advance, this challenge will only grow more complicated. Our Task Force has sought to contribute to the solution by providing the framework for a national strategy and an architecture for a decentralized system of robust information sharing and analysis that makes the most effective possible use of information while instituting guidelines and technologies to minimize abuses and protect privacy.

Descriptions of additional papers

Part Two: Working Group Analyses

Working Group I: Networking of Federal Government Agencies with State and Local Government and Private Sector Entities

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

Part Three: Appendices

Appendix A
Reliable Identification for Homeland Protection and Collateral Gains

This paper sets forth near-term recommendations for improving existing forms of identification and an agenda for longer-term research on creating more reliable means of identification while protecting civil liberties.

Appendix B
A Primer on Homeland Security Players and Information

Our primer offers a description of the roles, responsibilities, and authorities of the many different players who are part of the community we seek to bring together in the network, and of the reasons that information often is not shared as fully as it should be.

Appendix C
The Immune-System Model

In considering the issue of information flow among government agencies and, in particular, the problem of potentially flooding the system with too much information, we thought it would be useful to explore different models for how a system might work. One potential model is the human immune system, which is discussed in this paper.

Appendix D
Information Vignettes

Our vignettes describe different types of information that might come into the possession of various entities in the network of governmental and private sector actors. The scenarios allowed the Task Force to consider how such information would be handled today and how it should be analyzed and shared to maximize its utility and to optimize the capabilities of all the players in the network.

Appendix E
The Four Key Questions of Detection and Prevention: Who? How? Where? and When?

This paper describes the four key questions the government must typically answer when trying to thwart an attack on the homeland: Who? How? Where? and When? The model offered in this paper helped us to develop information strategies and identify some information technologies needed to meet the government's security challenge.

Appendix F
Technology Challenges for the Near Future

To understand better the kinds of privately held data that are needed to meet real security challenges, we developed a number of plausible scenarios that government officials might face. These scenarios helped the Task Force to consider what types of information are truly necessary; what technological capabilities the government needs to acquire in order to gain access to the information in a timely, useful way; and what potential civil liberties and other concerns must be addressed by policies governing the circumstances under which the information is acquired and used.

Appendix G
Technologies Required to Meet the Challenges

In "Technology Challenges for the Near Future" we describe 12 scenarios that we used to contemplate the technology and infrastructure issues that need to be addressed to improve national security. In this paper, we reduce the technology requirements to a finite number of specific capabilities. In Section 1, these capabilities are presented alphabetically to enable the reader to quickly look up the description, availability, and best-case time frame for implementation of each capability. In Section 2, we highlight the most critical capabilities.

Appendix H
The Landscape of Available Data

In this table, we present an overview of the data landscape that exists in the private sector. The overview includes data sources, the types of documents that are generated from those sources, the availability of the data, whether the data is personally identifiable, and what entities, if any, currently aggregate or have access to that data. The purpose of this table is to present insight into the types of data that exist as a byproduct of our digital society.

Appendix I
Government Requests for Private Sector Data: An Informal Survey

The purpose of this survey was to get a sense of the kinds of private sector data the government currently seeks for national security purposes, how it seeks that data, and some of the issues the private sector has with government use of its data.

Appendix J
Data Analytics Practices of the Private Sector

In considering how the government could make better use of information technology for counterterrorism purposes, we looked into how the private sector uses data for identity verification, risk assessment, and related purposes. This paper is the result of consultations with representatives of various companies on the use of data analytics in the private sector.

Internet-only information

Matrices of Laws Governing Access to Privately Held Data

A broad array of laws covers how and in what circumstances the government or commercial companies can acquire and use various types of private sector data. We have developed two matrices in which we set forth those laws in an accessible fashion. These matrices can be seen on the Task Force's website at www.markletaskforce.org.

PART TWO
Working Group Analyses

Working Group I: Networking of Federal Government Agencies with State and Local Government and Private Sector Entities

Working Group I is co-chaired by William Crowell and Tara Lemmey. Members of this Working Group are Alexander Aleinikoff, Robert Atkinson, Zoë Baird, Jim Barksdale, Eric Benhamou, Bruce Berkowitz, Robert Bryant, Ashton Carter, Robert Clerman, Wayne Clough, Mary DeRosa, Sidney Drell, Slade Gorton, Lauren Hall, Morton Halperin, Eric Holder, Robert Kimmitt, Governor Mike Leavitt, Terrence Maynard, Mary McCarthy, Judith Miller, James Morris, Craig Mundie, Michael Vatis, Rick White, and Winston Wiley. This paper was written by Michael Vatis.

Introduction

In the Task Force's first report, we sought to provide goals and guidelines for shaping a national security infrastructure that takes advantage of our country's strength in information technology, our understanding of the evolution of networks, and our desire not to choose between security and liberties, including privacy, but to have both. Our recommendations were directed at providing a road map for the development of human and electronic networks that would provide support and encouragement for the development of the new Department of Homeland Security (DHS). More broadly, we hoped to foster the creation of new interactions and understandings that would serve our nation not only in preventing and responding to terrorist threats but also in addressing the day-to-day needs of the many players engaged in protecting the well-being of our society.

Since then, we have consulted with field experts in order to better understand the current state of governmental activity and the successes achieved and barriers encountered to date. The stories and situations related to us by these experts helped us to be informed in our approach. For this paper we employed a technique of scenario-based envisioning. This approach allowed Working Group members and associates to walk through the complex and diverse situations common to homeland security issues with experts in the various arenas. We considered these situations in light of the Task Force's earlier recommendations and then identified the strengths and limitations of current processes.

Our goal was to discover where the present homeland security initiatives are optimized to achieve the dynamic and distributed network required to take on the challenge of distributed and complex threats, and where more work is needed. We wanted to find out how well we are achieving the goal of maximizing the potential contribution of all of the participants in the government's homeland security network, especially those at the state and local levels.

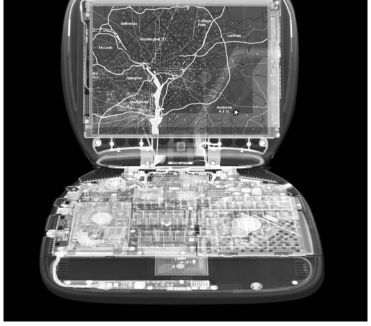
We believe that the more nodes that are interconnected in the network, the more powerful that network will be as an analytical force, as greater numbers of ad hoc groups can come together around matters of common concern, share information, and collaborate in their analyses. In technical terms, the power of the network (as related in Metcalfe's Law[1]) strengthens the value of connectivity as a function of the number of nodes, and the utility of large networks (Reed's Law[2]), particularly social networks, can scale exponentially with the size of the network (number of possible subgroups of network participants is 2^N, where N is the number of participants).

Ultimately, we believe the network the government creates for homeland security should mimic successful private sector networks. That is, it should provide more information on a timely basis to enable players to make better decisions and take more effective actions. In addition, enabling the homeland security participants in this network can give all players more time to potentially prevent terrorist attacks.

FIVE CENTRAL THEMES OF WORKING GROUP I RECOMMENDATIONS

1. **Optimizing for a more distributed, coordinated model**
   Although the need to move from a centralized to a distributed model appears to be widely accepted, there is still a significantly greater degree of centralized control in the government's current approach than we would like to see. This is particularly true with regard to the relation of the federal government to state and local participants. We recognize that moving from centralized control to a coordinated model is difficult, as it reshapes long-standing interactions. But we think it is critical to do this in order to realize the full benefits from the resources at the edges of the network.

[1] According to Metcalfe's Law, the value of a communications system grows as the square of the number of users of the system (N).
[2] According to Reed's Law, the utility of large networks, particularly social networks, can be scaled exponentially with the size of the network.



2. **Redundancy and inclusiveness leading to robustness**
   We would like to see more redundancy of analysis and inclusiveness in the system to cover the seams between actors. The government's current approach is too reliant on conventional coordination and the "sneaker net" (ad hoc coordination through personal relationships) to get information through the system in order to connect the dots. We recommend the implementation of systems of data publishing and subscription that would allow those participants most likely to make important connections to be able to access the relevant information. We would like to empower local analysis of situations. This requires the construction of regional analysis centers and the active engagement of analysts at the state and local level and in the private sector.

3. **Increasing the signal in the data noise**
   Much of the current public discussion focuses on sharing data, but it does not focus on how to get meaningful information from that data (or, to put it another way, to distinguish the signal from the noise). To do this, we would like to see more targeted tasking and specific requests for information and action both from the collectors and experts in the field, as well as from the central coordinating bodies. We also observe the need to create and broaden directories or connecting environments in order to find experts, local specialists, and private sector entities that can be helpful in combating terrorism. We also see the need to explore models that will help provide a framework for information collection and dissemination. One example is the immune-system model (see Appendix C), which the Working Group used as one point of inspiration.

4. **Designing for broad communication**
   There is still a great resistance to broader information-sharing across the network. While we respect the requirement and necessity for compartmentalization to shield highly sensitive information, we strongly recommend that all information collection and dissemination systems be designed to anticipate sharing some form of the information immediately. Our nation's experience during the Cold War revealed that we had to construct systems to encourage and even require the generation of sanitized reports from highly classified reports before information could be released. Analysts have to be trained and directed to create disseminable forms of information regardless of the classification of the source. Moreover, to remove some of the existing disincentives to broad communication, guidelines need to be established to discourage the unauthorized disclosure of disseminated information and uncoordinated action by players in the network.

5. **Setting clear objectives and evaluating performance**
   Many of the organizations and activities responsible for analysis and information-sharing are new and not yet fully staffed. We therefore cannot expect the sort of distributed, coordinated network that we envision to be created overnight. However, we believe it is important for the federal government to set clear objectives that it expects the players in the network to meet, and then to lead a process to evaluate the players' performance in meeting those objectives after enough time has passed for them to have had a chance to make meaningful progress. Such evaluation and continuing oversight are crucial to overcoming agency processes and cultures resistant to change. It is also important for each of the players to have clear rules and guidelines for the behavior and goals of its own personnel.

Background

In our initial report, we emphasized the need for a next-generation homeland security information network that would "empower local participants to contribute, access, use, and analyze data," while also allowing them to "identify, access, communicate with, and assemble other participants in both the public and private sectors" (p. 17).

Also, as we noted, "Most of the real frontlines of homeland security are outside of Washington, DC," and, "Likely terrorists are often encountered, and the targets they might attack are protected, by local officials" (p. 10). In a way, these local actors are critical sensors, capable of detecting the presence or activities of terrorists operating across the country. They are also the most immediate guardians of potential targets. But local officials and organizations can only fill these roles adequately if they know what they should be looking for. Thus, it is vital that the federal government develop the capacity to share terrorism-related information quickly with state, local, and private sector entities in order to optimize these entities' capability to serve as sensors and guardians. At the same time, we need to develop the capacity for these entities to share information with the federal government, as well as with each other, so that all the players have the information necessary to carry out their respective missions.

Information-sharing, however, is easier said than done. For starters, the culture of federal agencies traditionally has been to minimize the dissemination of information and to keep it within a specific domain rather than to share it widely. This has begun to change since September 11, at least in the context of countering terrorism. From our discussions with current and recently retired government officials, Working Group I has learned that more information is being shared among federal agencies than before September 11, particularly between the law enforcement and intelligence communities. This is a significant, and positive, development. However, there are still shortcomings in the sharing of information with local and state agencies.

More fundamentally, though, there are legitimate concerns and values that often inhibit broad sharing of information. These include the need to protect sensitive intelligence sources and methods and individual privacy, and to preserve the ability to effectively investigate, and potentially prosecute, terrorists domestically. Agencies legitimately fear that the more people who see certain information, the more likely it is that the information will be leaked to the media and the public—and to the terrorists—thereby jeopardizing counterterrorism operations or individual privacy. While these interests in the past have been asserted overly broadly as a reason to withhold information from other agencies, they are, at their core, legitimate concerns that must be accommodated before genuine and full information-sharing will take place.

The first objective of Working Group I, therefore, was to determine how concretely we might implement the objective set forth in the Task Force's initial report of creating a networked and nationwide community that maximizes the sharing of information between the federal government, on the one hand, and state, local, and private sector entities, on the other, and that, in doing so, addresses those legitimate concerns that impede information-sharing.

To do this, we felt it was necessary for Task Force members to possess a more complete and up-to-date understanding of the roles, responsibilities, and authorities of the many different players who are part of the community we seek to bring together in the network, and of the reasons that information often is not shared as fully as it might be. To this end, we developed "A Primer on Homeland Security Players and Information" (see Appendix B).

Second, in considering the problem of information flow among government agencies, we thought it imperative to acknowledge the danger of flooding the system with too much information for it to respond effectively. The reality is that every hour of every day, our intelligence and law enforcement agencies, health providers, private companies, and numerous other players receive information that may or may not be relevant to uncovering and preventing a terrorist attack. Were all this information to be disseminated to all the players in the network, the sheer volume of data would create such a high degree of noise and computational complexity that the likelihood of analysts finding useful correlations, or of local agencies taking meaningful protective action, would be virtually nil. On the other hand, of course, fear of flooding can be exaggerated, or used as an excuse to artificially limit information dissemination. The issue is not the amount of information per se, but the need for a system to distinguish signals from noise within the available information.

We therefore thought it vital that we gain a better understanding of how we might distinguish useful signals of potential terrorist activity from useless noise. To help us in this task, we developed "The Immune-System Model" (see Appendix C), which suggests thinking about this problem by way of analogy to our bodies' immune systems.

Third, we thought it important to develop and examine several concrete vignettes that describe different types of information that might come into the possession of one of the players in our nominal network. Through our vignettes we consider the following: (1.) How would the information likely be treated today by its initial recipient? (2.) Where in the system would it be analyzed today in conjunction with other information? (3.) With what other entities would it be shared today? (4.) Where are the roadblocks or speed bumps that prevent or impede necessary sharing? (5.) What additional players should be getting the information today in order to activate all the sensors in the system and increase the intake of relevant information? (6.) How might sanitization or other procedures be implemented so that sensitive information that must be protected (such as sources or methods) can be removed but the rest of the information shared with as broad a group of other players as possible and as quickly as possible? (7.) How can we avoid flooding the system with noise while ensuring that potential signals of terrorist activity are distinguished from the noise and shared widely? To this end, we developed "Information Vignettes" (see Appendix D). These vignettes served as the basis for our discussions and our findings, which are set forth below.

Fourth, we wanted to create visual depictions of how the information in these vignettes flows today, and how it should flow in order to fully activate and utilize all the potential sensors in the network and to maximize information flow. The visualizations help to show the nature of the network we are recommending, the location of roadblocks and speed bumps that need to be removed or smoothed over, and the processes that must be implemented to enhance information flow. The visualizations can be seen at www.markletaskforce.org.

Finally, it is important to keep in mind that our recommendations and information vignettes make certain assumptions about the system that should be in place in order for information to flow in the way we envision. We assume, for instance, that federal, state, and local agencies have the network technology in place to permit the easy and quick dissemination of information to other agencies. In many instances, however, that technology, while available on the market, is simply not in place, and putting it in place is not a trivial matter in terms of cost or time. We will address this reality, as well as integration-architecture solutions that can help. But first we detail the technology that would enable our recommendations to be effectuated fully.

IMPORTANT CAPABILITIES FOR IMPROVING INFORMATION FLOW

1. All players should be linked through a communications system that allows data to be shared.

2. All data on the network should be digital.

3. All data should be portable, using standards, such as HyperText Transfer Protocol (HTTP), Extensible Markup Language (XML), etc., and government agencies should create and incorporate standards for data representation and for the mapping of potential terrorist targets.

4. Data should be blinded, or suppressed from certain viewers or users, thus withholding certain parts of data from actors who lack the necessary permissions.

5. Data should be attached to a reputation system, which, for example, permits ranking of the credibility of the source of the data or the credibility of the analyst who provides the analysis of the data.

6. Data should carry with it a contract for use that specifies the permission level of the actors who can access data. All use of data should be tracked and auditable.

7. Data should carry pointers back to the source, enabling players in the network to contact the source for more information.

8. The system should be set up to be open and available 24/7: In some form, both technological and human components of the system should be accessible at all times.

9. Data should be authenticated.

10. Data should be anonymized when possible, that is, the personally identifiable information should be removed, but analysts should maintain the ability to perform link analysis, queries, and identity resolution.

The DHS, with the force of the President behind it, should serve as a convening authority and bring together the relevant agencies to institute these technologies as rapidly as possible. This would require making these technologies procurement priorities in each of the relevant agencies. Until the technology is in place, the DHS should also promote the use of interim solutions—some involving manual labor instead of technology, others involving technologies that don't require major new investments. We do not believe the government should wait until it has the perfect technology platform in place before it begins to move toward the networked model we advocate. Rather, the government should use a combination of standards that works across systems, policy, and middleware to get the network off the ground quickly.

Discussion

Optimizing for a more distributed, coordinated model

In order to create the sort of decentralized, coordinated network we envision, which fully utilizes the players on the edges of the network, we need to begin with a structure at the federal level that makes the sharing of information with state, local, and private sector entities a central part of its mission. (Structures needed at the state and local level are considered below.) In our initial report, we envisioned that the DHS would play this role. Since the issuance of our initial report, however, the President has announced the creation of the Terrorist Threat Integration Center (TTIC), which is to serve as a center for fusion, sharing among federal (though not state and local) entities, and analysis of terrorism from both foreign and domestic sources. The TTIC reports to the Director of Central Intelligence (DCI), and is an interagency center comprising representatives from the CIA, the FBI, the DHS, and other Executive Branch agencies. The TTIC opened its doors in May 2003, and is now supposed to perform many of the analytical functions that Congress had assigned to (and that our initial report recommended be performed by) the DHS's Intelligence Analysis and Infrastructure Protection Directorate (IAIP).

The TTIC's creation has caused confusion among state and local entities and within the federal government itself about the respective roles of the TTIC and the DHS. This confusion needs to be resolved. We therefore believe the President should clearly define, in an Executive Order, the respective responsibilities of the TTIC and the DHS with regard to intelligence analysis and the sharing of information between the federal government and state and local governments. TTIC officials have described to us how they have set up the organization for information-sharing. As described, the TTIC will obtain information from participating agencies, and if TTIC officials believe that other agencies should see the information, the TTIC will go back to the source and request permission to share it. The source will go through its customary agency procedures to determine whether the information is appropriate to share. We believe that this approach further locks the government into a system that has proven unsuccessful for producing the sharing of information in the past—both because it maintains a centralized approach, and because it does nothing to break down the unwillingness to share. Further, its success requires TTIC analysts (who, we are told, are typically young and inexperienced) to see the connections between different pieces of information or the benefit of that information for other players in the system. If the TTIC is to have principal responsibility for ensuring appropriate sharing of information, it must change its planned method of operating. It must facilitate a decentralized system of information-sharing, in which all participants are expected to share information directly with, and seek information from, others in the network. Otherwise, we cannot support it.

In addition, it is critical that information is shared by the TTIC on both a push and pull basis. Because we cannot expect information consumers always to know enough to request specific information, relevant information should be published and supplied to a list of information subscribers. To minimize information flooding, this information need not be published initially in great detail; the TTIC can simply provide enough directory-level metadata so that subscribers are aware of the available data, and can request further details should they feel it necessary, and only if they have the requisite permissions to see that data. Conversely, because we cannot expect the TTIC always to know when information might be useful to another information consumer, the TTIC should also allow those consumers to seek, or pull, information from the agency.

Finally, if the TTIC is to be a significant center for information integration and analysis, it is vital that the agency hire the most experienced, most highly qualified analysts possible. Moreover, it must provide continuing training and education not only in counterterrorism, but also in how these analysts must work within a network, empowering and optimizing the capabilities of all the other players in the system.

State, local, and private sector interaction

Neither the TTIC nor the DHS nor the FBI appears to be interacting sufficiently with state, local, and private sector entities to initiate the decentralized, coordinated-network that we envision. We believe that the DHS, despite the creation of the TTIC, should have the lead responsibility for ensuring that information from federal agencies (including the TTIC) is shared with state and local government agencies (at least non–law enforcement agencies) and private sector entities in a manner that they can use, and that is targeted to their needs. The DHS should also be the focal point for receiving information from those entities and sharing it with other federal agencies. The DHS should also be responsible for integrating threat and vulnerability analyses, and for determining what protective steps need to be taken, including the securing of potential terrorist targets and the issuance of public warnings.

While we believe that the DHS should be the primary vehicle for sharing federal information with state and local entities, the FBI, apparently, will retain its role as the principal vehicle for sharing information with state and local law enforcement agencies. This will be done primarily through the FBI's Joint Terrorism Task Forces (JTTFs), which are FBI-led groups of federal, state, and local law enforcement agencies that conduct joint terrorism-related investigations. While, once again, we might have preferred to see the mechanisms for sharing information with all state and local entities consolidated within DHS, this arrangement seems workable, and there is logic to having the FBI be the primary vehicle for sharing with state and local law enforcement since these agencies must often coordinate investigations into the same or related targets.

The main problem, however, is that while the FBI apparently shares information regularly with its JTTFs, this does not constitute sharing with state and local agencies. When state and local agencies assign representatives to work on the FBI-led JTTFs, those representatives are not permitted to freely share information from the JTTF with their home agencies without the permission of the FBI. And the FBI apparently is not regularly sharing information directly with the state and local agencies. Therefore, those agencies are not being fully activated as sensors and so are unable to collect and enter into the system relevant information that they might uncover. Because of this, we believe the FBI should develop and implement procedures to ensure the timely sharing of information with state and local law enforcement agencies, not just with JTTFs, so that those agencies might become full players in the network and be better capable of collecting information about terrorist threats. Moreover, if the FBI is to have the lead for interacting with state and local law enforcement agencies, it must also assume the lead responsibility for receiving information from those agencies and sharing it with other federal agencies.

In addition, the FBI could greatly increase the flow of relevant information from JTTFs (and, through them, from state and local law enforcement agencies) to other nodes in the network (including the TTIC and the DHS) if it identified one or more persons in each JTTF whose sole responsibility was to ensure that appropriate law enforcement information was rapidly sent or made available to other players in the network. While such positions are a routine component of many intelligence and military organizations, they are not traditional parts of law enforcement agencies, so the creation of new positions with this role would do a great deal to foster the flow of information throughout the network.

Moreover, it is not enough for the federal government simply to provide more information about threats. Rather, as Information Vignette 3 (see Appendix D) demonstrates, the DHS and JTTFs should, where possible, provide information that is specific and tailored to fit the particular circumstances of each local entity. Moreover, they should also request that state, local, and private sector entities search their records for stored information about those specific threats, and to be on the lookout for any new information about them. Stored information can be critical in providing context and analytical depth to more recently received information on specific threats. This coordinated analysis of stored and more recent information should be structured in such a way as to allow rapid dissemination to, and alerting of, the DHS, the FBI, the TTIC, and other agencies. This would ensure that state, local, and private sector entities are fully activated as sensors, without overburdening them with information that is not relevant to their situations or is too vague to be actionable. The federal agencies should also strive, where possible, to provide concrete guidance as to protective measures that the state and local agencies should take.

Concomitantly, both the DHS and the FBI must build the capability, and instill a culture of willingness, to respond to requests for information from state and local entities. Those entities have their own sense, based on knowledge of their communities, of vulnerabilities and potential threats within their jurisdiction—and they need to be able to tap into the information held by the federal government in order to be effective. Yet too often, state and local entities do not know whom to call to get relevant information. Even more worrisome is the fact that they are often rebuffed when they do know whom to call.

Accordingly, we believe the DHS and the FBI should establish clear mechanisms for responding to requests for threat and vulnerability information from state and local officials, and that they should establish a culture that makes responding to such requests a priority. Establishing a system for rewarding personnel who do a good job of sharing is one way of changing the culture. Ultimately, these sharing mechanisms should be automated and simplified through the use of directories and the publishing of metadata, allowing state and local officials to look for and pull relevant information from federal databases. This would require technology that identifies who has permission to access certain information. It would also require the requisite security and auditing procedures. (The automation of permissioning and auditing of which government officials have accessed which information would also provide a major deterrent to misuse of information. It could also deter leaks, as discussed below.) Agencies should make such technology a procurement priority. In the short term, until the requisite technology is introduced, federal agencies should at least make clear whom state and local agencies can call to obtain information by establishing online directories which, over the long run, can be built into automated systems.

The "line at the border" and the need for transparency

The above measures would do a great deal to create a distributed, coordinated network. But such a network would be inadequate if it did not also protect against encroachment on our nation's cherished liberties. The creation of the TTIC as an all-source intelligence fusion and analysis center with access to both foreign intelligence and domestic intelligence and law enforcement information concerning U.S. persons confronts us with the question of what will replace the previous line (or, the so-called "line at the border") that defines the differential rules for foreign and domestic information collection. Foreign intelligence agencies have traditionally operated abroad with relatively few constraints on their collection activities. Domestic law enforcement and counterintelligence agencies, on the other hand, traditionally have operated under much stricter rules designed to safeguard the rights and liberties of U.S. persons and residents.

Since at least the mid-1980s, with the growth of international terrorism and international narcotics trafficking, the work of foreign intelligence agencies and that of domestic law enforcement and counterintelligence agencies has increasingly overlapped. As a result, these communities have had to work more closely and share more information than ever before. The creation of the TTIC takes this to a new level, and it is imperative that we have an open, public debate about what the new lines are that will replace the line at the border that existed to protect the civil liberties upon which our nation is based. Since September 11, and with legal changes such as the USA PATRIOT Act, which modified or eliminated previous limitations on information-sharing between the law enforcement and foreign intelligence communities (such as limits on the sharing of criminal grand jury or wiretap information with non–law enforcement agencies), a significant erosion of the line at the border has begun without the simultaneous development of a new line (and new guidelines) to protect civil liberties.

It is critical that such guidelines be developed, in order to maintain the trust of the American people. Any perceived misuse of terrorism-related information or inappropriate activity in the domestic realm by foreign intelligence agencies is likely to produce a backlash that will make our recommendations about sharing increasingly difficult to implement. Moreover, even without actual misuse of information, a perceived lack of transparency about the rules fosters visions of a Big Brother government abusing civil liberties.

The information vignettes and network visualizations developed by our Working Group illustrate just how much domestic information the TTIC—and, through it, the U.S. foreign intelligence community—might have access to. Legal issues aside, this raises concerns about the appropriate role of our foreign intelligence agencies with respect to information about U.S. persons that might or might not be relevant to international terrorism. In our information vignettes, for example, the reports about activities of U.S. persons might not ultimately relate to international terrorist plots, yet the information might have been shared with the TTIC and other agencies in the intelligence community (and even with the intelligence agencies of foreign countries).

Recently, too, it was reported that a Department of Defense (DoD) contractor obtained passenger lists for JetBlue flights, and contracted with a data aggregator to run those lists against consumer data (including social security numbers, income levels, number of children, and vehicle ownership), apparently in order to test the viability of passenger profiling. This raises many questions of importance to this Task Force, including: What authority did the contractor or its client agency, the DoD, have to access such data about U.S. persons, and would a higher threshold apply to the collection of such data by a domestic agency such as the FBI? Did the DoD or its contractor comply with relevant provisions of the Privacy Act of 1974 concerning the establishment of systems of records about individuals?

The President should set out, in an Executive Order, clear guidelines governing the TTIC's authority to receive, retain, and disseminate to intelligence agencies (both U.S. and foreign) information gathered in the U.S. about U.S. persons. The Order should also contain guidelines to govern the intelligence agencies' ability to set requirements for (or to "task") domestic collection of information. These guidelines should, to the maximum extent possible, be unclassified and put out for notice and comment so that the American public can have insight and confidence in the way domestic information is collected and used by the government.

Redundancy and inclusiveness leading to robustness

When it comes to intelligence fusion and analysis, the discussion is often cast as a choice between centralizing this function in one or several agencies (in Washington, DC, invariably) and decentralizing analysis among all relevant players. In fact, this is a false choice. We need both centralized and decentralized analysis. We need, for example, an agency like the DHS or the TTIC that is capable of pulling together relevant intelligence and law enforcement information so that it can put together as many pieces of the puzzle as possible and gain a full view of the terrorist threats it is looking at. But we also need other bodies, at the edges of the network, that are capable of gathering pieces together. Redundancy, or complementarity, of analysis is in itself a good thing, given that intelligence analysis is largely a matter of trying to assess the probabilities of connections among people or of events from uncertain facts that are susceptible to different interpretations. In addition, the reality is that the TTIC and a local or state agency might be working on different puzzles, or different parts of the same large puzzle, albeit with some pieces in common. The TTIC might be looking at the activities of foreign terrorist groups and their plots against U.S. interests in general, while a local police agency might be looking at a specific criminal group that is only one small part of a terrorist group. We would not want, nor can we reasonably expect, a single entity to be able to perform both sorts of analysis. These entities also produce a range of different analytical products for a wide variety of audiences or consumers, ranging from strategic intelligence analysis for the President or Cabinet-level officials to tactical leads for the police officer on the street. Our approach to analysis therefore must be inclusive of a range of different actors and analytical centers.

The JTTFs represent one form of decentralized analysis that already exists. As noted, however, the JTTFs are largely limited to law enforcement agencies. Other interdisciplinary analytical groupings should also be encouraged, both among government entities and private companies—such as Information Sharing and Analysis Centers (ISACs), which are discussed below. These groupings should be tied into the network and encouraged to communicate directly with one another as well as with the DHS and the FBI.

In order for these decentralized entities to be a true part of a network rather than becoming their own stovepipes of information, it is critical that they adopt interoperable standards and formats for communicating, storing, and retaining information so that they are able to share easily and quickly with one another. To ease the burden of a massive change of legacy systems, agencies could agree in the interim to publish metadata in a standard format for use in a directory service that points agencies to the holder of specific information they can then request access to. In addition, guidelines are needed that address not only how information should be shared, but also when it should be shared, and with whom. Accordingly, we believe the DHS should work with state and local government entities to create additional decentralized analytical centers, and that it should foster their ability to communicate not only with the DHS and the FBI, but also directly with one another.

Of course, with the creation of new analytical centers, the current shortage of qualified analysts will only get worse. Therefore, part of the effort must include a drive to recruit and train analysts who will have the necessary expertise and skills for the mission. In this regard, it may also be useful to consider federal funding of college and graduate scholarships or grants, potentially tied to a requirement of federal service, for studies in fields at the center of the homeland security intelligence problem. The federal Cyber Corps Scholarship for Service, which provides grants for students who study cyber security in exchange for a promise of federal service after graduation, is one potential model.

Beyond state and local government, private sector entities must also be brought into the network. To date, some industry sectors have formed ISACs for the purpose of analyzing and sharing information among companies and between the sector and the federal government. These ISACs were originally formed to deal with cyber-related information. Since September 11, however, many have broadened their scope to deal with terrorism-threat information as well. But the ISACs have a mixed record when it comes to the amount of information actually shared among companies or with the government. Moreover, existing ISACs are generally limited to critical infrastructure sectors (such as electrical energy, information technology and telecommunications, and financial services). As terrorists increasingly seek out soft targets (sites or events at which they can take innocent lives without great risk to themselves), it is important that the federal government have the ability to communicate quickly and broadly with non-infrastructure companies.

Thus, we believe the DHS should work with private companies to improve the two-way flow of terrorism-related information between government and industry, including with non-infrastructure companies and companies not currently members of ISACs. It should work with existing ISACs to expand their scope beyond cyber threats to deal with terrorism-threat information as well, and to share more information with the government and with other industry ISACs. The DHS should also foster the creation of new ISACs or other mechanisms to bring together non-infrastructure companies that might be targets of attack or that might, in the course of their business, collect information related to terrorism-related activity. The DHS should also work with ISACs to establish information-sharing standards and, where necessary, provide seed funding.

To augment the regional analytical centers and ISACs we also think it would be useful to create and broaden directories, or connecting environments, in order to allow analysts and counterterrorism operators to find experts, local specialists, or private sector entities that can be helpful in combating terrorism. We therefore recommend that the DHS work with state, local, and private sector entities to foster the creation of such directories.

A robust sharing of information must be pursued consistent with civil liberties interests and under strict controls. Predicates for pursuing information must be developed in guidelines, and a rich resource of directories that point to where information can be found must be developed.

Increasing the signal in the data noise

While more information clearly needs to be shared, we must also avoid flooding the network with too much data, causing the real signals to be lost in the noise. If agencies are inundated with too much data, they will be less likely to pay due attention to the most important pieces of information. Judgment as to what is signal and what is noise therefore is, inescapably, required to avoid flooding the system and overloading the sensors. This judgment must be exercised at several different locations: at the agency that originally collects the information and decides what is worth reporting up its own internal chain and to other agencies; at the focal points for intelligence fusion, including the TTIC and the DHS; and at the focal points for dissemination to state, local, and private sector entities.

But because it is not always easy, or possible, to distinguish signal from noise when information is first collected, we must ensure that even when information is not actively disseminated, or pushed, to other entities, it can be easily found, or pulled, by the appropriate agencies when other relevant information comes to light. Given the vast amounts of data that are already collected and entered into the system—and the vast amounts of additional data that must be collected—we cannot rely on analysts to remember all of the information that they have seen, especially information that seemed unimportant at the time. If information is not stored digitally, and if it is not retained, it might as well never have been collected. We therefore urge that standards be developed to ensure that information is digitized, stored, and retained, and that it is searchable at a later date. This applies not only to data that is collected by the federal government, but also to data that is, at least initially, at a state or local agency.

Accordingly, we believe the DHS should convene an expert group from government, industry, and academia to establish common or interoperable data formats and standards for state and local entities so that information they make available to the federal government can be easily stored and searched. Given the financial woes of most states and cities, the necessary upgrade of their information systems is unlikely to take place without federal funding. Therefore, the administration and Congress should fund an initiative to enable state and local agencies to digitize their information as part of a national information-sharing network, and to enhance the security of that information. Because of the potentially enormous expense of this task, the DHS should work with state and local agencies to establish a prioritized list of which records should be digitized first, based primarily on which state and local agencies' records are likely to be most useful to counterterrorism analysis.

Moreover, digitization of information makes security of that information, and of the networks within which it resides, a much higher priority. Therefore, substantial attention should be devoted to improving the security of electronically stored and transmitted information.

In addition, the Office of Management and Budget (OMB) should establish clear rules for federal acquisition and use of information in distributed databases, based on the principle of a clearly authorized purpose, consistent with the acquiring organization's mission. The rules should address when and under what conditions federal agencies can access such information, what information they can retain, and for how long. Regarding retention, federal agencies will have little capability to maintain large volumes of information in current and corrected form, or to determine its validity over time. It may, therefore, be better for those agencies to retain pointers and directories to information of value rather than retaining the information itself, so that they can return to the information when needed, and allow the holder of the information to refine and update it over time. These pointers could be available across a broader cross-section of the analytic community. For example, federal agencies such as the DHS, the TTIC, or the FBI may need to retain, in a central database, some information about persons on watch lists, but the broader set of data from which that information may have been obtained, and which is not immediately relevant for watch list purposes, should not be retained centrally and should be accessed only when needed. (See "Working Group II: Building an Effective, Sustainable Partnership Between the Government and the Private Sector," which addresses the issue of federal agency access to, and retention of, privately held information in detail.)

Designing for broad communication

One of the principal reasons information is not shared more widely with state and local governments and with private sector entities is fear that the information ultimately will be leaked to the media and the public, jeopardizing intelligence sources and methods, compromising law enforcement investigations or prosecutions, or violating individual privacy interests. These are legitimate concerns. But these concerns can be ameliorated if federal agencies put in place regular processes for producing information in a way that it can be shared even if it comes from sensitive law enforcement or intelligence sources. With these new needs, new methods of creating documents must quickly be developed so a version is created at the outset that can be shared more broadly. Instead of a culture of classification, we need a culture of distribution, in which the rewards go to those whose information has been found most valuable by people across the network. We need to reward those who figure out exactly what it is that others in the distributed system need to see, and who make sure they get it in a form they can use.

In government today, agencies focus on sanitizing the information. Sanitization occurs when an agency removes from a report any sensitive information that it believes cannot be shared widely with other agencies without undue risk to sources and methods or some other legitimate interest, but provides the gist of the information so that recipient agencies can take appropriate investigative or protective actions or utilize the information in their analyses. Currently, as our information vignettes demonstrate, some federal agencies perform some sanitization to remove source and method information. (The NSA, for instance, regularly produces tear-line versions of reports, with a "Top Secret" version disseminated on paper to a small group of recipients, and a "Secret" version disseminated to a broader group electronically.) But even the sanitized version of information is often still classified, and is usually designed for dissemination only to other federal agencies. Sanitization does not generally occur as a matter of course for many agencies, and few agencies regularly produce a sanitized version of information that is unclassified and appropriate for wide-scale dissemination to state, local, and private sector entities. The sanitization process is cumbersome and takes time. To foster additional information-sharing then, new processes for conveying information in a manner that doesn't reveal sources, methods, or when necessary, sanitization, need to become a regular part of the process.

In addition, it is important to develop mechanisms for producing disseminable versions of information without losing data fidelity or making the information so general or vague as to be useless. The use of anonymization techniques and pointers, which direct the recipient agency to the person or organization from which more information might be available if the recipient has the right permissions, can help address this problem.

The most logical place for these processes to occur initially is with the agency that collects the information. But one agency, the DHS, should review these decisions and have the authority to serve as the backstop, the guarantor that as much information as possible is being shared, in a process to resolve disputes between its view of the need to share the information and the originating agency's desire to prevent it.

We believe that all federal agencies responsible for collecting terrorism-threat information should see state and local government agencies and, in some instances, private sector entities, as regular consumers of that information, and should produce unclassified reports of relevant information that can be disseminated to all these entities. If the originating agency does not believe any such dissemination is possible without causing undue risk of damaging counterterrorism operations, it should have to note in writing, and with specificity, why this is the case, and provide that written explanation to the DHS. This would serve to encourage the production and dissemination of reports and foster effective oversight.

Furthermore, we believe the originating agency should provide copies of both the classified and unclassified versions of the information to the DHS and to the FBI, so that the DHS and the FBI can disseminate the unclassified versions to state and local agencies and private sector entities, as discussed below. The DHS should be responsible for disseminating versions of threat reports that it receives to non–law enforcement agencies at the state and local level and to private sector entities. The FBI should be responsible for disseminating such reports to state and local law enforcement agencies.

While we believe this process would greatly foster the dissemination of actionable information to, among others, state and local agencies, it does not entirely eliminate the risk of unauthorized disclosure and the harm that such disclosure can cause. Even sanitized information could, if made public, reveal important clues about the state of the government's knowledge about a terrorist group or plot.

Moreover, as information is shared among agencies with overlapping jurisdictions, there is a risk that uncoordinated action by one agency in response to that information could impede or disrupt a sensitive counterterrorism operation by another agency. If one federal agency, for example, shares information about a terrorist group that it has been investigating undercover for a long time, and a second agency undertakes its own investigation of that group, the second agency's actions could disrupt the first agency's investigation and cause the loss of vital intelligence.

Finally, a recipient of information not suitable for public disclosure—for example, information of uncertain credibility about a potential terrorist threat to a landmark or infrastructure asset—could take action or make public statements that cause undue public alarm if the threat turns out to be unfounded. Additional measures must therefore be taken to minimize the risk of unauthorized disclosure of information and ensure coordination by recipient agencies before information is acted on.

These problems are not susceptible to easy answers. Auditing technology could be deployed to track the flow of information to different players, thus deterring leaks. The auditing tools should use strong means of authentication that have forensic value—that is, they can be used in court to prove access. There should also be a tool that summarizes the information that was accessed and the actions taken with the information (for example, if it was printed, forwarded, or edited). But this technology would be of little help in tracing an unauthorized disclosure if information is leaked orally. Another possibility would be to make certain federal funding for information-sharing purposes contingent on the adherence to certain rules prohibiting unauthorized disclosure. But efforts to enforce such rules would almost certainly meet political resistance when agencies are threatened with the loss of funds. A third potential mechanism would be the establishment of "deconfliction" centers (populated by representatives of relevant agencies) that would ensure the coordination of investigations and operations by multiple agencies at various levels of government. While none of these measures is perfect, a combination of such efforts might reduce the chance of unauthorized disclosure or uncoordinated action, and thereby foster a healthy environment for the sort of broad communication that we envision. Accordingly, we recommend that the DHS lead an interagency group to study and institute a variety of such mechanisms —including both technology and policy.

Setting clear objectives and evaluating agency performance

In all of the preceding recommendations, we urge the implementation of certain measures by various entities to improve information analysis and sharing among the many players in the network. Many of these entities—such as the DHS and the TTIC—are new. Other, established entities are being expanded or given new responsibilities. The newness of these agencies and responsibilities makes the work we have set out all the more challenging: Agencies need to hire personnel, implement procedures, and acquire technology to perform their new missions. At the same time, though, this newness presents an opportunity for the Executive Branch and Congress to evaluate how federal agencies perform the important tasks discussed above, and to make any necessary changes in the division of responsibilities before agency roles and missions become so entrenched as to make major adjustments politically or administratively difficult or infeasible. The Executive Branch could also work with state and local governments and private companies to evaluate their performance, and should take appropriate steps to encourage any necessary improvements.

In order to evaluate agencies' performance, we believe the Executive Branch should set forth specific and clear objectives for improved analysis and information-sharing, based on the recommendations above, which each federal agency should be required to meet by December 31, 2004. At the conclusion of this period, the Executive Branch and Congress should evaluate how agencies have performed in meeting those objectives. If an agency has not performed adequately, the Executive Branch and Congress should consider making any necessary changes.

We also think the DHS should include state and local government entities in a regular process for assessing how well information is being shared with them, akin to the process the intelligence community currently uses for having customers of intelligence evaluate collectors. Concomitantly, the DHS should work collaboratively with state and local governments and private sector entities to set objectives for them to meet in analysis and information-sharing as well, and it should work with them to jointly evaluate these entities' performance after December 31, 2004, and thereafter on an ongoing basis.

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.

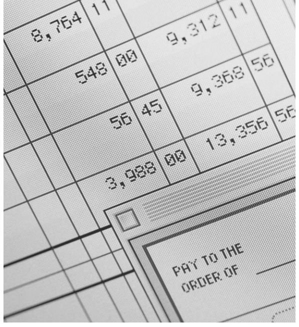
In our initial report, we stated, "The government will need access to public and private sector data for national security. The Department of Homeland Security (DHS) should develop innovative service delivery models for using information held within and outside government (on trade or specific cargo, 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 counterterrorism 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 counterterrorism 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 Section 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.

PART THREE
Appendices

Appendix A

Reliable Identification for Homeland Protection and Collateral Gains

This paper is presented by the Subgroup on Reliable Identification for Homeland Protection and Collateral Gains, which is chaired by Amitai Etzioni. Members of the subgroup are Robert Atkinson, Stewart Baker, Eric Benhamou, William Crowell, David Farber, Mary McKinley, Paul Rosenzweig, Jeffrey Smith, James Steinberg, Paul Schott Stevens, and Michael Vatis.[1] This paper was written by Amitai Etzioni.

Executive summary

There is strong evidence that having reliable means of personal identification would greatly enhance many of the new security measures introduced since September 11, as well as those that were in place before the attacks. Despite some attempts to make our means of identification more reliable, many of those routinely used in the U.S. are still highly unreliable. We realize that means of identification cannot be made foolproof. However, we believe that very substantial improvements can be made that will greatly enhance our security and that the improvements will have what we call collateral gains (advantages in treating other serious national problems).

In this paper, our focus is on developing purposeful means of identification—issued by governments and by private industry—that can enhance our ability to resolve an individual's identity. Identity resolution should be balanced with the need to maintain accuracy and liability for the principal uses of the means of identification, even though the means of identification may still be used for purposes beyond what was authorized by the issuer.

The members of our subgroup come from different backgrounds. Some have held positions in federal agencies such as the Department of Justice (DOJ), the Federal Bureau of Investigation (FBI), the National Security Agency (NSA), the Central Intelligence Agency (CIA), the Immigration and Naturalization Service (INS), and the Department of Defense (DoD). Some of us are privacy advocates, some elected officials, some policy researchers, some CEOs of high-tech companies. But we all agree that it is necessary to make means of identification more reliable, especially those used in high-security, high-risk areas. We do not call for the introduction of national ID cards; rather, we call for making the multiple means of identification people use when seeking to enter controlled areas—such as airplanes, buildings, and, for incoming immigrants and foreign visitors, the U.S.—more reliable. We believe that we should not rely on any one means of identification, but rather that multiple means of identification are needed, depending on the purposes at hand and the desired level of security. We believe that as the security level of the purpose increases, so too should the reliability of the identification.

Recommendations

We recommend that the Department of Homeland Security (DHS) form a task force whose purpose will be to examine proposals (ours and others) for making the means of identification used within the federal government's jurisdiction (transportation security, border security, immigration, and critical-infrastructure protection) more reliable and to implement—or foster, when the authority for implementation is outside the agency's domain—the needed measures along the lines detailed in this report.

An interagency task force for reliable identification, led by the Office of Management and Budget (OMB), should also be formed. This task force should be composed of representatives of the DHS, the National Institute of Standards and Technology (NIST), the Department of Treasury, the Department of State, the FBI, the NSA, the DoD, and the Department of Transportation, among other agencies, and should examine how these agencies' programs are affected by technical or process issues regarding current means of identification. The task force should collaborate with the DHS in identifying ways to make the means of identification used by all elements of the government, and for privately owned critical infrastructure, more reliable.

Governmental remedies

Because of the severity and urgency of the situation, short-run measures should be introduced first. Meanwhile, more reliable means of identification, which will have longer implementation times, will be studied to determine whether they may later be put into practice. For governmental remedies, we recommend a two-phase process for making more reliable means of identification. The first phase should focus, albeit not exclusively, on how

[1] We are indebted to Deirdre Mead for her extensive research assistance and to Dennis Bailey, Jerry Berman, Marc Dunkelman, Shane Ham, Lara Flint, Joanna McIntosh, Neville Pattinson, Ari Schwartz, and Tom Wolfsohn for their valuable suggestions.

the federal government can assist in making state driver's licenses and state-issued identification cards more reliable as quickly as possible, as they are the most widely used forms of identification in the U.S. However, some of our recommendations will help make other means of identification—such as passports and visas—more reliable as well.

The second phase should initially focus on studying whether biometric and cryptographic technologies may be used to make driver's licenses and other forms of identification more reliable, and on determining which technology, if any, is appropriate and how the technology and enrollment processes may be implemented, given the primary purposes and uses of these means of identification. These studies should also address ways in which we can protect privacy and civil liberties while achieving more reliable means of identification. If an appropriate technology is identified, the technological wherewithal is available, enrollment processes have been carefully refined, and privacy concerns have been addressed, biometrics might be added to driver's licenses and other means of identification. Finally, some of us believe that a pure biometrics system may, in the long run, be preferred; others feel this idea is highly dubious and subject to error or fraud in the base technologies, the enrollment processes, or the people implementing the processes. Hence, at this stage, pure biometric technology should merely be studied.

All improvements to our means of identification require attention to the following three elements: (1.) the processes (the enrollment process, higher levels of validation, network verification of the information on a form of identification, and the introduction of audit trails); (2.) the personnel (improved training, selection, and oversight); and (3.) the technologies involved (biometrics; smart cards; scanning devices to verify the information on the card against information on the network; cryptography, etc.).

RECOMMENDED MEASURES TO MAKE DRIVER'S LICENSES MORE RELIABLE

Phase One

1. The federal government should conduct research on affordable methods of improving identification systems and making the entire identification mechanism more verifiable. The government should encourage states to implement the studies' findings and adopt interstate standards, and to put them into practice through the use of grants.

2. In each of the jurisdictions, the fines and penalties for individuals who possess, attempt to obtain, or sell counterfeit or false identification should be increased, as should the fines and penalties for individuals who knowingly supply such identification or knowingly allow people who are using it to enter controlled areas.

Processes

1. Paper breeder documents should be standardized.

2. Birth- and death-certificate records should be digitized and searchable in all states. One existing program that addresses this need and therefore deserves further support is the E-Vital program, which establishes a common process through which birth- and death-record information can be analyzed, processed, collected, and verified. Yet we believe that the holder of such data should have privacy-protection measures and enforcement policies in place that address issues such as who may access the data and for what purposes. For instance, to protect civil liberties, audit trails should be established.

3. States should verify that the social security number a person presents when applying for a driver's license is not someone else's. The Department of Transportation should develop an approach to providing the needed funds so states will be encouraged to undertake this verification step.

4. Federal legislation should tie the expiration date of a driver's license or state-issued identification card to the expiration dates of a foreign visitor's visa, as some states have already done.

5. State driver's licenses and identification cards should meet minimum uniform standards concerning their data content and the verifiability of the credential.

Personnel

1. State motor vehicle agencies should provide their employees with ongoing, detailed training in how to spot counterfeit or false documents. They should also provide law enforcement personnel with guidelines on how to check the validity of driver's licenses.

2. State motor vehicle agencies should launch aggressive oversight, auditing, and anticorruption policies to help prevent fraud and to make it easier to detect fraud when it occurs in the driver's license issuing process.

Phase Two

Technology

We suggest the need for various studies. These would best proceed on two levels: (1.) meta-analysis, overview, and codification of what is known (the results of various ongoing studies in the private sector and in the government); and (2.) the issuance of Requests for Proposals (RFPs) to invite additional studies that would cover well-known lacunae or those lacunae the analysis of the first level—the summaries of the state of the art—would reveal.

Private sector remedies

We believe private sector alternatives to making means of identification more reliable should also be examined. DHS officials should convene a panel of representatives from corporations to determine incentives that would encourage the private sector to develop for use various purposeful cards (credit cards, medical cards, etc.) that are more reliable and verifiable—for example, those incorporating, on a voluntary basis, the use of pictures or biometrics. Among the ideas to be examined is whether such cards could be used to provide secondary verification of identity.

Accountability and privacy protections

Concerns about privacy and other civil liberties should be addressed in all matters, including all studies, regarding the development of more reliable means of identification. For personal data, such as digitized birth- and death-certificate records, we emphasize that those who hold the data should have privacy-protection measures in place to address issues such as who may access the data and for what purposes. There must also be enforcement policies. For instance, audit trails, which could detect unauthorized use of data and thus help deter it, should be established.

We also recommend that the DHS set up a body of public and private sector members to review proposals and measures regarding more reliable means of identification for homeland security purposes. This body should also examine the measures' effectiveness and their privacy implications. It should operate under the criteria specified in the Federal Advisory Committee Act.

Reliable identification is essential to homeland protection

The prevalence of means of identification that are readily falsified or are obtained in a fraudulent manner is a particular vulnerability of our homeland security. Unless substantial improvement is made in this area, many new systems—and many other programs that help protect the public—will continue to be severely hampered. These include the foreign student tracking system (SEVIS) and the National Security Entry-Exit Registration System (NSEERS)—both of which will eventually be part of the U.S. Visitor and Immigrant Status Indication Technology program (U.S. VISIT)—as well as the Computer Assisted Passenger Prescreening System (CAPPS II) and current watch lists maintained by the FBI, the CIA, and the Bureau of Citizenship and Immigration Services (BCIS).

Press reports suggest that the reluctance of the White House and Congress to deal with the means by which people are identified stems from concerns that an action taken in this area would entail the introduction of a national ID card,[2] which faces strong opposition from the left and right and from much of the U.S. citizenry. We cannot stress strongly enough that we are not recommending such a course of action. Our concern here is with what we call purposeful means of identification: means issued by governments and by private industry for specific purposes. People are not required to carry these means of identification with them at all times and to show them upon demand, as is the case with national ID cards used in other countries, such as Belgium and Spain.

Many different types of purposeful means of identification, not necessarily cards, are used by people seeking access to controlled areas—airplanes, secure facilities, most public buildings, and numerous private ones. For the 40 million foreigners who travel to the U.S. each year for vacation, to attend school, or to conduct business, the U.S. itself is a controlled area.

In addition, we believe that our country should not rely on any one means of purposeful identification, but rather that multiple means of identification are needed, depending on the purposes at hand and the desired level of security. The credentials required to obtain a library card at a local public library, for example, should be less than those required of someone who will be responsible for transporting haz-

[2] See, for example, Mark Helm, "As Term Nears End, Armey Not Afraid to Speak His Mind," *Washington Post*, 18 August 2002, A7; Bill Miller, "Homeland Security Cost Weighed," *Washington Post*, 17 July 2002, A8; and Bill Miller and Juliet Eilperin, "House GOP Leaders Unveil Homeland Bill," *Washington Post*, 19 July 2002, A4.

ardous materials across the country. We believe that having more than one identifying document is necessary for the protection of privacy and civil liberties and, furthermore, that relying on any single document for identification makes the system more easily manipulated by terrorists.

Next, as we illustrate here, many of the means of identification routinely used in the U.S. are still highly unreliable. Deliberations about ways to improve them often focus on technical aspects only. Yet all three elements of how individuals are identified—the processes, personnel, and technologies involved—need to be stressed and improved.

We realize that means of identification cannot be made foolproof, but we believe that very substantial improvements can be made and that these will greatly enhance our security. These improvements will have what we call collateral gains, or advantages in treating other serious national problems.

Unreliable means of identification severely hamper homeland security

We next present evidence in support of our observation that, despite some recent improvements, the prevailing means of identification—which are commonly relied upon in the U.S.—are woefully inadequate.

The 100-percent failure rate of border security

Robert J. Cramer, managing director at the General Accounting Office's (GAO) Office of Special Investigations (OSI), reported to our group the results of an investigation the GAO conducted between September 2002 and May 2003: In every instance when agents attempted to enter the U.S. from Western Hemisphere countries using counterfeit driver's licenses and birth certificates with fake identities, they were successful. The border-patrol agents, without exception, failed to realize that the documents were not authentic. For these security tests, OSI agents used widely available computer-graphics software—which can be found in most average homes—to create counterfeit documents.

In the course of this investigation, OSI agents used counterfeit documents and false identities to enter the U.S. from four countries. It is important to keep in mind that U.S. citizens—or people claiming to be U.S. citizens—seeking to enter the U.S. from Western Hemisphere countries are not required to show a passport. Instead, they are required to prove U.S. citizenship. This may be done through a state-issued birth certificate or a baptismal record, and photo identification—for instance, a driver's license. Or, as the GAO states, "Since the law does not require that U.S. citizens who enter the U.S. from Western Hemisphere countries present documents to prove citizenship, they are permitted to establish U.S. citizenship by oral statements alone."[3] Teams of two OSI agents tried to enter the U.S. from Canada three times, from Mexico two times, from Jamaica one time, and from Barbados one time. Each time, agents were able to cross the border—whether at an airport, a land-border crossing, or a seaport of entry—when border-patrol agents failed to recognize that the documents were counterfeit.[4]

Federal buildings and airports are highly porous

In April and May of 2000, the GAO's OSI agents tried to gain access to 19 federal buildings and two airports using counterfeit law enforcement credentials that were either acquired from public sources or created using commercial software packages, information from the Internet, and an ink-jet color printer. Agents gained entry to 18 of the 21 sites on their first attempt; the other three sites were successfully entered on the second attempt. The buildings to which the agents gained entry included some of the most sensitive, and presumably most secure, facilities in our country: the CIA headquarters, the Pentagon, the FBI headquarters, the Department of State, the DOJ, and others.[5]

Upon entering these buildings or the airport terminals, the undercover agents carrying counterfeit credentials declared that they were armed law enforcement officials. They passed through security without being screened, even though one agent always carried a valise. Robert H.

[3] Prepared Testimony of Robert J. Cramer, managing director, OSI, GAO, before the House Judiciary Subcommittee on Immigration, Border Security, and Claims on Counterfeit Documents Used to Enter the U.S. From Certain Western Hemisphere Countries Not Detected, 108th Cong., 1st Sess., 13 May 2003 (GAO-03-713T).

[4] Ibid., and Prepared Testimony of Robert J. Cramer, managing director, OSI, GAO, before the Senate Committee on Finance on Weaknesses in Screening Entrants Into the U.S., 108th Cong, 1st Sess., 30 January 2003 (GAO-03-438T).

[5] Prepared Testimony of Robert H. Hast, assistant comptroller general for investigations, OSI, GAO, before the House Judiciary Subcommittee on Crime, on Breaches at Federal Agencies and Airports, 106th Cong., 2nd Sess., 25 May 2000 (GAO/T-OSI-00-10).

Hast, assistant comptroller general for investigations with the OSI, reported, "At the 21 sites that our undercover agents successfully penetrated, they could have carried in weapons, listening devices, explosives, chemical and biological agents, devices, and/or other such items or materials."[6]

Another troubling finding was that at 15 of the 16 facilities, agents were able to stand directly outside the suites of agency heads and cabinet secretaries. The five times agents attempted to enter the suites, they were able to do so. Undercover agents also were able to enter restrooms near the agency heads' or cabinet secretaries' suites, where they could have left dangerous materials without having been detected.[7]

Airport officials did not detect the counterfeit documents either. Airline ticket agents readily gave the undercover OSI agents law enforcement boarding passes, and although the procedures for getting through security varied at the two airports, none of the agents nor their valises were screened by security personnel.[8]

In response to these findings, 19 of the 21 agencies and airports that were part of the original GAO study responded that they had taken specific actions to enhance their security.[9] However, a task force investigation into Washington, DC–area airports in 2001 revealed that those airports' general security systems remained lax. The task force, formed by U.S. Attorney Paul McNulty of the Eastern District of Virginia, examined the records of airport employees who held Security Identification Display Area badges, which allow access to secured areas of Dulles International and Reagan National Airports.[10] As McNulty reported to the House of Representatives, the investigation found that "75 airport workers used false or fictitious social security account numbers to obtain security badges, and that afforded them unescorted access into the most sensitive areas of our airports." He went on to say, "Many of these airport workers also used the same false or fictitious social security number to obtain Virginia driver's licenses, fill out immigration forms, or apply for credit cards."[12]

The Washington, DC–area airports were not the only ones at which individuals used fraudulent means of identification to obtain security passes. After the September 11 terrorist attacks, an investigation, directed by the DOJ, into employees at the Salt Lake City International Airport found that "61 individuals with the highest-level security badges and 125 with lower-level badges … misused social security numbers" to obtain security badges or fill out employment-eligibility forms.[13]

Military facilities are like an open book

When the GAO's OSI agents used false means of identification—a fake ID card from a fictitious agency within the DoD—they were able to enter areas controlled by the military (areas in which weapons are stored between stages of transport across the country). Moreover, the undercover agents were allowed unhampered access to the weapons themselves. This observation is based on OSI Managing Director Cramer's report to our group. The GAO report on this matter has apparently proven either so damaging to national security or so embarrassing to the government—or both—that it has been withdrawn from circulation.

Fraudulent documents are used to enter the U.S.

INS officials intercepted more than 100,000 fraudulent documents each year between fiscal years 1999 and 2001. These documents included border-crossing cards, nonimmigrant visas, alien-registration cards, and U.S. and foreign passports and citizen documents, as well as other documents.[14] While every intercept of a fraudulent document is a success, many are not caught. This is evidenced

[6] Ibid.

[7] Ibid.

[8] Ibid.; and Letter from Robert H. Hast, managing director, OSI, GAO, to the Honorable Lamar Smith, Chairman of the House Judiciary Subcommittee on Crime regarding Security Improvement Inquiry, 31 August 2001 (GAO-01-1069R).

[9] Letter from Hast (GAO-01-1069R). One agency, the CIA, did not provide a specific response to the inquiry, and the other agency, the U.S. Courthouse and Federal Building in Orlando, Florida, was not part of the follow-up. However, the GAO reports it contacted the U.S. Marshals Service and the General Services Administration, which are responsible for the security of judicial facilities and federal buildings.

[10] DOJ Press Release, "Attorney General Statement Regarding Airport Security Initiative," 23 April 2002. Available at http://www.usdoj.gov/opa/pr/2002/April/02_ag_246.htm. Accessed 25 June 2003.

[11] Prepared Testimony of Paul J. McNulty, U.S. Attorney for the Eastern District of Virginia, before the House Judiciary Subcommittee on Crime, Terrorism, and Homeland Security and the Subcommittee on Immigration, Border Security, and Claims, 107th Cong., 2nd Sess., 25 June 2002.

[12] Ibid.

[13] Office of the Inspector General, Social Security Administration, *Social Security Number Integrity: An Important Link in Homeland Security*, Management Advisory Report, May 2002 (A-08-02-22077).

[14] Prepared Testimony of Richard M. Stana, director, Justice Issues, GAO, before the House Judiciary Subcommittee on Crime, Terrorism, and Homeland Security and Subcommittee on Immigration, Border Security, and Claims on Identity Fraud, 107th Cong., 2nd Sess., 25 June 2002 (GAO-02-830T).

by the fact that in a 20-month period between October 1996 and May 1998, the INS reported that "about 50,000 unauthorized aliens were found to have used 78,000 fraudulent documents to obtain employment."[15]

We cannot assess to what extent this problem has been alleviated since September 11, but the following reports suggest that it is far from resolved. Raids in the Seattle area in September 2002 netted enough computer equipment and specialty paper to print more than 800 fraudulent documents, including driver's licenses, social security cards, green cards, and Mexican driver's licenses.[16] In Washington, DC, raids resulting from an ongoing investigation, which began in April 2002, have netted more than 1,000 fraudulent documents and nearly 50 arrests.[17] In one bust during this ongoing investigation, authorities confiscated more than 500 fake residency cards, social security cards, driver's licenses, and other IDs at a single residence. Cynthia O'Connell, acting director of the Identity Fraud Unit of the Bureau of Immigration and Customs Enforcement (BICE), reported in August 2003 that "there are not enough agents to do it all, especially after September 11."[18]

Terrorists use them too

Many of the September 11 hijackers and their associates have been found to have used counterfeit social security numbers (ones that were never issued by the Social Security Administration [SSA]). Meanwhile, one of the hijackers used the social security number of a child, and other hijackers used numbers that had been associated with multiple names.[19] This fake or counterfeit information seems to have been used by the hijackers to obtain driver's licenses. Some of the hijackers held multiple licenses from states including Virginia, Florida, California, Arizona, and Maryland. Only one of the hijackers appears not to have possessed a state-issued form of ID, according to Senator Richard Durbin at his hearing on driver's licenses in April 2002.[20] It should be noted, too, that Timothy McVeigh used a fake ID to rent the Ryder van that exploded in front of the Murrah Federal Building in Oklahoma City in April 1995.[21]

In short, the urgent need for more reliable means of identification for homeland security is evident. Our current means of identification are inefficient, tedious, and labor intensive. They impose a nontrivial transaction cost on ID verification. A side effect of this inefficiency is that we cannot verify IDs as often as we need to—or as often as we should—and this makes our current means of identification less effective than they should be.

New security measures and systems presume reliable means of identification

In other papers included in this report, we suggest an array of new measures to improve our homeland security. And in the Task Force's first report, we called on analysts to conduct wide scans to identify vulnerabilities and to utilize that knowledge to focus on known concerns. This process includes identifying potential targets and the means that could be used to attack them, as well as analyzing information about individuals and groups of people (including their goals, capabilities, and networks) who pose a threat to our country.[22] Reliable means of identification are necessary for the analysts to identify the individuals and groups who pose a threat to homeland security.

Again, we stress that we are not claiming that more reliable means of identification would solve all security problems. Nor are we implying that because false IDs can be readily obtained, it is impossible for law enforcement to find wanted terrorists. We do not claim that new systems, such as SEVIS and U.S. VISIT, or older systems, such as watch lists, are blind. We merely state that these systems would become much more effective if the processes of issuing IDs, and the technologies used to issue them, were substantially improved.

[15] Ibid.

[16] Diane Brooks, "Raids Net Pile of Fake IDs," *Seattle Times*, 14 September 2002, B1.

[17] Warren A. Lewis (interim director, Washington District Office, BICE, DHS), letter to the editor, *Washington Post*, 17 May 2003, A24.

[18] Mary Beth Sheridan, "Raids Don't Stop D.C. Street Trade in Fake U.S. IDs," *Washington Post*, 3 August 2003, A1.

[19] Prepared Testimony of James G. Huse, Jr., inspector general, SSA, before the House Judiciary Committee's Subcommittee on Crime, Terrorism and Homeland Security and Subcommittee on Immigration, Border Security, and Claims, 107th Cong., 2nd Sess., 25 June 2002.

[20] Statement of Senator Richard Durbin before the Senate Governmental Affairs Committee's Restructuring and the District of Columbia Subcommittee on Fake or Fraudulently Issued Driver's Licenses, 107th Cong., 2nd Sess., 16 April 2002.

[21] Ibid.

[22] Markle Foundation Task Force, *Protecting America's Freedom in the Information Age: A Report of the Markle Foundation Task Force* (New York, NY: Markle Foundation, October 2002), 25–26.

Driver's licenses are still the weakest link in a weak chain

Driver's licenses and state-issued identification cards are classic examples of multipurposeful means of identification that deserve special attention. The vast majority of Americans over 16 years of age possess a driver's license, one of the few identification documents that is widely accepted as proof of ID or age. When boarding a plane, cashing a check, purchasing alcohol, or conducting similar activities, most Americans are asked to show ID. Other forms of ID, such as passports or military IDs, are held by much smaller segments of the population. When asked for identification, most Americans present a driver's license. Of course, driver's licenses are not created for this purpose, and their reliability level is inadequate for the security uses for which they are commonly employed.

Before September 11, it was very easy to obtain driver's licenses in the U.S. using false or counterfeit documents, although it was more difficult in some states than in others. One could even purchase a counterfeit driver's license on the street or on the Internet. Terrorists took advantage of these weak documents. Seven of the September 11 hijackers obtained Virginia driver's licenses by submitting false information to prove residency in the Commonwealth. The hijackers (and surely many others) took advantage of the fact that proof of residency could be obtained with a notarized affidavit from another Virginia resident. According to Paul J. McNulty, two of the hijackers paid an illegal immigrant $100 to vouch for their residency.[23]

In short, driver's licenses and state-issued identification cards for nondrivers are being very widely used for security-identification purposes—to board airplanes and to enter public and private buildings, including legislatures, courts, government agencies, and numerous corporations. Yet driver's licenses are still a very unreliable means of identification. Since September 11, the stakes involved in having a reliable ID system have been raised significantly.

Some loopholes have been closed in the wake of the September 11 attacks (for example, in Virginia, the notarized affidavit was taken off the list of acceptable documents for proof of residency; and in Florida, Governor Jeb Bush ordered that driver's licenses for foreigners expire at the same time as their visas), but false driver's licenses can still be obtained easily.

Between July 2002 and May 2003, the GAO's OSI agents conducted security tests in seven states and in Washington, DC, to determine whether state motor vehicle agencies would issue driver's licenses to applicants who presented counterfeit "breeder" documents,[24] such as counterfeit birth certificates, driver's licenses, and social security cards. As with the other GAO investigations, undercover OSI agents created fictitious identities and counterfeit documents using off-the-counter computers, printers, and software. The investigation found that department of motor vehicles (DMV) officials generally did not recognize that the documents they were presented were counterfeit. Therefore, DMV officials issued genuine driver's licenses to the inspectors using the fictitious identifying information on the counterfeit breeder documents. In instances where DMV officials noted irregularities in the counterfeit documents, they still issued driver's licenses to the undercover agent and returned the counterfeit documents to him or her.[25] It remains clear that despite attempts by some states to make their driver's license systems more reliable, much more work remains to be done.

Additionally, there are still many people ready and willing to sell stolen or fake social security numbers and counterfeit birth certificates, which are then used to obtain false or counterfeit driver's licenses. In August 2003, it was reported that phony ID cards, including social security cards and driver's licenses, still could be purchased in Washington, DC, for anywhere between $20 and $135.[26] The low cost suggests these IDs are readily available.

Efforts to make identification more reliable in the short run are most likely to involve driver's licenses and state-issued identification cards—and thus motor vehicle agencies. There is no sense in ignoring that driver's licenses and state-issued identification cards are used for homeland protection. Therefore, it is important to identify the weaknesses in the current identification system. (In a later section, the subgroup will point to ways to improve driver's licenses and state-issued identification cards.) This discussion will focus on three areas of weakness in particular: the processes, personnel, and technologies involved.

[23] Prepared Testimony of McNulty.

[24] Breeder documents are basic documents that an individual needs to present to obtain other documents, such as driver's licenses or passports. Breeder documents include birth certificates, social security cards, and baptismal records.

[25] Prepared Testimony of Robert J. Cramer, managing director, Office of Special Investigations, GAO, before the Senate Committee on Finance on Counterfeit Identification and Identification Fraud Raise Security Concerns, 108th Cong., 1st Sess., 9 September 2003 (GAO-03-1147T).

[26] Sheridan, "Raids," A1.

WEAKNESSES OF DRIVER'S LICENSES AS RELIABLE MEANS OF IDENTIFICATION

Processes

1. Fraudulent breeder documents (for example, birth certificates, social security cards, baptismal records, etc.) often pass for the real thing. The wide availability of sophisticated graphics software programs and high-quality color printers, as well as how-to books, makes it easy to create counterfeit breeder documents.[27]

2. A state that issues a driver's license based on counterfeit breeder documents threatens the reliability of the entire system, as driver's licenses issued in one state are honored by all others. Wrongdoers seek out states with the weakest protections against false identification.

3. States have differing rules about the issuance of driver's licenses or state-issued identification cards to foreign visitors. Some states tie the expiration date of the foreign visitor's license to his or her visa expiration dates, while other states allow foreigners' driver's licenses to expire at the same intervals as citizens' licenses.

4. Each state issues its own license, and there are no standard minimum requirements. For instance, some states place the driver's photo on the left side of the card; others on the right. States also use a wide range of authentication features, including holograms, bar codes, multiple photos, and magnetic strips. With these differences, Transportation Security Administration (TSA) personnel, police, retail clerks, and bartenders in one state may not know what a license in any of the other 49 states looks like, nor how reliable a document it is.

Personnel

1. Some employees at motor vehicle agencies have been easily bribed into issuing false driver's licenses.[28]

2. It is often difficult for the personnel issuing driver's licenses to identify counterfeit or false breeder documents, as the GAO's recent investigation notes.[29]

3. State motor vehicle agency personnel do not always follow security procedures and are not always alert to the possibility of fraud, as the GAO's recent investigation notes.[30]

Technology

1. Many of the identifying features currently used in driver's licenses are not the most reliable; for instance, a person's eye color can be altered through the use of contact lenses, and weight often varies from what is listed on the card.

2. Most driver's licenses are easy to tamper with or forge. As with breeder documents, the wide availability of sophisticated graphics software programs and high-quality color printers, as well as how-to books, make it easy to create counterfeit IDs.[31]

These weaknesses in the current driver's license system need to be addressed to make our means of identification more reliable. Improvements will not only help our homeland security but will also have collateral gains, which will be discussed later.

Recommendations

This section of the report focuses on actions that should be taken by the government and might be taken by the private sector to make more reliable means of identification. In other words, we are seeking to strengthen the forms of ID that we currently have or may want to develop (in the case of private sector cards). Once again, it is important to stress that we are discussing ways to strengthen multiple means of identification—especially those means used for security purposes—and that we are not advocating a single identification system. As we have stated, as the security level of the purpose for which the card is used increases, so too should the reliability of the identification. Thus, it may often be necessary to rely on multiple means of identification. For instance, a driver's license should be more reliable than a college ID card, since a driver's license is used to gain entry into areas

[27] How-to books, such as John Q. Newman, *The ID Forger: Homemade Birth Certificates and Other Documents Explained* (Port Townsend, WA: Loompanics Unlimited, 1999), are available for purchase from mainstream retailers like Amazon.com.

[28] See, for example, Allan Legel, "Ex-Clerk Accused of DMV Fraud," *Washington Post*, 10 January 2003, B2; Christopher Quinn, "Bribery in Driver's Tests?" *Atlanta Journal Constitution*, 19 January 2002, 1A; and Ronald Smothers, "State Report to Outline Lapses in Security at DMV Offices," *New York Times*, 7 November 2002, A28.

[29] Prepared Testimony Cramer (GAO-03-1147T).

[30] Ibid.

[31] How-to books, such as Max Forge's *How to Make Driver's Licenses and Other ID on Your Home Computer* (Port Townsend, WA: Loompanics Unlimited, 1999), are available for purchase from mainstream retailers like Amazon.com.

such as airports and federal buildings, while a college ID is used to gain entry into a dining hall. These two purposeful means of identification serve widely differing functions—and each card is needed for its specific purpose. Our goal in creating more reliable means of identification is to fashion procedural speed bumps that make life unreliable for terrorists, but not to unduly burden law-abiding Americans in the process.[32]

Before we move on to examine ways in which the government and private sector can help make means of identification more reliable, a few general recommendations about how to proceed deserve to be mentioned.

We recommend that the DHS form a task force whose purpose it is to examine proposals (ours and others) to make means of identification used within its area of jurisdiction (transportation security, border security, immigration, and critical-infrastructure protection) more reliable and to implement—or foster, when the authority for implementation is outside its domain—the needed measures along the lines detailed in this report.

An interagency task force for reliable identification, led by the OMB, should also be formed. The interagency task force should be composed of representatives of the DHS, the NIST, the Department of Treasury, the Department of State, the CIA, the FBI, the NSA, the DoD, and the Department of Transportation, among other agencies, and should examine how those agencies' programs are affected by technical or process issues regarding current means of identification. This task force should collaborate with the DHS in identifying ways to make the means of identification used by all elements of the government, and for privately owned critical infrastructure, more reliable.

Governmental remedies

Because of the severity and urgency of the situation, short-run measures should be introduced first. Meanwhile, more reliable means of identification, which have longer lead times, will be studied to determine whether they may later be put into practice. This section will primarily focus on driver's licenses and state-issued identification cards, since they are, by far, the most widely used forms of identification issued by governmental agencies; however, some of these recommendations will help make other means of identification, such as passports and visas, more reliable as well.

There are numerous possible approaches from which to choose. We recommend a two-phase process toward making more reliable means of identification. The first phase will focus, albeit not exclusively, on how the federal government can assist in making state driver's licenses and state-issued identification cards more reliable as quickly as possible.

The second phase should initially focus on studying whether biometric and cryptographic technologies might be used to make driver's licenses and other forms of identification more reliable, and on determining which technology, if any, is appropriate and how the technology, verification, and enrollment processes may be implemented, given the primary purposes and uses of these means of identification. These studies should also address ways to protect privacy and other civil liberties while achieving more reliable means of identification. If an appropriate technology is identified, the technological wherewithal is available, enrollment processes have been carefully refined, and privacy concerns have been addressed, biometrics might be added to driver's licenses and other means of identification. Finally, some of us believe that a pure biometrics system may, in the long run, be preferred; others feel this idea is highly dubious and subject to error or fraud in the base technologies, the enrollment processes, or the people implementing the processes. Hence, at this stage, pure biometric technology should merely be studied.

Behind these specific recommendations is the assumption that all improvements to our means of identification require attention to three elements: the processes (the enrollment process; higher levels of validation; verification of the information on the card against information held on a network, at least when the ID is used for access to sensitive facilities; and audit trails); the personnel (improved training, selection, and oversight); and the technologies involved (biometrics, smart cards, scanning systems for network verification, cryptography, etc.). Each of these issues will be examined separately below.

RECOMMENDED GOVERNMENTAL REMEDIES FOR IMPROVEMENTS TO OUR MEANS OF IDENTIFICATION

Phase One

> We recommend that the federal government conduct research on affordable methods of improving identi-

[32] We thank Jerry Berman, president of the Center for Democracy and Technology, for making this point.

fication systems and making the entire identification mechanism more verifiable. We believe that the research should devote due attention to concerns about privacy and civil liberties. The government should encourage states to implement the studies' findings, to adopt interstate standards, and to put them into practice using grants.

These studies, which should address the three elements—processes, personnel, and technologies—will be of great assistance to states that are facing budget crunches and may not be able to afford to conduct such studies on their own.

We recommend that in each jurisdiction, the fines and penalties for individuals who possess, attempt to obtain, or sell counterfeit or false identification should be increased, as should the fines and penalties for individuals who knowingly supply such identification or knowingly allow people who are using it to enter controlled areas.

Processes

Paper breeder documents should be standardized, and birth- and death-certificate records should be digitized and searchable in all states.

The GAO's September 2003 report on the ability of undercover agents to obtain genuine driver's licenses using counterfeit documents highlights the problems with breeder documents.[33] Birth certificates are particularly problematic because they are issued by numerous jurisdictions and vary widely in format. This will make it easier for DMV officials—and other officials who issue means of identification, such as passports—to recognize counterfeit documents. We also recognize the argument that standardization of the documents may make breeder documents easier to fake in the long run. Digitizing breeder documents would allow DMV officials and others—such as Department of State officials who issue passports—to access birth- and death-certificate records electronically and reduce questions about the authenticity of paper documents.

The holders of this data should have privacy-protection measures (including audit trails) and enforcement policies in place, in order to control access to the data and to define specific purposes for which access to the data would be granted.

We believe the E-Vital program should be well funded once initial testing of the program shows its merits.

Only some states have made progress in making birth- and death-certificate records electronic. The good news is that the federal government has launched an initiative in this area; the bad news is that the initiative is still in its early stages. The federal initiative, called E-Vital, is establishing a common process through which birth- and death-record information can be analyzed, processed, collected, and verified.[34] This initiative will create a federal information repository of birth- and death-certificate records that will be electronically searchable. Because deaths will be certified online, this initiative will greatly decrease the amount of time it takes for a person's death to be officially reported to the SSA. However, there are both institutional and financial hurdles to overcome. Marsha Rydstrom, the SSA's project manager for E-Vital, said that the program faces problems in states that resist measures by the federal government to regulate management of state data. And there are funding issues, since the program's cost could range between $.5 and $5 million in each state, depending on the state's current capabilities.[35]

We believe that an elementary step in ensuring the validity of driver's licenses is to verify the social security number a person presents as his or her own when applying for a license.

State motor vehicle agencies are supposed to collect social security numbers from driver's license and state-issued identification-card applicants.[36] Motor vehicle agencies are allowed, but not required, to access the SSA's online database to verify the identity of the applicant. Prior to September 11, only 12 states used the Social Security Online Verification System (SSOLV) to verify the authenticity of social security numbers submitted to their DMV, according to the SSA.[37] States may choose to verify the authenticity of the driver's license applicant's social security number in two ways: first in real time, through an online check, and

[33] Prepared Testimony of Cramer (GAO-03-1147T).

[34] For more information, visit http://www.whitehouse.gov/omb/egov/gtog/evital.htm.

[35] Judi Hasson, "Electronic Death Records 'Vital' at SSA," *Federal Computer Week*, 1 April 2002.

[36] Under the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, states are to collect the social security number of driver's license applicants on their application. See U.S. Code, vol. 42, sec. 666a (1996). The GAO reported in 2002 that six states still were not collecting the social security numbers of driver's license applicants. See GAO, *Child Support Enforcement: Most States Collect Driver's SSNs and Use Them to Enforce Child Support*, Report to the Subcommittee on Human Resources, Committee on Ways and Means, House of Representatives, February 2002 (GAO-02-239).

[37] Office of the Inspector General, SSA, *Congressional Response Report: Terrorist Misuse of Social Security Numbers*, October 2001, 5 (A-08-02-32041).

second through batch checks, in which multiple checks are performed and reported at a later time, generally within 24 to 48 hours.[38] The number of states currently using the system stands at only 24 states and Washington, DC, according to the GAO.[39] That is, the majority of states still do not undertake this minimal verification step. Thirty-four state governments have entered into agreement with the SSA to use either the batch or online identification system, according to the American Association of Motor Vehicle Administrators, but problems with the performance and reliability of the SSOLV system have prevented any new states from being able to use the SSOLV since the summer of 2002.[40]

According to the GAO, one reason states do not use the SSOLV is cost.[41] Since states are strapped for funds and the verifications would require additional time, money, and work, we recommend that the Department of Transportation develop an approach to providing the needed funds, so states will be encouraged to undertake this verification step. Electronic birth- and death-certificate records will help immensely in solving this problem, though measures to verify social security numbers should not be stalled while E-Vital is still being tested.

We recommend that federal legislation tie the expiration date of the driver's license or state-issued identification card to the expiration date of the foreign-visitor's visa, as some states have already done.

States have varying rules for issuing driver's licenses to noncitizens. Some tie the expiration of the driver's licenses to the expiration of the visa, while others use the same expiration interval as that used for U.S. citizens.

We recommend that state driver's licenses and identification cards meet minimum uniform standards concerning the data content and the verifiability of the credential.

Driver's licenses vary greatly from state to state. Some states, such as Massachusetts, place multiple pictures on driver's licenses—larger and smaller versions of the same picture. In some states the picture appears on the left side of the license, while in other states it is located on the right side. Some states use a single bar code on their licenses; others use multiple bar codes; and some licenses do not have bar codes at all. The use of holograms, too, is inconsistent. These uniform standards can also address problems regarding the ease with which driver's licenses can be fraudulently altered or forged. For access to sensitive facilities (such as certain government buildings), verifying the information on the credential by comparing it to information on a network would increase the reliability of the credential. And while we recommend that all states adopt similar standards, there would still be room for variations among the licenses—for example, in the use of a state seal or motto on the license.

Personnel

We recommend that state motor vehicle agencies provide their employees with ongoing, detailed training about how to spot counterfeit or false documents. They should also provide law enforcement personnel with guidelines on checking the validity of driver's licenses.

As noted in a recent GAO report on the use of counterfeit documents to obtain licenses, many DMV officials do not recognize counterfeit documents when they are presented.[42] Periodically, a state could conduct spot checks to see whether officials spot the false documents and whether they follow protocol in those instances. For example, in states that require DMV officials to confiscate documents they believe are counterfeit or false, are officials complying with these guidelines? To better meet these responsibilities, state motor vehicle agencies should launch aggressive oversight, auditing, and anticorruption policies to help prevent fraud and to make it easier to detect fraud in the license-issuing process.

Phase Two

Technology

We need to develop studies to determine whether biometric and cryptographic technologies might be used to make driver's licenses and other forms of identification more reliable. Further research should examine available and new technology and make clear which, if any, is appropriate to improve our means of identification. We should examine the enrollment processes and their implementation, incorporating assumptions

[38] GAO, *Social Security Numbers: Improved SSN Verification and Exchange of States' Driver Records Would Enhance Identity Verification*, Report to Congressional Requesters, September 2003, 1 (GAO-03-920).

[39] Prepared Testimony of Cramer (GAO-03-1147T).

[40] GAO, *Social Security Numbers,* 12 (GAO-03-920).

[41] Ibid.

[42] Prepared Testimony of Cramer (GAO-03-1147T).

about the primary purposes and uses of the particular means of identification.

We believe multiple studies are needed, on two general levels: (1.) meta-analysis, overview, and codification of what is known (the results of various ongoing studies in the private sector and in the government); and (2.) the issuance of RFPs to invite additional studies that would cover well-known lacunae or those lacunae the analysis of the first level—the summaries of the state of the art—would reveal.

Some examples of current technologies are smart cards, two-dimensional bar codes, scanners for network verification, and magnetic strips. Biometric data already exists on driver's licenses, and for years biometric data has been used to link an individual to an identification card. For instance, driver's licenses include a photo and other identifying information, such as height, weight, and eye color. Unfortunately, these biometrics are not the most reliable: Individuals can gain or lose weight, or lie about it, and eye color can easily be changed using contact lenses. The addition of new forms of biometric data on driver's licenses—data that is difficult to change and is specific to the individual—might increase our ability to identify individuals more reliably and accurately, especially when a higher level of security is needed.

Any analysis should address ways to protect privacy and other civil liberties while achieving more reliable means of identification. Recommendations for the collection, storage, and use of biometric data should be addressed, as should the possible unintended consequences of collecting it.

Private sector remedies

We believe the government should explore private sector alternatives to making our means of identification more reliable. We urge DHS officials to convene a panel of representatives from corporations to determine incentives to encourage the private sector to use various purposeful cards (credit cards, medical cards, etc.). These cards could be purchased voluntarily by consumers, could be more reliable and verifiable, and could use photos or biometrics along with other identifying information.

Among the options to be examined is whether various new cards could be used to provide secondary verification of identity. The private sector has shown repeatedly that it can and does create successful means of identification. For instance, many corporations are devising their own purposive means of identification, some of which are low-tech and others high-tech. And some companies will not even allow an employee to enter the premises if he or she has forgotten the company-issued ID, even if the employee can present a driver's license to security officials.[43]

Private sector initiatives have been launched to develop more reliable means of identification, with ATM cards as one example of this. Below we explore issues surrounding the private sector producing a more reliable means of identification, whether companies could make the identification more widely available and acceptable while providing incentives to people for its use. We also examine whether this method might ease some concerns about identification, by proposing new means of identification that is less intrusive, not more, and helping to convince the public that improving identification will increase security.

Although it appears that the private sector is interested in having more reliable means of identification, the question remains: Would "high-security" cards catch on? These would be cards that could be purchased for approximately $65 to $100, from various companies, and with which one could cash checks and pass through building and airport security, among other things. This does presuppose that the government would partner with the private sector, accepting means of identification developed and used by the private sector. Would these cards ease the problems at hand? What incentives might be needed to encourage the private sector to develop high-security cards?

We are suggesting that if private sector cards, obtained on a voluntary basis, could reliably identify individuals, then routine identification (not to be confused with security checks before entering highly secured areas) could become more reliable, with little or no cost to the government. Moreover, the stigma now attached to some identification methods could be reduced, due to the voluntary nature of the purchase. In addition, private sector cards could also be used for nonsecurity purposes. If the private sector card were a smart card and were embedded with a computer chip and encryption technology, ATM and credit card functions could be added to the card as well.

[43] We thank Eric Benhamou, chairman of the board of directors with 3Com Corporation, for this point.

Accountability and privacy protections

We believe that if accountability if found deficient (or excessive), the remedy is to adjust accountability but not to deny a measure altogether.

New measures that are introduced to enhance security and, more generally, to assist in law enforcement are often examined in terms of whether they are of merit as separate and distinct solutions. However, judging the legitimacy, or value, of a public policy measure entails more than establishing whether it significantly enhances public safety, is minimally intrusive, undermines further our already endangered civil rights, or makes it more difficult to deal with other public needs. The legitimacy and value of a policy must also be based on a judgment of those who employ new powers: Are they sufficiently accountable to the various overseers—ultimately, the citizenry? Some powers are inappropriate no matter what oversight is provided. However, for the issue at hand, the main question is whether there is sufficient accountability.

Concerns about privacy should be addressed in all matters regarding more reliable means of identification. We believe that studies of ways to make means of identification more reliable should also include the quest for ways to protect privacy and civil liberties.

As we mentioned earlier, for personal data, such as digitized birth- and death-certificate records, we believe that the owner of the data should have privacy-protection and enforcement measures in place that address access issues. For instance, audit trails should be established that could detect unauthorized use of data and help deter it.

We also recommend that the DHS set up a public-private body to review more reliable means of identification measures to be used for homeland security purposes as they emerge, and also to examine the measures' effectiveness and privacy implications. This body should operate under the criteria specified in the Federal Advisory Committee Act.

Collateral gains

If more reliable means of identification were available for national security purposes, then a great number of other safety and nonsafety issues could be alleviated. Collateral gains would be possible; we examine some of them in this section.

Protecting the innocent

A major example of the miscarriage of justice is the well-established and widely known fact that people are misidentified and jailed for crimes they did not commit. With more reliable means of identification, the incidents should decrease in which innocent people are barred from flying, driving, entering the U.S., and obtaining security-sensitive jobs.

Identity theft and credit card fraud

The Federal Trade Commission (FTC) reported that it received more than 160,000 complaints of identity theft in 2002[44]; and this year alone, the FTC anticipates receiving some 210,000 complaints.[45] These reported complaints are low-end estimates of the prevalence of identity theft. A September 2003 FTC survey estimated that within the past year, more than three million Americans discovered that their personal information had been misused; it also found that the total annual cost to identity-theft victims is about $5 billion.[46] If means of identification were more reliable, then such fraud could be more difficult to commit and easier to detect.

Voter fraud

Identification difficulties can lead to problems with voter fraud. In many states, deceased voters remain on the voting rolls and individuals with false or counterfeit identification can often vote in person or often request absentee ballots. Picture identification is not consistently required. If means of identification were more reliable, then voter fraud could be easier to detect.

Fugitives

While the exact number of felons at large is not available, some estimates have been made: In 2002, the FBI said it

[44] FTC, Identity Theft Data Clearinghouse, "Information on Identity Theft for Consumers and Victims from January 2002 Through December 2002." Available at http://www.consumer.gov/idtheft/reports/CY2002ReportFinal.pdf. Accessed 11 July 2003.

[45] FTC, *Overview of the Identity Theft Program: October 1998–September 2003*. Available at http://www.ftc.gov/os/2003/09/timelinereport.pdf. Accessed 10 September 2003.

[46] FTC, *Identity Theft Survey Report*, prepared by Synovate, September 2003. Available at http://www.ftc.gov/os/2003/09/synovatereport.pdf. Accessed 10 September 2003.

was looking for about 12,000 fugitives at any one time.[47] The lack of reliable means of identification makes it difficult for law enforcement officials to catch fugitives who have skipped court appearances or those with warrants out for their arrest. If a police officer pulls over a speeding driver in Oregon and checks the driver's license, the officer is unable to determine whether there is a warrant out for that person's arrest in another state; there is also no way of telling if the driver is using a false or counterfeit ID. If means of identification were more reliable, law enforcement would better be able to accurately identify the driver. Likewise, more reliable means of identification would help when individuals who are prohibited from driving—due, for example, to several DUI convictions—get behind the wheel of a car.

Employment

Convicted sexual predators often depart the jurisdiction of their offenses only to apply later for jobs at child-care centers or schools elsewhere in the nation. While their names may be compiled in a national network, such a database is useless if the predator has counterfeit or false identification. In much the same way, abusive health-care workers—particularly those caring for the elderly—will often apply for jobs caring for the vulnerable, even after having been previously caught and terminated. Efforts to warn other health-care providers will be more successful with more reliable means of identification.

Other programs

Lack of reliable identification can create great expense for other government programs, such as those for student loans, affordable housing, and food assistance, and can lead to a loss of revenue in terms of individual income tax payments. By using electronic benefits transfers, the government has cut down on fraud in some of these systems; reliable identification can help even more, especially during the enrollment process of these programs.

[47] FBI, "General Frequently Asked Questions." Available at http://www.fbi.gov/aboutus/faqs/faqsone.htm. Accessed 12 September 2003.

Appendix B

A Primer on Homeland Security Players and Information

by Mary DeRosa and James Lewis

Introduction

To make specific recommendations about a network for sharing homeland security information, it is necessary to understand what the information is and the players who collect and use it. This memorandum attempts to provide some basic, practical information about who collects homeland security information, how they collect that information, and who uses it. In Section 1, we will discuss information collection and introduce some key collectors. In Section 2, we will provide examples of some information users and discuss their information needs. When recommending a network for information-sharing, we also have to recognize and address the dangers of disclosure of certain types of information. Section 3 of this primer will therefore explain some of the policies and values behind protecting information from disclosure.

Section 1: Information collectors

In this primer we discuss information in four categories: (1.) information collected for federal law enforcement; (2.) intelligence; (3.) information collected by federal agencies in the course of their duties (other than law enforcement and intelligence); and (4.) information from state and local police and government agencies. We will not discuss information collected by the private sector, which also can be crucial to developing terrorism warnings.

Law enforcement information

Law enforcement information is information collected to investigate, solve, and prosecute crimes. Law enforcement is primarily reactive. That is, although sometimes law enforcement operations prevent crimes, usually they solve crimes after they occur. Federal law enforcement officers investigate crimes and work with the Department of Justice (DOJ), including U.S. Attorneys' offices, to indict and prosecute criminals. In the course of investigations and prosecutions of suspected terrorists, law enforcement officials gather a great deal of information about terrorists. For example, from investigations of the 1993 World Trade Center bombing, the 1998 embassy bombings, the attack on the *USS Cole*, and other terrorism investigations over the past decade, the Federal Bureau of Investigation (FBI) collected significant information about Al Qaeda's structure, methods, and membership. Such information is usually recorded in evidence reports, but it can also be in court papers such as indictments.

THE MOST SIGNIFICANT INFORMATION-COLLECTION METHODS USED BY FEDERAL LAW ENFORCEMENT AGENCIES

1. **Forensic/crime scene and other physical evidence**
   Fourth Amendment protections apply to searches and seizures of physical evidence in private places.

2. **Interviews and interrogation**
   Interviews can be of witnesses or suspects. There are well-known constitutional constraints on the questioning of suspects in custody.

3. **Criminal and other public sector databases**
   Agents will refer to databases, such as the National Crime Information Center (NCIC), to check on criminal background and other information about people of interest in an investigation.

4. **Private sector data**
   Sometimes agents will purchase, request, or demand by some legal process (for example, subpoena or warrant) data from the private sector on individuals. This could include credit, financial, travel, communications, or other similar data.

5. **Physical surveillance**
   Physical surveillance in public places can raise First Amendment issues if it chills the exercise of protected speech.

6. **Human sources (HUMINT)**
   These can be paid or volunteer sources who develop relationships with specific agents. There are detailed procedures for their recruitment and use.

7. **Electronic surveillance**
   Wiretaps and most other electronic surveillance for federal law enforcement are conducted pursuant to Title III of the Omnibus Crime Control and Safe

Streets Act of 1968, which requires a judge to find probable cause that a specific crime has been, is being, or will be committed and that the wiretap will obtain communications about that crime.

8. **Undercover (covert) operations**
   These extremely sensitive operations involve law enforcement personnel infiltrating criminal groups. They are time-intensive and often very expensive.

Many of these tools and techniques are the same as those used to collect intelligence, which will be discussed below. The differences are in the legal authorities for, and restrictions on, gathering the information; the purpose for collection; and the ultimate use of the information. Law enforcement agents must always be attentive to constitutional protections of the people they investigate. If evidence is collected in a manner that violates a constitutional protection, it can be excluded from use at trial.

To those collecting it, law enforcement information is evidence, which leads to some problems with sharing the information. First, because those collecting the information are focused on solving a particular crime, they sometimes will ignore—or at least fail to record—information that could be relevant to preventing future terrorist attacks but does not relate to that particular crime. One example of this is the case of convicted terrorist Abdul Hakim Murad. Prior to September 11, the FBI learned, as part of a criminal investigation, that Murad had been involved in plots to blow up 12 U.S.–owned airliners over the Pacific Ocean and to crash an aircraft into the Central Intelligence Agency (CIA) headquarters. But information about those plots was not relevant to the crimes with which Murad was charged. Information about those plots did not show up in Murad's indictment or in any other form that would have allowed analysts to assess it in light of other information about terrorist plots. The information, essentially, was lost.

An even greater problem with sharing of law enforcement information is the strong incentive for law enforcement personnel to keep investigations and evidence secret because of a concern about protecting eventual prosecution. The value of protecting the secrecy of ongoing investigations will be discussed in greater depth in Section 3. It is clear, though, that this legitimate concern affects the culture of law enforcement information-gathering generally, and that it leads to hoarding of information that could be shared without harming eventual prosecution.

The Federal Bureau of Investigation (FBI)

The FBI has the broadest law enforcement jurisdiction of any federal law enforcement agency. It has the authority to investigate any federal crime that is not exclusively within the jurisdiction of another agency and is the federal law enforcement agency responsible for investigating terrorist crimes. The FBI also has an intelligence mission, discussed below, which, in the area of counterterrorism, has increased significantly since September 11.

The FBI has 56 field offices in major cities across the country and smaller resident agencies in some smaller locations. Each field office operates with a great deal of autonomy. Agents in field offices initiate and run investigations and operations on their own, although they need to seek authorization for certain activities—such as undercover operations—from headquarters. The primary documentation for field-office criminal investigations is the FD-302 report (an official report of evidence collection—such as a witness interview or report of surveillance—that can be used in court). Traditionally, FD-302 reports are closely held and not shared with other field offices. Field offices also record information from discussions and investigations in less formal memoranda. Since September 11, memoranda containing information that could be related to terrorism are usually forwarded to a local Joint Terrorism Task Force (JTTF) (a team of state and local law enforcement officers, FBI agents, and other federal agents whose purpose is to pool expertise and share information) and the FBI headquarters.

FBI field-office counterterrorism personnel work with JTTFs throughout the country. There are currently 84 JTTFs (an increase from 35 in 2001). JTTFs are headed by a supervisory agent from the local FBI field office, and are, more often than not, located with the FBI field office. Historically, the information-sharing has often been in one direction, with the FBI being reluctant to inform state and local agencies of operations or investigations for fear of interference that could harm those investigations. As a result, JTTF representatives from agencies other than the FBI agree not to share information they receive from the JTTF with their agency unless they receive approval from the JTTF head.

At FBI headquarters in Washington, DC, oversight and direction for counterterrorism criminal investigations come from the Counterterrorism Division. This division is responsible for all counterterrorism matters, whether criminal or intelligence. Supervisory special agents in field offices determine what information to share with headquarters and report it to the Counterterrorism Division.

The Financial Crimes Enforcement Network (FinCEN)

The Financial Crimes Enforcement Network (FinCEN) is a Treasury Department agency. It was established in 1990 to administer the Bank Secrecy Act (BSA), support law enforcement agencies, and analyze information from banks and other sources. Banks and other financial institutions provide FinCEN with information on financial transactions. The BSA's record-keeping and reporting requirements create a financial trail for investigators to track terrorist activities and assets, and FinCEN's data-collection authorities have been expanded by a number of laws aimed at money-laundering, the most recent being the USA PATRIOT Act. FinCEN has approximately 200 employees, most of whom are analysts. FinCEN also has 30 to 40 long-term detailees from different law enforcement and regulatory agencies.

FinCEN emphasizes the use of network and information-processing technologies. The agency uses data extraction, data mining, and analytical software tools on the data it receives under the BSA. It uses data from the Suspicious Activity Report system (also known as SARs) in combination with other intelligence, law enforcement, or commercial information to identify trends and patterns in money-laundering and BSA data.

FinCEN defines itself as a network of law enforcement, financial, and regulatory agencies on the international, federal, state, and local level. It links law enforcement agencies and financial institutions to allow them to share information on suspicious financial transactions. The Hawala system of informal money transfers that is widely used in Pakistan and the Persian Gulf poses a challenge for FinCEN, which relies primarily on information received from banks and other financial institutions.

Intelligence

The purpose of intelligence is to provide warning, help assess threats and vulnerabilities, identify policy opportunities, and assist policymakers in national security decision-making. Unlike information collected for law enforcement, the purpose of intelligence collection is to prevent harm. Because of the potentially devastating effects of a terrorist attack, counterterrorism is seen increasingly as more of an intelligence challenge than a law enforcement challenge. The tools and techniques for collecting intelligence are similar to those used for law enforcement, but the authorities are different. Intelligence collected abroad on foreign persons does not raise Fourth Amendment search-and-seizure issues. Intelligence collected on U.S. persons or within the U.S. however, does raise some of these constitutional issues. But when the purpose of the collection is for national security, courts have allowed greater flexibility for intelligence collection than for law enforcement, particularly when the threat can be shown to be a foreign power.

The head of the U.S. intelligence community is the Director of Central Intelligence (DCI). The DCI is responsible for coordination and policy direction for the entire intelligence community, which includes entities within the Department of Defense (DoD) and several other Executive Branch departments. The DCI has direct authority for the programs, staff, and budget of the CIA. As mentioned above, intelligence collection uses most of the same methods as law enforcement.

THE MOST SIGNIFICANT METHODS OF INTELLIGENCE COLLECTION

1. **Human sources (HUMINT)**
   Many post–September 11 analyses have noted the weak collection capabilities for human intelligence on non-traditional threats such as terrorism and weapons of mass destruction.

2. **Imagery**
   This is primarily satellite imagery, but also includes imagery from manned and unmanned aircraft and other sources.

3. **Electronic surveillance**
   This includes intercepts of telephone and other electronic communications. The authority for electronic surveillance conducted in the U.S. is the Foreign Intelligence Surveillance Act (FISA). If surveillance involves a U.S. person, the FBI conducts it. The FISA requires the government to obtain a secret court order from a special court, the Foreign Intelligence Surveillance Court (FISC). The government must show probable cause that the target is, or is an agent of, a foreign power. No such authorities are required for surveillance originating or occurring outside the U.S.

4. **Interviews and interrogation**
   Information obtained in this manner normally is disseminated as HUMINT reports. Here the person conducting the interview is key: If he is unaware of important pieces of missing data in the terrorism picture, he may fail to ask a relevant question, or may fail to record a piece of valuable information.

5. **Seized materials**
   Items seized or turned over to intelligence agencies, such as computers, records, equipment, or maps, must be "exploited," or analyzed, by technically competent persons who are also aware of the analytic picture. This effort takes a long time to complete; but shortcuts can result in conclusions that are unreliable.

6. **Covert action**
   These are activities that are not primarily for intelligence collection, although they often produce intelligence. They are extremely sensitive operations directed by the President and designed to influence political, economic, or military conditions abroad, where it is intended that the U.S. role will not be acknowledged publicly.

The information collected from these sources is called "raw intelligence." Raw intelligence must be combined with other intelligence and analyzed to get a sense of its credibility, reliability, and significance. The results of this analytical process are called "finished intelligence." Our intelligence structure gives the intelligence collectors ownership of the information they collect, and collectors protect raw intelligence jealously. Indeed, the national security classification system allows the originator of a piece of intelligence to place the designation "Originator Controlled," or "ORCON," on a piece of intelligence. This means that the intelligence cannot be distributed further without the originator's approval. This insistence on control is due in part to the fear that without such control the information will be leaked or inadvertently released and a critical source or method will be compromised. This concern is discussed in greater length in Section 3. At least as important, controlling information is often seen as a way to preserve bureaucratic power.

When raw intelligence is controlled in this way, the real loser is intelligence analysis. Each intelligence organization has, to a greater or lesser degree, its own analysts. These agencies, in the past, have preferred to have only their own analysts see their raw intelligence. As a result, there were many analysts with parts of the story, but little real all-source analysis. Since September 11, the intelligence community has recognized this problem, and there have been some improvements. The Terrorist Threat Integration Center (TTIC), discussed below, is designed, in part, to address this problem.

Finished analytical products are distributed more freely than raw intelligence. Many reports are circulated routinely among groups of cleared policymakers and other officials. Other more sensitive products—such as anything about a covert action, or intelligence that would reveal a particularly sensitive source—are never distributed in electronic form and are kept within a tight circle of cleared officials.

Sometimes intelligence from a sensitive source is "sanitized." That is, less-sensitive material is extracted so that a broader audience can view the remainder. The sanitized version of intelligence can sometimes have a lower classification (for example, "Secret" rather than "Top Secret"), or it can even be unclassified. Sometimes this is done on a paper report with a tear line. Below the tear line is sensitive information that would reveal the source; above the line is data extracted from the report that is less sensitive. These paper reports are actually torn apart, and the top portion is distributed more broadly. Often, however, policymakers and other officials find the sanitized data on these and other reports to have limited usefulness because it lacks context or key information.

The Central Intelligence Agency (CIA)

The CIA is responsible for collecting foreign intelligence, primarily outside of the U.S., through human sources and other means; for analyzing and disseminating that intelligence; for conducting and coordinating counterintelligence activities outside of the U.S.; and for conducting covert actions approved by the President outside of the U.S. CIA offices relevant to homeland security are the Directorate of Operations (DO), the Directorate of Intelligence (DI), and the DCI Counterterrorist Center (CTC).

The DO is the service responsible for gathering human-source intelligence around the world. It does this primarily by recruiting HUMINT sources and by collaborating with host-country intelligence services and police services. The DO is also the CIA directorate responsible for overseas covert action. DO sources and operations are among the most sensitive information in the intelligence community, and the DO is notoriously reluctant to share information—even within the CIA. Information comes directly to the DO headquarters from field offices, and DO personnel prepare a report about that information. Raw products that would identify a human source never leave the DO, and typically only the most senior CIA analysts see the DO report. To the extent information about human sources and about covert actions is disseminated, it is done only on paper, not electronically.

The DI is the CIA analysis office. Analysts from the DI gather information from the CIA and other sources and conduct strategic analysis. The mission of the office is to provide timely and objective assessments to senior U.S.

policymakers in the form of finished intelligence products, including written reports and oral briefings.

DCI William Casey created the CTC in the late 1980s after a series of high-profile attacks by international terrorists. The CTC reports to the DCI and, technically, is not part of the CIA bureaucracy, although it is housed at, and is supported administratively by, the CIA. The CTC's mission is to assist the DCI in coordinating the counterterrorism efforts of the intelligence community by coordinating and conducting counterterrorist operations and exploiting all-source intelligence in order to produce in-depth analyses of terrorist groups, methods, and plans. Since 1996, the CTC and the FBI's counterterrorism directorate have been exchanging senior-level officers, although before September 11, this collaboration did not always result in successful information-sharing between the two entities. One criticism of the CTC has been that it has operated mostly with the DO and has emphasized operations over collection and analysis.

The National Security Agency (NSA)

The NSA collects signals and communications intelligence on foreign targets of concern to the U.S. The NSA collects an immense amount of traffic, and one of its key daily tasks is to reduce millions of intercepts down to a few thousand for analysts to review. Computers do this filtering using specialized software. Linguists and analysts with area or subject expertise then review the much smaller set of filtered intercepts to determine their importance. At the end of this daily process, a small number of intercepts is found to be useful.

The NSA prepares processed reports, some of which are available in the routine traffic circulated among agencies. Other, more sensitive reports are closely held and handled in special dissemination channels. On rare occasions, the NSA will also provide raw traffic (for example, translated text of actual intercepts) to senior policymakers. Intelligence analysts at other agencies rely on input from the NSA in developing their own analyses, and the NSA can be tasked by agencies to collect intelligence on specific problems or to search databases. The NSA has finite collection and analytical resources, so high-priority assignments can bump long-term or less-important collection projects. Signals and communications intercepts provide very valuable intelligence, but sophisticated targets like Al Qaeda use a variety of techniques to evade interception. NSA material is usually highly classified, not only because of the sensitivity of the material, but also because of the sensitivity of the collection techniques. Currently, signals and communications intelligence is one of the most important sources of information that the Department of Homeland Security (DHS) uses to issue alerts, but the actual intelligence upon which the alert is based is not shared with local authorities.

The Federal Bureau of Investigation (FBI)

The FBI is the agency responsible for collecting intelligence on terrorists in the U.S.; it is the only U.S. domestic-intelligence agency. U.S. policy and regulation restrict foreign-intelligence agencies from collecting intelligence on U.S. persons. The FBI collects intelligence related to foreign threats, such as international terrorism, pursuant to FISA and the "Attorney General's Guidelines for FBI National Security Investigations and Foreign Intelligence Collection." (These guidelines are largely classified.)

As with law enforcement collection, the FBI organizes its intelligence collection by cases or investigations called "foreign counterintelligence," or "FCI," investigations. For the most part, field agents do not specialize in intelligence collection. An agent in a field office can, at any time, be conducting a criminal investigation of terrorism or an FCI terrorism investigation. Important intelligence gathered in field offices is shared with FBI headquarters. Headquarters officials make decisions about sharing intelligence with other intelligence agencies and policymakers.

A number of institutional issues has made the FBI historically ineffective as an intelligence agency. Most important, the FBI is fundamentally a law enforcement agency. Its culture is that of a law enforcement agency, and the system rewards success in law enforcement such as arrests, prosecutions, and convictions. The disciplines of law enforcement and intelligence differ in critical ways, and FBI special agents primarily are taught the law enforcement view of how and why information is collected. Senator Shelby, in his "Additional Views" to the Joint Congressional Investigation of September 11, referred to the "tyranny of the casefile." He meant by this that FBI agents are trained to think in terms of a case, which causes them to focus on discrete individuals or organizations. Information about an individual believed to be in Hezbollah, for example, could be viewed as part of the Hezbollah case and would not necessarily be considered as part of an investigation of Al Qaeda. Thus, agents or analysts become experts on one group, but correlations, trends, and patterns involving both can be lost.

FBI agents are also unfamiliar with being a tool for policymakers or other national security decision-makers. They are simply not accustomed to—and in fact their culture discourages—a focus on a customer other than the prosecutor. Finally, the FBI has not traditionally valued, rewarded, or even understood analysis, which is critical to intelligence.

Since September 11, the FBI has recognized many of these deficiencies and has made significant changes to address them. For example, it has greatly expanded its office of analysis and has enhanced analyst training. There is now an analysis branch in the Counterterrorism Division at headquarters, which focuses on strategic assessments and reports to policymakers. The FBI has also established the position of reports officer. The reports officer's job is to extract relevant information from FBI criminal and FCI investigations, turn it into Intelligence Information Reports (IIRs), and disseminate it as widely as possible. The FBI is hiring reports officers who will be assigned to field offices to support local law enforcement and intelligence community needs. Perhaps the most significant new development is the FBI's elevation of intelligence in its management structure. The FBI recently created and filled the new position of Executive Assistant Director for Intelligence at FBI headquarters. It has also appointed Assistant Special Agents in Charge (ASACs) of intelligence in each field office, and is creating separate intelligence units in all field offices.

Despite these steps, some policymakers and experts believe that the FBI's mix of law enforcement and intelligence functions is inherently ineffective. They advocate creating a separate domestic intelligence agency—similar to the U.K.'s Security Service (MI-5)—that would be responsible for collecting and analyzing domestic intelligence.

The Terrorist Threat Integration Center (TTIC)

The newly created TTIC is intended to be a center for fusion and analysis of terrorist-threat intelligence information from all sources, domestic and foreign. President Bush announced the plan to create the TTIC during his 2003 State of the Union address, and its doors opened in May 2003. The TTIC's director is a CIA official who reports to the DCI, but it is a joint venture that includes personnel from the CIA, the FBI, the DHS, and several other entities of the intelligence community. It is currently housed in the CIA complex. According to the White House announcement and testimony by administration officials, TTIC personnel have unfettered access to raw and finished intelligence about terrorist threats. The TTIC does not collect intelligence.

The TTIC's mission is to integrate and disseminate terrorist threat–related information and analysis. Its analytical staff—which consists primarily of junior analysts—includes about 100 members (as of July 2003), but that number is expected to increase significantly over the next year. The TTIC's analytical focus is on preparing two daily reports: the President's Terrorism Threat Report (PTTR) and the Terrorism Threat Matrix (TTM). The PTTR is a highly sensitive analysis for the President of the daily threat information. The TTM is a compilation, without analysis, of the terrorist-threat information received in the previous 24 hours; it is distributed to senior officials in all federal government intelligence agencies with a homeland security mission.

To assist TTIC in its information-dissemination responsibilities, the DCI, the Attorney General, and Secretary of the DHS signed, in March 2003, the Memorandum of Understanding on Homeland Security Information Sharing, which commits all agencies participating in TTIC to take steps—such as minimizing use of originator controls and producing sanitized versions of intelligence—to increase intelligence-sharing. In practice, it is not yet clear that this agreement has had a significant effect. For example, TTIC analysts may not disseminate information they receive without originator permission. Moreover, most analysts must keep an array of computer terminals under their desks in order to access information from different U.S. government sources and cannot perform one search against multiple-agency databases simultaneously.

One significant information-sharing advance the TTIC has implemented is the TTIC Online website. This website hosts TTIC analysis and links to other counterterrorism reports. It reaches analysts with the appropriate clearances at all major departments and agencies with a homeland security mission, including JTTFs around the country. Currently, TTIC Online contains information at the Top Secret/SCI level. The website is, therefore, available only to people with the highest clearances and in the most secure environments. However, the TTIC plans to replicate TTIC Online on less-sensitive networks, to provide less-sensitive information and analysis to a broader community of analysts and other consumers.

Because it is in its infancy, there are still many questions about the TTIC's role and functions. It is not clear, for example, how great a role FBI personnel will actually play in the TTIC, although the intention is that it will be significant. It also remains to be seen how TTIC personnel will interact with intelligence collectors to set collection priorities. Nor is it clear how much information the TTIC will receive from the DHS or other non–intelligence agencies that collect information, from state and local governments, or from the private sector.

Another significant question is how effectively the TTIC will disseminate intelligence to all players responsible for preventing or responding to terrorist attacks. It is sure to provide information to the DHS and the FBI. Less certain is whether the TTIC will have any direct relationship with state, local, or private sector entities.

The Information Analysis and Infrastructure Protection Directorate (IA&IP) of the Department of Homeland Security (DHS)

The Homeland Security Act established the IA&IP in the DHS, headed by an Under Secretary, with an Assistant Secretary for information analysis. The legislation envisions an intelligence entity that would receive and analyze information from within the DHS and from law enforcement, intelligence, state, local, and private sector entities. It would analyze that information and use it to do the following:

1. Assess the nature and scope of threats and potential vulnerabilities.
2. Perform risk assessments.
3. Identify priorities for protection and support measures.
4. Develop a national plan for securing key resources and critical infrastructure and recommend measures to protect them.
5. Provide warnings of terrorist attacks.
6. Disseminate information within the DHS and to other federal, state, local, and private sector entities responsible for homeland security to assist in prevention, and response to, terrorism.

The statute is explicit that, except as otherwise directed by the President, the DHS is to have access from any federal agency to all information and intelligence—including raw intelligence—about terrorist threats and vulnerabilities of the U.S. to terrorism. The directorate does not have authority to collect intelligence.

When the legislation was passed, many assumed this office would be responsible for all-source fusion and analysis of intelligence for homeland security. With the creation of the TTIC, it is unclear how much the DHS entity will conduct its own analysis and how much it will rely on the TTIC. The directorate will almost certainly duplicate the TTIC's functions to some degree.

Other federal agencies

A significant amount of the information collected by the federal government that is relevant to homeland security comes from agencies whose primary function is not intelligence collection or law enforcement. Most of these agencies are now in the DHS, but some very significant ones are in other departments. These agencies collect the information they need to carry out their primary function (immigration or border control, tracking infectious diseases, collecting taxes, issuing visas, etc.). The information collected in the process is often records of applications or transactions (visa or immigration information, shipping manifests, etc.). It can also be reports of diseases in people or agriculture, or information necessary for government programs (tax or social security records). Most of this information is not classified. However, accessing some of it, such as IRS records, raises significant privacy concerns.

The Bureau of Immigration and Customs Enforcement (BICE) of the Department of Homeland Security (DHS)

The BICE at the DHS is the enforcement arm of the Border and Transportation Security Directorate (BTS) (the operational directorate within the DHS responsible for securing the nation's borders and transportation infrastructure). The BICE combines the enforcement functions of several former border and security agencies, including the Immigration and Naturalization Service (INS) and the United States Customs Service, and focuses on enforcement of immigration and customs laws.

In the course of its enforcement work, the BICE collects significant, valuable information about terrorists and their organizations, drug and contraband smuggling, human trafficking, illicit trading of weapons of mass destruction, money-laundering and financial crimes, threats to government facilities, and other matters relevant to homeland security. The BICE has its own office of intelligence, which collects and analyzes this information and shares it with the DHS's IA&IP.

The BICE also has a variety of databases with information on immigrants and visitors to the U.S., which can assist law enforcement and intelligence agencies in fighting terrorism. These include the Student and Exchange Visitor Information System (SEVIS), which manages and maintains data about foreign students and exchange visitors; the National Security Entry-Exit Registration System (NSEERS), which contains detailed registration information about foreign visitors of elevated national security risk—primarily nationals of certain high-risk countries; and the United States Visitor and Immigration Status Indication Technology (US VISIT) system, a new system that will manage data, including biometric identifiers and entry, exit, and status information, on all visitors to the U.S.

The BICE's Law Enforcement Support Center (LESC) is a national enforcement-operations center located in Vermont. Its purpose is to share information with federal, state, and local law enforcement agencies about the immigration status of aliens suspected of, arrested for, or convicted of criminal activity. The LESC gathers information from eight DHS databases, including SEVIS, NSEERS, US VISIT, and other former INS, Customs Service, or Federal Protective Service databases. It also has access to several national and state criminal-information databases.

The Department of Health and Human Services, Centers for Disease Control and Prevention (CDC)

The CDC is the lead federal agency for preventing disease. Its primary function is to provide useful information to enhance health decisions. The CDC carries out its duties primarily by interacting with state and local health providers. The CDC has more than 100 health-surveillance programs nationwide, most of which track specific diseases or trends in clusters of diseases, such as food-borne illnesses and hospital infections. It is developing a larger network-based system to monitor and communicate information about outbreaks of disease, including biological attacks. The CDC's National Electronic Disease Surveillance System (NEDSS) is an initiative to create information-system and data standards for integrated and interoperable surveillance systems at federal, state, and local levels. At this time, many state and local health agencies use different data formats or even depend on paper and fax machines, complicating any effort to develop a national health-monitoring system. As the NEDSS progresses, its purpose will be to improve the ability to identify and track emerging infectious diseases and potential bioterrorism attacks. The NEDSS will put local and state public-health, clinical, and laboratory data into a larger national monitoring network. The CDC's work in this area predated September 11, but has increased in intensity recently.

State and local government agencies

State and local government entities play a critical role in collecting homeland security information. Terrorists live in, and plan attacks throughout, the country. States and localities often have information that is a piece of a puzzle about terrorist activities. One place these clues can be found is in state databases that contain DMV or other license records, records of arrests, or court records.

More important, state and local personnel cover more ground than the federal government could hope to. The FBI has only 11,400 agents nationally. There are many hundreds of thousands of local police and sheriff's office personnel around the country. If terrorists are casing potential targets or attempting to acquire tools or training to commit terrorist acts, state and local police officers are likely to hear about it first. Also, in the course of their regular law enforcement duties, these officers often uncover activity that could be related to terrorist planning. Police officers and local security officials at ports, airports, rail stations, and on highways are sometimes in the best position to detect the movements of suspicious people and dangerous cargo. The problem is that there is little regular, coordinated sharing of this local information with federal and other officials who are in a position to fit it into a larger context.

A local police report about strangers lurking around a train containing hazardous material, for example, is likely to go no farther than the local precinct. If the report is contained there, the mosaic of a terrorist plan to use that train or those materials for an attack will be harder to recognize. Some states and regions have developed law enforcement or terrorism-related databases with information about criminal or suspicious activity that can be accessed by law enforcement officials in terrorism investigations.

State and local public health and agricultural officials are most likely be the first to see signs of a biological attack. Public health agencies, coroners, medical examiners, pharmacists, and health care providers see particular ailments or symptoms that are associated with such an attack. The challenge is to obtain access to this information in a time period that is useful. Some states have methods of tracking this information. Wisconsin, for example, monitors some drug disbursements at state pharmacies. (In 2002, the state issued an alert when officials detected greater-than-normal sales of Imodium at Walgreens pharmacies. Fortunately, in that case, the increase was due to a sale on Imodium.[1]) A more sophisticated method, however, is New York City's Department of Health and Mental Hygiene's cutting-edge Syndromic Surveillance System, which analyzes more than 50,000 pieces of information daily, including information about 911 calls, emergency-room visits, pharmacy purchases, and worker absenteeism. The system looks for unusual patterns that can alert officials to the early stages of a disease outbreak.[2] This kind of tracking is still an exception, but it is increasing.

Section 2: Information users

Every player in homeland security is an information user. Indeed, all of the collectors described in the previous section need to use information from other sources to do their jobs well. This section describes only three information users, each with substantial but different information needs.

The Department of Homeland Security (DHS)

The DHS is intended to be the one agency accountable for protecting the U.S. from terrorism. Its mission, according to the statute that created it, is to prevent terrorist attacks, reduce the vulnerability of the U.S. to terrorism, and minimize damage from terrorist attacks in the U.S. If it is to accomplish all of this, the DHS needs virtually all information that exists about threats of terrorism and U.S. vulnerabilities.

To stop potential terrorists from entering the U.S., the Border and Transportation Security Division needs an up-to-date watch list with accurate information about suspected terrorists. The Emergency Preparedness and Response Division needs information from states and localities about local emergency capabilities and plans. The Infrastructure Protection Office requires specific and reliable information from a variety of sources about infrastructure vulnerabilities and specific threats to infrastructure. In fact, each operational entity in the DHS must have significant information beyond what it collects itself to do its job.

In addition, to provide useful threat advisories and warnings to state and local government, the private sector, and the public, the DHS needs specific, accurate, reliable, and timely warning information about terrorist plans. And, because it is the one entity that must see the full picture about terrorism in order to set its policies and priorities, the DHS must also have a steady diet of long-term strategic analysis about terrorist plans, trends, and methods.

Because the DHS has operational responsibility for all of these homeland security functions, it is in the best position to know and direct what intelligence and analysis it needs to do its job. Whatever the respective responsibilities of the DHS Information Analysis Office and the TTIC, the DHS will have to receive a massive and steady stream of every kind of homeland security information. This will have to include the information from other federal agencies and the state and local governments described in Section 1, and from the private sector.

The Department of Defense Northern Command (NORTHCOM)

The U.S. Northern Command, established in October 2002, assumed responsibility for the U.S. military's homeland security activities within the U.S. The Northern Command's headquarters are at Peterson Air Force Base in Colorado Springs. The Northern Command is one of nine combatant commands in the U.S. military. (These regional commands include personnel from all four military services under the command of a single, senior flag officer.) The geographical scope of the Northern Command's responsibility includes the continental U.S., Alaska, Canada, Mexico, parts of the Caribbean, and U.S. coastal waters out to 500 nautical miles. The command's geographic focus on the domestic U.S. is a significant departure for the U.S. military, which has focused on overseas warfare since the Civil War.

The Northern Command is very new, and the precise role it will play in homeland security is not yet clear. The Northern Command's mission is as follows: (1.) to conduct operations to deter, prevent, and defeat threats and aggression aimed at the U.S. within the area of its responsibility; and (2.) to provide military assistance—including consequence-management operations—to civilian authorities.[3] The assistance mission—supporting civilian authorities in responding to, and managing the consequences of, natural and man-made disasters—is not new. The DoD has played a significant support role in security for major domestic events such as the

[1] *Strengthening Federal-State Relationships to Prevent and Respond to Terror: Wisconsin*, Dennis L. Dresang, The Century Foundation, June 1, 2003, http://www.tcf.org/publications/homeland_security/kettlpapers/Dresang.pdf

[2] "An Early Warning System for Diseases in New York," Richard Perez-Pena, *New York Times*, April 4, 2003.

[3] See http://www.northcom.mil.

Olympics and Super Bowls, and after disasters, including September 11. The deterrence, prevention, and defeat role is less clearly defined and still evolving.

In carrying out its missions, particularly its responsibility to deter, prevent, and defeat threats to the U.S., the Northern Command will need significant intelligence, from a range of sources, on terrorist threats to the U.S. One of the principal functions of the Northern Command staff is to anticipate terrorist plots and develop plans for responding to them. This requires intelligence from all sources that is as complete as possible. The Northern Command has created its own Combined Intelligence Fusion Center at its headquarters in Colorado, where analysts and officials from a number of DoD and other agencies review and analyze threat information from foreign and domestic sources.

State and local agencies

State and local governments also have a great need for homeland security information, but in their case the full picture will not always be necessary. These governments need the kinds of information that allow them to protect the people, infrastructure, and property in their communities and to contribute effectively to prevention and response efforts.

State and local police, fire, and emergency officials must have accurate and timely information about threats to their area. If the warning is general or vague, these officials cannot make informed decisions about what to protect. Without specific information about methods the terrorists are using or targets they are interested in, these officials can try to cover everything—but given limited resources, they will most likely end up making a best guess. Although these warnings must be as specific as possible, they rarely will need to contain source-identifying information. State and local officials can and should rely on the federal government to make credibility decisions about intelligence sources.

Similarly, police and security personnel can be much more effective at lending their eyes and ears to prevention of terrorism if they know what to watch for. If they are told to look for terrorists who are lurking at rail yards or looking for hazardous chemicals, they will be more useful than if they are told simply to watch for terrorists in their areas. Again, such warnings will rarely need to contain source-sensitive information. In some cases, when local police departments are participating in counterterrorism law enforcement investigations, there is a greater need for specific information. This has led to problems because only very few of these officials have security clearances. Still, many of these concerns can be addressed with use of sanitized intelligence.

If a biological terrorist attack occurs, local health departments and health care officials will need information to handle it and reduce its impact. Doctors need almost real-time notice about symptoms to look for and how to handle these diseases. Public-health and other state officials need accurate and timely information to make decisions about quarantines and other possible precautions to prevent epidemics.

One issue that state and local government entities face in getting information from the federal government is what some refer to as the Gray Davis problem.[4] Federal government players fear that if they provide local officials with more information, that information will be revealed or misused for political reasons, sometimes to the detriment of investigations and public safety.

Section 3: Reasons to protect information

Protecting information that could harm national security if disclosed

Maybe the greatest challenge for an effective homeland security information network is to find a way to share information that is currently restricted because of national security classification. The classification system is designed to protect certain military, foreign policy, and intelligence information that, if disclosed, could harm national security. The U.S. government seeks to protect this information by, first, having an official identify it and, second, ensuring that the information identified is shared only with personnel who have a need to know it to perform their duties and are cleared to see it by a personnel-security process. The current classification system starts with three levels of classification: “Confidential,” “Secret,” and “Top Secret.” These levels are associated with the degree of damage to national security that would result if the information were revealed. On top of these levels are a number of other

[4] The reference is to Governor Gray Davis of California’s public announcement, soon after September 11, that there were threats to the Golden Gate and other California bridges. The announcement was based on what federal officials believed to be uncorroborated and unreliable intelligence.

protections, such as Special Access Programs (SAPs) in the DoD and the Department of Energy, and Sensitive Compartmented Information (SCI) programs in the foreign intelligence agencies. These programs set up smaller, more tightly controlled lists of people who are cleared for access to certain kinds of information.

The current system of security classification is cumbersome, often misapplied, and significantly overused. Serious questions remain about the process for making initial classification decisions and about oversight of those decisions, despite some reforms in the 1990s that were based on recommendations of high-level commissions that studied the system. At the same time, the concept of "need to know" is eroding because of the increased automation of information and the ease with which it is distributed. Indeed, because terrorist networks are diverse and constantly adapting, addressing the terrorist threat requires a wide-ranging, fluid information-sharing process. This is, in some ways, incompatible with the concept of "need to know." In short, one can never really know who "needs to know" certain information.

There is no question, though, that some types of information, if disclosed, would damage national security. Despite all of the flaws of the current classification system, there is great value in what it attempts to do, which is to protect this information from disclosure. There are several categories of information that would cause damage if disclosed. They have varying degrees of relevance to a homeland security information network. Some of the categories are as follows:

1. The conduct of effective diplomacy often requires that U.S. positions on negotiations or diplomatic efforts—or sometimes even the fact of those diplomatic efforts—remain secret.

2. Technical information about the design of certain systems—such as weapons, cryptologic, and imagery systems—if revealed, can provide adversaries with the ability to avoid, counteract, or recreate these systems.

3. Revealing the sources and methods used to collect and process intelligence—from signals, imagery, people, or other sources—can compromise the usefulness of those sources and methods because adversaries can learn how to avoid them. If this happens, U.S. intelligence is damaged until an alternate source can be developed. Sometimes, the result is that very expensive collection systems are suddenly stripped of their operational value. Osama bin Laden's realization that the U.S. could intercept some satellite telephone conversations, for example, led him to stop using that communications channel except as a means to confuse and misinform U.S. intelligence.

4. Plans for the conduct of military operations, or the existence of ongoing sensitive intelligence operations, if exposed, not only will compromise those operations, but could endanger lives and cause serious damage to U.S. foreign policy.

5. Protecting the names and other identifying information about individuals who have provided information to the U.S. with the expectation that it will be held in confidence is critical. Revealing these identities can put the source and his or her family at substantial risk. In addition, the loss of sources can impede the ability of U.S. agents to collect human-source information in the future because the U.S. will not be able to assure potential sources that their identities will be protected.

   To build a homeland security network that includes the maximum amount of relevant information will require demonstrating to a national security community—whose culture strongly emphasizes secrecy—that these critical categories of information can be protected. Some distribution restrictions for particularly sensitive information are inevitable. The CIA's DO, for example, will fight to the death putting the CIA's most sensitive information on a network. To keep this compartmentalization to a minimum will require cultural changes. In particular, far greater emphasis is needed on training initial classifiers not to overclassify and to focus as much attention on effective sanitization of the intelligence as on classification. That is, they must learn how to create a report that does not include the truly sensitive information (but contains enough information to be useful to others using the network), so that it can be distributed more widely.

Protecting privacy

Americans traditionally have resisted allowing the federal government to access their private information. They fear, with some historical support, that greater government access to private information will lead to abuse. Although the free flow of information to the government and between government entities is critical to fighting terrorism, greater access by government personnel to private information about U.S. citizens' activities can create an atmosphere in which abuse of rights is easier and, therefore, more likely.

After significant abuses (by the FBI, the CIA, and military intelligence agencies, among others) in the Vietnam and Watergate eras were revealed in the early 1970s, the federal government instituted a number of reforms designed to control government behavior by restricting government collection, sharing, and use of private information on U.S. persons. Some of these restrictions were as follows:

1. The number of intelligence agencies permitted to collect information on U.S. persons was restricted. With a few exceptions, the FBI was the only agency given this role. Foreign intelligence agencies generally were prohibited, by a combination of law and Executive Branch policy, from such collection.

2. A wall was erected between law enforcement and intelligence collection. The constitutional protections provided to subjects of law enforcement collection are greater than with intelligence collection, which involves national security. To be sure that law enforcement officials did not use the less-rigorous standards for intelligence collection simply to make their job easier, there were restrictions—particularly with electronic surveillance—on use of intelligence tools or products for law enforcement.

3. The FBI was restricted by DOJ policy from collecting publicly available information simply for leads or in order to create dossiers on U.S. citizens. The FBI was required to allege some tie to a crime before it could conduct surveillance in public places, surf the Internet, or access publicly available commercial databases.

Since September 11, many of these restrictions have been relaxed, either by changes to law or policy. For example, although the FBI remains the only agency authorized to collect intelligence on U.S. persons, significantly more of that intelligence is now shared with foreign intelligence agencies. The TTIC, which is now responsible for fusing and analyzing domestic and foreign intelligence on terrorism, is under the authority of the DCI and is housed at the CIA. In addition, the USA PATRIOT Act and changes to DOJ policy allow intelligence information and tools to be used more freely by law enforcement personnel, and DOJ guidelines now permit the FBI to conduct surveillance in public places or perform Google searches, for example, without alleging criminal activity.

Relaxation of these restrictions was, for the most part, inevitable and necessary, given the importance of the free flow of information to the fight against terrorism. The challenge, though, is that there are now significantly fewer institutional protections against government misuse of private information. At the same time, advances in technology have improved immeasurably the government's ability to collect and use private information. Therefore, in designing a network that would promote free flow of information to any number of users, there must be new mechanisms for protecting private information. Technological protections that would, for example, keep private information out of the hands of officials who don't need it, and keep tabs on those who do, can play a significant role in privacy protection. New guidelines and oversight to control the behavior of officials who do have access are just as important.

Protecting the ability to arrest and successfully prosecute terrorists

Federal law enforcement officials guard information about ongoing investigations jealously, which can sometimes hamper other efforts to fight or respond to terrorism. For example, FBI officials are reluctant to share information with local officials about investigations in their region, which sometimes leaves those officials in the dark about local threats. (Health and other officials have said that FBI officials investigating the 2001 anthrax attacks handled information in a way that set back efforts to alleviate the threat to public health and safety.) Also, in the past, the FBI has resisted informing even senior national security policymakers or intelligence officials about information that it uncovers as part of an ongoing terrorism investigation.

Although some of this reluctance to share can be attributed to FBI culture and the agency's unfamiliarity with other disciplines, there are also legitimate concerns about sharing information on ongoing investigations. These investigations often are intricate and have developed over long periods of time and at great expense. If the circle of people who know about an investigation expands to include local officials, there is a risk that, intentionally or inadvertently, those officials will act on the information. Actions by local officials, such as conducting surveillance or arresting or detaining suspects in a federal investigation, could alert terrorists to the investigation.

The example of the anthrax attack demonstrates a difficult problem with counterterrorism, which is both a law enforcement and a public safety challenge. To obtain a conviction at trial, prosecutors must be able to demonstrate that evidence is what they say it is. To do this, physical evidence, crime scenes, and witnesses must be handled very carefully. Involvement with evidence by officials not involved in the investigation threatens a prosecution.

There are also legal issues with sharing some law enforcement information. Federal Rule of Criminal Procedure 6(e) prohibits law enforcement and prosecutorial officials from revealing information collected during a grand jury proceeding. The DOJ and the FBI have at times taken an overly broad view of what constitutes grand jury information. In addition, the USA PATRIOT Act clarified that Rule 6(e) does not restrict the sharing of grand jury information with federal intelligence agencies. Nonetheless, the restriction does exist, and law enforcement officials understand that to violate it could damage an eventual prosecution.

When it comes to sharing information with senior policymakers, law enforcement officials have an additional concern about protecting criminal investigations from inappropriate political influence. The reality or perception of such influence can affect the credibility and legitimacy of an eventual prosecution.

Appendix C

The Immune-System Model

by Tara Lemmey

Background

In our initial report, we stated the following: "To create a national infrastructure that is aware, robust, and resilient to the many challenges we face in the 21st century, we have to harness the power and dynamism of information technology by utilizing the strengths and mitigating the weaknesses of our networked society" (p. 11). We also identified 11 key principles for building this kind of infrastructure. Those principles included empowering local participants, creating network-aware scenarios, facilitating a connected culture, and ensuring safeguards and guidelines for protecting civil liberties. In order to achieve a dynamic infrastructure, we need to consider viable models of implementation and reasonable means for deploying these models across all of the various players in the network.

Some of the criteria we considered while looking at the models were as follows: (1.) scalability to the national level; (2.) provision for organic growth and graceful collapse; (3.) ability to take advantage of existing systems and culture; (4.) evolvability of the system based on current state; (5.) assurance that everyday operations benefit from homeland security measures; and (6.) respect for historic safeguards where possible.

The most critical elements are as follows: (1.) time optimization to allow for the most advantageous decision-making and action; (2.) effective use of the entire landscape of resources; and (3.) computational feasibility.

The problem of too much data and conflicting needs

Much of the current conversation centers on the use of data as a panacea. Our daily relationship with the Internet has encouraged the thinking that all things are "findable," meaning that, given all of the information, we should be able to find the threats in the data. The ability to find things using network technology is now simpler. With search technologies like those used by Google, one can locate data points in space and look for explicitly proposed correlations such as "Tom and Jerry and cartoons."

Discovering some patterns in the data is, of course, possible, and one should go ahead with the pursuit by reasonable computational means. But automating the discovery of all implicit correlations in a data set (in order to generate all—and only—the significant correlations) is, in general, intractable. And because automating the discovery of implicit correlations in the data is intractable, generating a complete solution is also intractable.

For example, if we had 10 terms in a data set and we were looking for all significant pair-wise correlations, we would quickly find that we would have to look at 100 possible relationships. For three term correlations, we would have to consider 1,000 possibilities. Given a 1,000,000-term data set and looking for three term correlations, we're looking at something on the order of 10^{18} possible correlations. The sheer volume of data coming from all of the possible sources creates such a high degree of noise and computational complexity that the likelihood of finding useful correlations is nil. On the other hand, after an event has happened, the correlations we'd be looking for would be explicit. There are ways around this combinatorial explosion, which we explore here, but the key is to have an idea of what you want before you start.

In addition, there are a number of issues that will limit the application of pattern search in large-scale databases. For example, all of the data is never going to end up in the same place, and some data will never show up anywhere in such a searchable format. Furthermore, we have already seen congressional distaste for approaches like that of the Department of Defense's Terrorism Information Awareness program, and we can expect that resistance to global data fishing–expeditions will only harden over time. The recommendations of our Task Force, therefore, will have to balance privacy and security concerns in whatever solutions we propose.

Another major issue is the conflicting set of requirements and constraints on the use of data at the various governmental and nongovernmental agencies. As detailed in "A Primer on Homeland Security Players and Information" (Appendix B), although some of the limiting factors can be overcome through policy or culture modifications, the bulk of these limitations are appropriate to protect privacy, sources, methods, successful prosecutions, military operations, etc. These requirements limit the ability of some data to be shared in raw form, but they should not limit our ability to act on—or add to—the data if the systems are functioning using all available resources.

A biological approach to resilient information systems

We can learn something about how to address the complexity of the threat-identification problem by looking at the human immune system, which has evolved in several distinct phases as it has had to cope with the complexity of deterrence of foreign biological invaders. The hard-won evolutionary adaptations of the immune system are directly relevant to our task. That said, the immune system should serve as inspiration, not as a direct analogy.

Evolutionarily, the immune system has faced the challenge of distinguishing between "self" and "nonself." It is estimated that the immune system must recognize on the order of 10^{16} different kinds of pathogens, while there are only about 10^{5} cells that make up our bodies. How does the immune system go about identifying the difference between what should and what should not be present in our bodies?

The immune system discovered a neat trick. In early development, it produces an enormous diversity of lymphocites carrying randomly generated antibodies, enough to recognize on the order of 10^{16} cells, including the cells that should be in the body. Then, it runs all of these lymphocites through the thymus, where all cell types that are supposed to be in the body are represented. Any lymphocite that responds to a cell in the thymus is destroyed. The only lymphocites that make it out of the thymus are those that do not respond to the body's own cells. Thus, the immune system explicitly trains up on cells that are supposed to be there, and treats everything else as a potential invader that needs to be checked.

When we are born, the antibodies produced by the immune system are somewhat sloppy recognizers—they will bind to anything that looks similar to the pathogen they are specifically generated to recognize. Later in life, as cells respond to foreign invaders, they become more and more refined in their responses to the specific invaders that they have encountered, thus fine-tuning their recognition function.

The point of looking at the immune system is to learn what it has to tell us about the tractability of different approaches to threat detection and intervention. Current government policy is to try to determine all of the bad things that could happen, a task which is in principle intractable. To take the lesson from the immune system, we should apply our information-handling resources to the task of explicitly representing the "normal" behavior of systems, filtering that out, and then paying particular attention to anything that is left.

Patterns in the data must be compared to a model to determine whether they are good patterns or bad patterns. The question is simply whether that model will be of the bad or the good patterns. The immune system teaches us that trying to produce an adequate model of the bad is intractable. Therefore, we should build a model of the good, and treat as suspect any event that does not fit that model. This is a far more tractable approach, and one that can rely on the local expertise of every public-safety worker out there in determining what normal behavior means for the systems under their care.

CENTRAL LESSONS FROM THE IMMUNE SYSTEM

1. A central insight from the immune system concerns the tractability of explicitly modeling good versus explicitly modeling bad.

2. There is a critical need for a greater understanding of "self" (the "normal" behavior of the systems under one's care) by all players at the federal, state, local, and private sector levels. Some surveillance systems are already based on this tenet of characterizing "self," though perhaps not intentionally. Credit-scoring and financial-systems models are examples of such an approach. Applying ourselves to representing the normal behavior of our systems is a specific and accomplishable task we can undertake now.

3. No two immune systems are identical. A population consists of a diverse set of representations of both "self" and "nonself." This implies that a collective homeland immune system should benefit enormously from its large population of local experts, who create diversity in the analyses and perspectives brought to bear on the problem of threat detection and attack prevention. This makes the case for distributed analysis—including analysis at the local level.

4. When the immune system recognizes a foreign pathogen, it produces a great many variants on the

pattern and circulates them so that anything similar will be flagged for attention. Circulating variants of a detected threat, or a generalized threat schema based on the variants, can allow people serving as low-level sensors to become more sophisticated in their signal-seeking.

5. The health sciences have learned that, because the immune system is so elegant, one of the most productive ways to improve health protection is to help the immune system to do its job more effectively. Vaccinations, antitoxins, and other immune-boosting response mechanisms improve the system's efficiency. We should consider methods of tuning up the systems that we already have in place and of training our sensors to be far more responsive to signals and triggers.

6. Scenario-based training helps. Vaccination makes use of the immune response, which boosts the immune system's ability to recognize a potentially lethal threat. It does this by presenting that threat in a nonlethal form.

7. "Self," or "normal operation," can and should have a broad definition. A good deal of what might be considered "abnormal" is not necessarily bad. The context is critical and best supplied by the most local sensor.

8. The immune system strives to achieve a delicate balance between under- and overprotection: If it is too aggressive in attacking entities, it risks attacking things that are supposed to be there, leading to autoimmunity diseases; if the immune system is too tolerant, it will fail to protect the body against potentially dangerous pathogens. Thus, a challenge is to use the extended network to approach the homeland security problem with sufficient aggressiveness, while maintaining proper respect for privacy and other core civil liberties: In the process of protecting against terrorist threats, we must not produce a system that results in a form of social autoimmunity.

9. The primary "success" of HIV/AIDS lies in the virus's ability to attack and disable the immune system itself, thus dismantling the system that recognizes foreign invaders. In the same way, the primary recognizers in our own homeland security system are vulnerable.

10. We must keep in mind that despite all of its complexity, elegance, and sophistication, there is not perfect coverage in the immune system, and some pathogens still manage to get through and cause a great deal of damage.

Internet resources
http://www.howstuffworks.com/immune-system.htm
http://www.niaid.nih.gov/final/immun/immun.htm
http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=11390983
http://www.bbc.co.uk/health/immune/
http://medic.med.uth.tmc.edu/edprog/Immuno/Immune.Works.2003_filesframe.htm#
http://www.cdc.gov/od/nvpo/intro4.htm
http://www.niaid.nih.gov/publications/vaccine/undvacc.htm
http://press2.nci.nih.gov/sciencebehind/immune/immune00.htm
http://uhaweb.hartford.edu/BUGL/immune.htm
http://answers.google.com/answers/main?cmd=threadview&id=218013

Books
Cellular and Molecular Immunology, by Abdul K. Abbas, Jordan S. Pober, and Andrew H. Lichtman
Molecular Biology of the Cell, by Bruce Alberts et al.
Immunobiology: The Immune System in Health and Disease, by Janeway C. A., P. Travers, M. Walport, and M. Shlomchik
How the Immune System Works, by Lauren Sompayrac

Appendix D

Information Vignettes

The following information vignettes describe different types of information that might come into the possession of players in our nominal network. Creating these vignettes using concrete scenarios allowed the Task Force to consider how information should be analyzed and shared to maximize its utility and to optimize the capabilities of the players in the network.

Vignette 1: Information-sharing between and within government agencies

A BIOTERROR THREAT

A source of the FBI's Chicago field office tells his handler that plans are underway to create a national crisis by infecting small numbers of individuals in disparate locations with a virulent virus acquired from sick hogs. The informant says that someone will drive from Chicago to St. Louis, transporting a cooler containing a number of sealed packages, and will hand over the cooler in St. Louis to another operative, who will then drive to an undisclosed location. The source believes the packages could contain the virus. The FBI considers the source to be reliable, but does not believe he could have access to this kind of information.

SECRET

Federal Bureau of Investigation
Chicago Field Office

March 30, 2003

URGENT REPORT

TO: Director Mueller
Deputy Director Gebhart
Executive Assistant Director D'Amuro
Assistant Director Mefford
Section Chief Doe
Unit Chief Bob/Bob

FROM: SAC Johnson/TBJ

RE: Case no. 176543-E

In the course of investigating alleged smuggling operations (electronics, clothing, and CDs) being carried out by a group of local, ethnic Middle Easterners representing themselves as a "mutual assistance group," Special Agent Morrison developed the following information:

According to a sensitive source who has been reliable in giving the FBI timely leads on the smuggling activities undertaken by a number of males of Middle East origin, there is a plan afoot to spread a sickness around the U.S. and create a national crisis. The idea is to infect a number of people in different cities around the country with a virus that terrorist scientists have extracted from sick hogs.

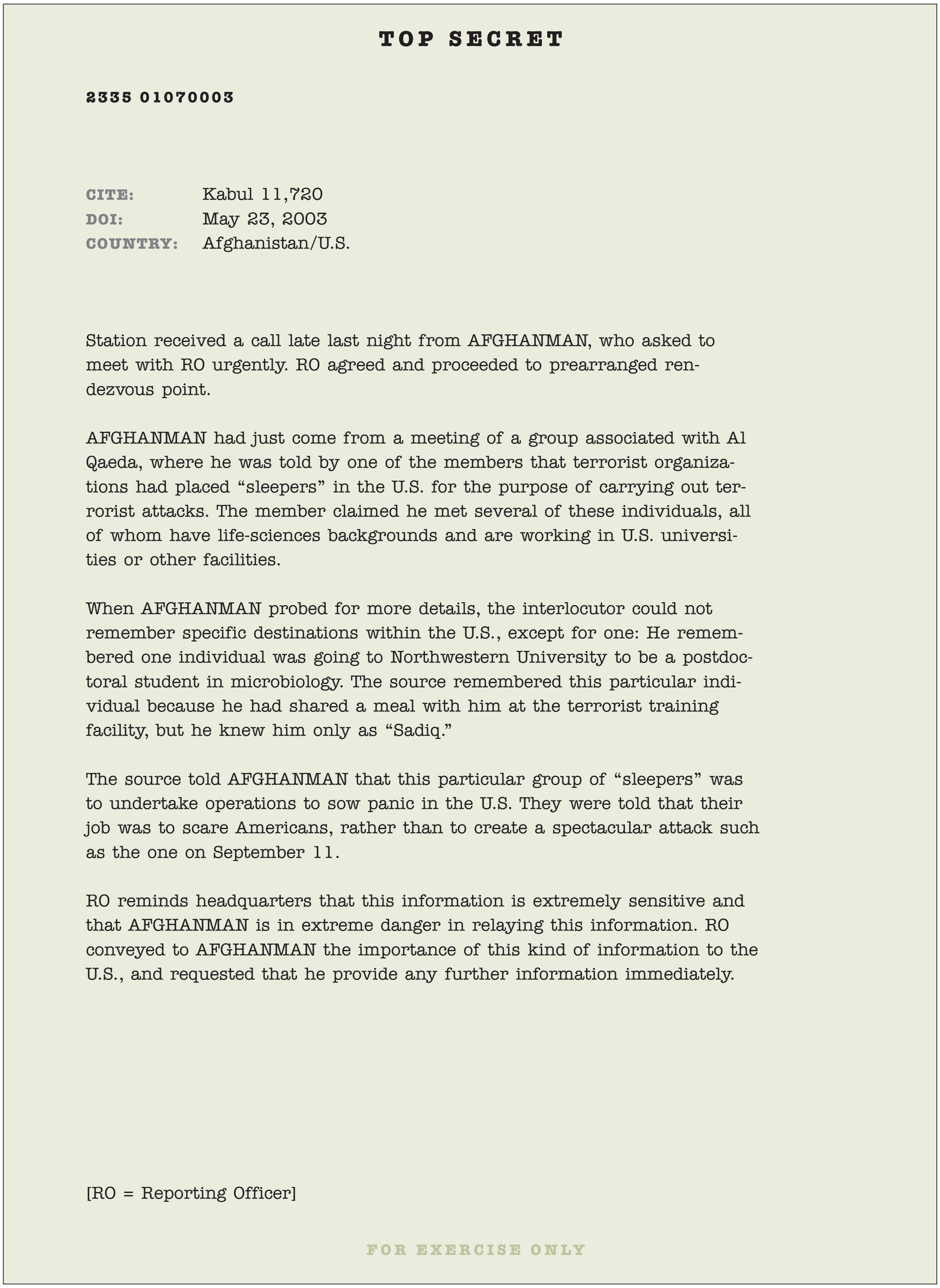
The source then told Mr. Morrison that someone would drive from Chicago to St. Louis with a cooler containing several packages, and would hand the cooler over to someone in St. Louis. That individual would then drive somewhere else and hand the cooler to another operative. The source believed the packages could contain the virus.

The source is placed in the middle of criminal activity related to smuggling. The group with which he is connected appears to be a regular criminal organization with no signs of terrorist connections. Special Agent Morrison does not believe, therefore, that this source would have access to terrorist plans.

Given the headquarters guidance to lean forward on any matters relating to terrorism, however, we are passing this on in case it helps to connect some dots.

[SAC = Special Agent in Charge]

FOR EXERCISE ONLY



TOP SECRET

2335 01070003

CITE: Kabul 11,720
DOI: May 23, 2003
COUNTRY: Afghanistan/U.S.

Station received a call late last night from AFGHANMAN, who asked to meet with RO urgently. RO agreed and proceeded to prearranged rendezvous point.

AFGHANMAN had just come from a meeting of a group associated with Al Qaeda, where he was told by one of the members that terrorist organizations had placed "sleepers" in the U.S. for the purpose of carrying out terrorist attacks. The member claimed he met several of these individuals, all of whom have life-sciences backgrounds and are working in U.S. universities or other facilities.

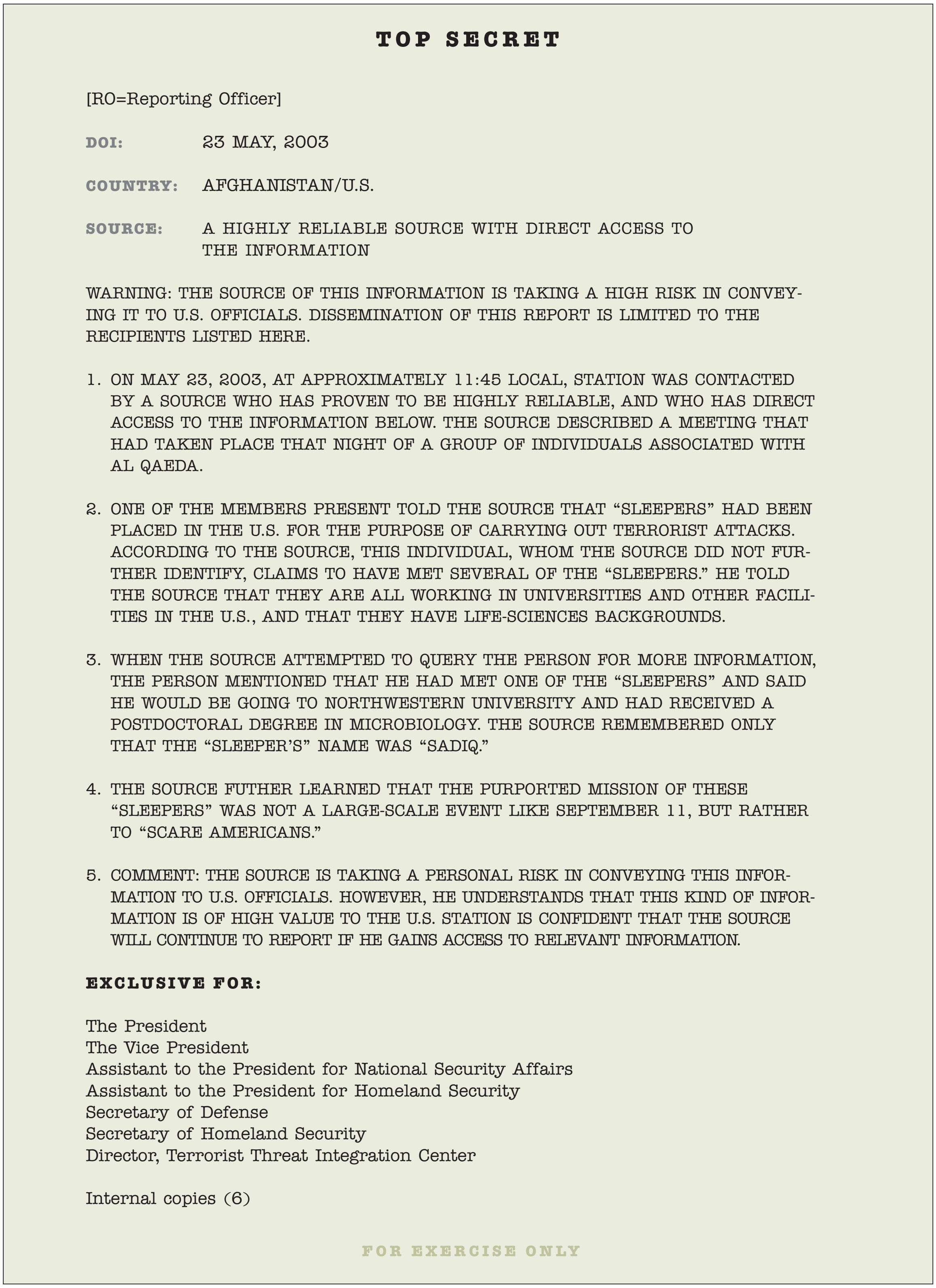
When AFGHANMAN probed for more details, the interlocutor could not remember specific destinations within the U.S., except for one: He remembered one individual was going to Northwestern University to be a postdoctoral student in microbiology. The source remembered this particular individual because he had shared a meal with him at the terrorist training facility, but he knew him only as "Sadiq."

The source told AFGHANMAN that this particular group of "sleepers" was to undertake operations to sow panic in the U.S. They were told that their job was to scare Americans, rather than to create a spectacular attack such as the one on September 11.

RO reminds headquarters that this information is extremely sensitive and that AFGHANMAN is in extreme danger in relaying this information. RO conveyed to AFGHANMAN the importance of this kind of information to the U.S., and requested that he provide any further information immediately.

[RO = Reporting Officer]

FOR EXERCISE ONLY



TOP SECRET

[RO=Reporting Officer]

DOI: 23 MAY, 2003

COUNTRY: AFGHANISTAN/U.S.

SOURCE: A HIGHLY RELIABLE SOURCE WITH DIRECT ACCESS TO THE INFORMATION

WARNING: THE SOURCE OF THIS INFORMATION IS TAKING A HIGH RISK IN CONVEYING IT TO U.S. OFFICIALS. DISSEMINATION OF THIS REPORT IS LIMITED TO THE RECIPIENTS LISTED HERE.

1. ON MAY 23, 2003, AT APPROXIMATELY 11:45 LOCAL, STATION WAS CONTACTED BY A SOURCE WHO HAS PROVEN TO BE HIGHLY RELIABLE, AND WHO HAS DIRECT ACCESS TO THE INFORMATION BELOW. THE SOURCE DESCRIBED A MEETING THAT HAD TAKEN PLACE THAT NIGHT OF A GROUP OF INDIVIDUALS ASSOCIATED WITH AL QAEDA.

2. ONE OF THE MEMBERS PRESENT TOLD THE SOURCE THAT "SLEEPERS" HAD BEEN PLACED IN THE U.S. FOR THE PURPOSE OF CARRYING OUT TERRORIST ATTACKS. ACCORDING TO THE SOURCE, THIS INDIVIDUAL, WHOM THE SOURCE DID NOT FURTHER IDENTIFY, CLAIMS TO HAVE MET SEVERAL OF THE "SLEEPERS." HE TOLD THE SOURCE THAT THEY ARE ALL WORKING IN UNIVERSITIES AND OTHER FACILITIES IN THE U.S., AND THAT THEY HAVE LIFE-SCIENCES BACKGROUNDS.

3. WHEN THE SOURCE ATTEMPTED TO QUERY THE PERSON FOR MORE INFORMATION, THE PERSON MENTIONED THAT HE HAD MET ONE OF THE "SLEEPERS" AND SAID HE WOULD BE GOING TO NORTHWESTERN UNIVERSITY AND HAD RECEIVED A POSTDOCTORAL DEGREE IN MICROBIOLOGY. THE SOURCE REMEMBERED ONLY THAT THE "SLEEPER'S" NAME WAS "SADIQ."

4. THE SOURCE FUTHER LEARNED THAT THE PURPORTED MISSION OF THESE "SLEEPERS" WAS NOT A LARGE-SCALE EVENT LIKE SEPTEMBER 11, BUT RATHER TO "SCARE AMERICANS."

5. COMMENT: THE SOURCE IS TAKING A PERSONAL RISK IN CONVEYING THIS INFORMATION TO U.S. OFFICIALS. HOWEVER, HE UNDERSTANDS THAT THIS KIND OF INFORMATION IS OF HIGH VALUE TO THE U.S. STATION IS CONFIDENT THAT THE SOURCE WILL CONTINUE TO REPORT IF HE GAINS ACCESS TO RELEVANT INFORMATION.

EXCLUSIVE FOR:

The President
The Vice President
Assistant to the President for National Security Affairs
Assistant to the President for Homeland Security
Secretary of Defense
Secretary of Homeland Security
Director, Terrorist Threat Integration Center

Internal copies (6)

How the information would likely be handled today

The FBI electronic communication

Assuming that established procedures are followed, this report would go to the Chicago Joint Terrorism Task Force (JTTF)[1] and to the FBI's headquarters in Washington, DC. There, one of two things would happen: Either the information from the report would be transferred to an Intelligence Information Report (IIR)—a formal intelligence report that FBI headquarters distributes internally and externally, at least to Terrorist Threat Integration Center (TTIC)—or the TTIC might become aware of this information via an informal email from personnel at FBI headquarters. (Given the undeveloped nature of the information in this report, however, and the fact that there is a continuing field investigation, this report might not become an IIR.)

Assuming that the FBI did prepare an IIR, the IIR would be placed on TTIC Online, a top-secret, secure network for counterterrorism information. (TTIC Online is now available to the appropriately cleared individuals in the intelligence community who have access to the network.)[2] The report would then be available to cleared Department of Homeland Security (DHS) personnel, who would pull it off of the TTIC Online system. In addition, if the TTIC produced an analytical product that included the FBI information, that product would go to the DHS (as discussed below). TTIC Online is also available to all JTTFs nationwide that are equipped with Sensitive Compartment Information Facilities (SCIFs)—this is most, if not all of the JTTFs. Therefore, cleared JTTF personnel could have pulled this report off of that system. However, neither the FBI report nor the information it contains would have gone to the Chicago Police Department or to other state or local law enforcement, health, or agricultural agencies around the country. It is important to note that dissemination to JTTFs is not the same as dissemination to the agencies represented on the JTTFs, since the agency representatives agree not to share information with their own agencies without the permission of the FBI.

The CIA report

Before the information contained in this report could be shared outside the Directorate of Operations (DO)—even within the CIA itself—it would have to be sanitized to remove all code words and any information that could help identify the source or place him in a specific setting such as a particular meeting.

A few headquarters personnel would know who AFGHANMAN was. Nonetheless, this information would not be shared with policymakers or, normally, with analysts. The sanitized report, which would still be classified "Top Secret," would contain a sentence describing the reliability of the source and his likely access to the information.

Members of the intelligence community who work on homeland security matters probably would first become aware of this CIA intelligence report through a "gist" published in the daily Terrorism Threat Matrix (TTM). The TTM is a compilation, without analysis, of the terrorist-threat information received within the previous 24 hours, which is distributed to senior officials. This matrix is available to all federal government intelligence agencies with a homeland security mission. Only personnel with "Top Secret/Code Word" clearances may view the TTM. The information in the CIA report would have been discussed at a morning secure video teleconference among designated officials from the homeland security agencies.

The Terrorist Threat Information Center analysis

If all went according to procedures, at this point, analysts in the biological weapons analysis group at the TTIC and/or the Counterterrorist Center (CTC) would probably put the information in the CIA report together with the information in the preceding FBI IIR or email notification. The TTIC might note this in its President's Terrorism Threat Report (PTTR)—the agency's daily analytic report for the President—after receiving permission from the originators of the information (in this case the FBI and the CIA). The TTIC might also inform personnel at the DHS, the CIA, and FBI headquarters of the two pieces of reporting.

[1] JTTFs are led by the FBI, and comprise representatives from other federal agencies as well as state and local law enforcement. They are usually headed by the deputy at the local FBI field office.

[2] Because TTIC Online contains intelligence at the "Top Secret/Code Word" level, the network access terminals must be located in special Sensitive Compartmented Information Facilities (SCIFs).

Additional sharing needed

In the case of this vignette, the two bits of information would most likely find their way to a common place in the federal government—probably the TTIC, and also the FBI and the CTC—where they could be correlated and analyzed. The most significant failure that this vignette demonstrates is that neither the initial reports nor the analytical product would likely be shared with state and local actors. To make the fullest and most effective use of the information, and to optimize all of the players in the network, some version of the information would need to be shared with state and local entities so that they might serve as additional sensors and collect and contribute additional information. In this scenario, state and local law enforcement should know that coolers could be a vehicle for transporting biological weapons, and local health and agricultural agencies should know to look for signs and symptoms of a hog virus. If any of these agencies came across relevant information, that information could be brought into the network and shared in some fashion with other relevant entities.

The necessary additional output includes a sanitized, unclassified TTIC analysis (any information that might reveal the FBI or CIA source or otherwise impede their investigation and collection efforts would be removed) that would go to the CDC and to regional, state, and local entities responsible for health and agricultural matters, and perhaps also to private sector agricultural entities, probably via the DHS. Also in this case, the TTIC information should flow to state and local law enforcement agencies, most likely from the JTTFs. It is important that the task forces or entities receiving such information have common practices and guidelines for information flow, security, and reporting. In addition, sanitized information should include a marker indicating the name of a person who can be contacted for further information.

In this vignette, it might also be the case that the initial FBI information and/or CIA information should be shared by the DHS with the CDC and other entities, and by the JTTFs with state and local law enforcement, to activate those sensors even earlier, without waiting for the TTIC's analysis of the two pieces of information together. Whether to do so in any particular case requires judgment: It is important not to overload the system of sensors with too much noise (information that might not be important), but also important to make sure the sensors are quickly alerted to signals (credible, actionable information) when they are distinguishable from noise.

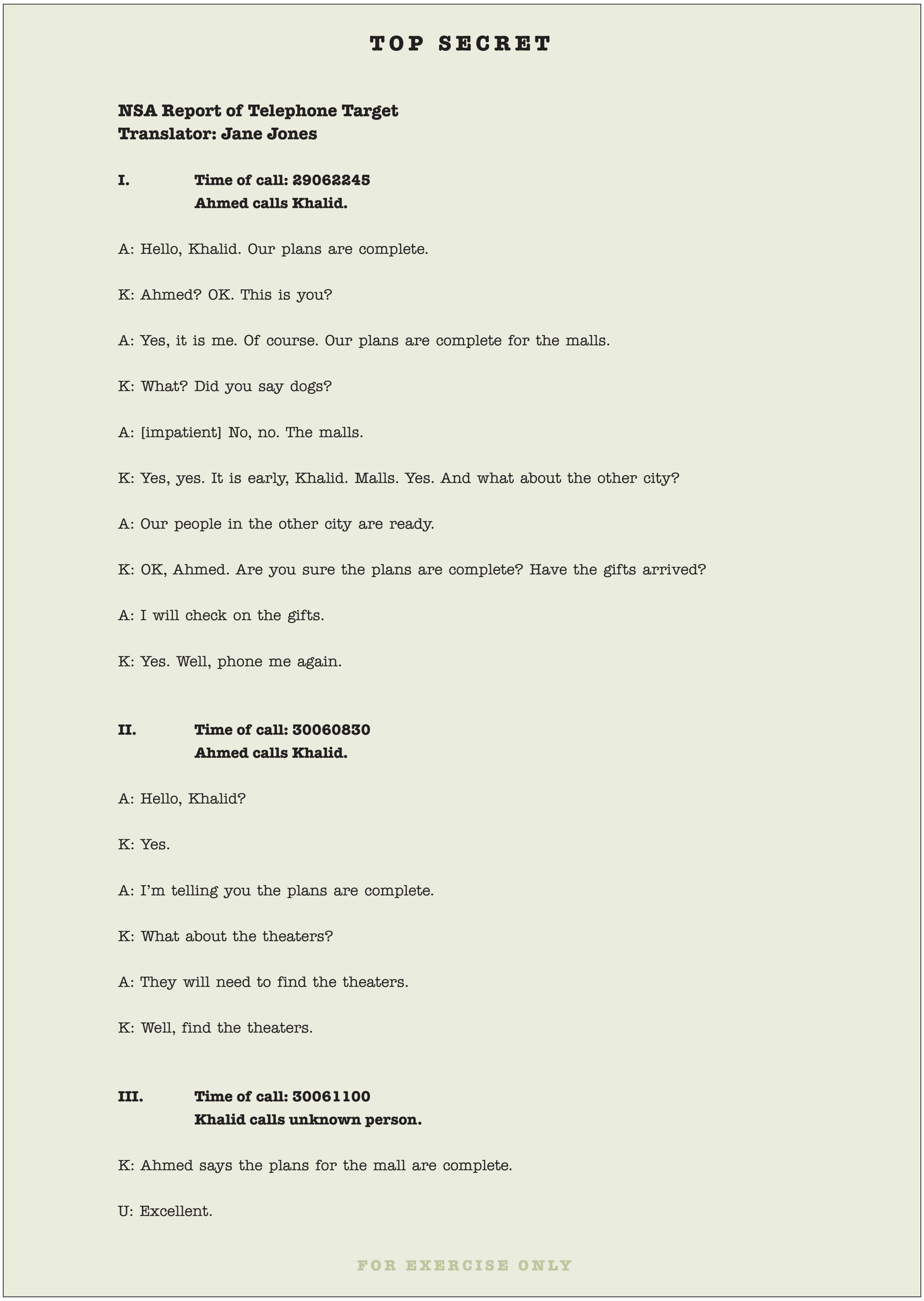
If the additional output (at least the information from the TTIC analysis, and possibly the original FBI or CIA information before that) is communicated effectively with the state and local entities and the CDC, they will be sensitized to collect more useful information. This second level of input from these entities must come back to the federal government, probably again through the DHS (from non–law enforcement entities) and the JTTFs (from law enforcement). Again, it is vital that there be some uniformity of reporting format and interoperability of communication methods.

Vignette 2: Information-sharing between and within government agencies

A THREAT TO MALLS

The National Security Agency (NSA) issues a report saying that sensitive intercepted communications (in this case, phone calls) among known Al Qaeda leaders abroad indicate that final preparations are being made for terrorist operations against targets in the U.S. Speakers have mentioned "malls," or perhaps "the Mall," and have referred to "the other city." In one conversation, they have also mentioned "movie theaters."

Meanwhile, two months prior, the Pittsburgh police received a tip from an anonymous source saying that one Mr. William Joseph, a local businessman, is involved in a plot to stage some sort of terrorist attack in the city. The Pittsburgh police shared the report with the FBI field office. The police and the FBI have since met to discuss the case and have agreed to share the investigatory and surveillance burdens. The FBI will obtain any subpoenas that are required.



TOP SECRET

NSA Report of Telephone Target
Translator: Jane Jones

I. Time of call: 29062245
Ahmed calls Khalid.

A: Hello, Khalid. Our plans are complete.

K: Ahmed? OK. This is you?

A: Yes, it is me. Of course. Our plans are complete for the malls.

K: What? Did you say dogs?

A: [impatient] No, no. The malls.

K: Yes, yes. It is early, Khalid. Malls. Yes. And what about the other city?

A: Our people in the other city are ready.

K: OK, Ahmed. Are you sure the plans are complete? Have the gifts arrived?

A: I will check on the gifts.

K: Yes. Well, phone me again.

II. Time of call: 30060830
Ahmed calls Khalid.

A: Hello, Khalid?

K: Yes.

A: I'm telling you the plans are complete.

K: What about the theaters?

A: They will need to find the theaters.

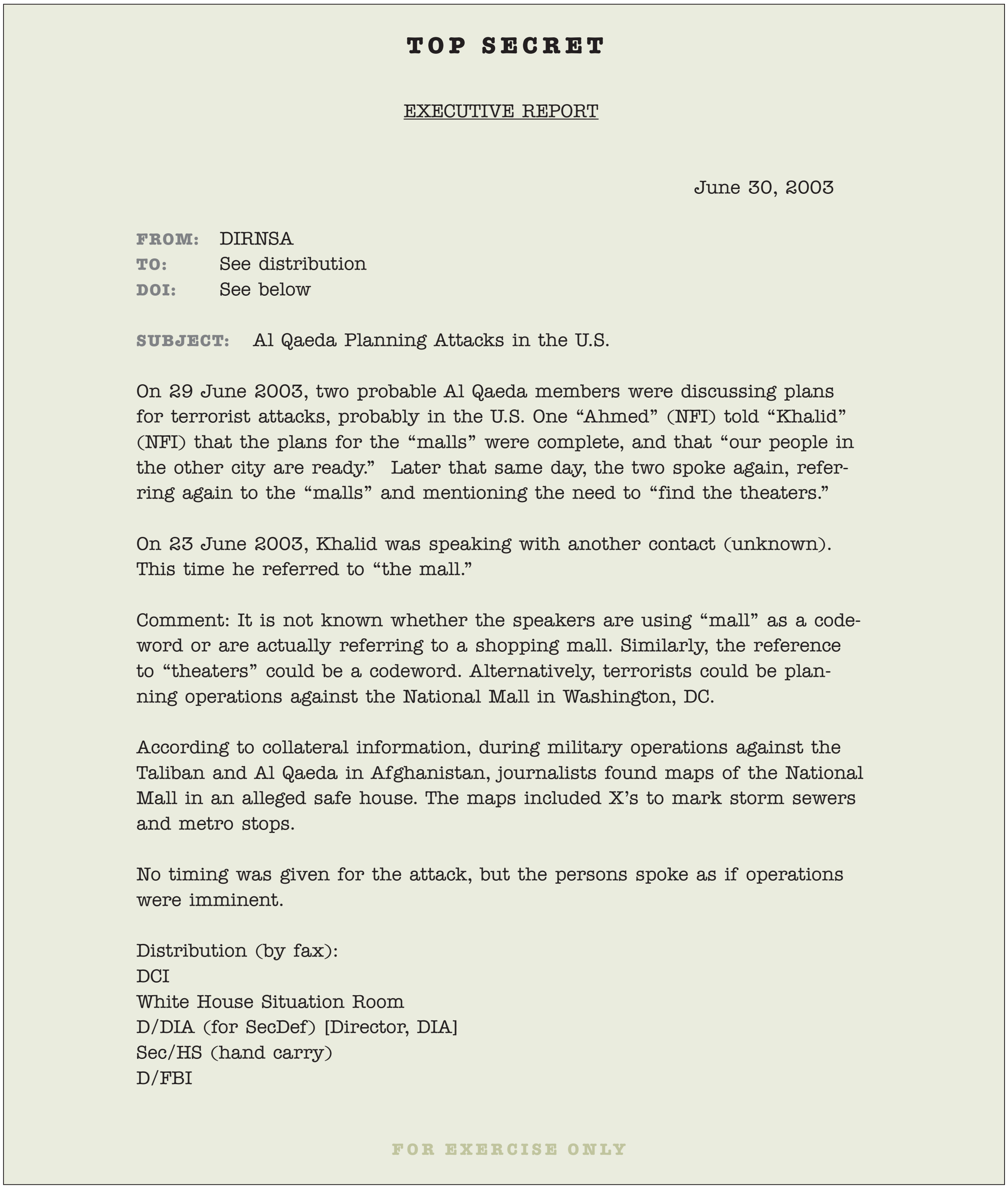
K: Well, find the theaters.

III. Time of call: 30061100
Khalid calls unknown person.

K: Ahmed says the plans for the mall are complete.

U: Excellent.

FOR EXERCISE ONLY



TOP SECRET

EXECUTIVE REPORT

June 30, 2003

FROM: DIRNSA
TO: See distribution
DOI: See below

SUBJECT: Al Qaeda Planning Attacks in the U.S.

On 29 June 2003, two probable Al Qaeda members were discussing plans for terrorist attacks, probably in the U.S. One “Ahmed” (NFI) told “Khalid” (NFI) that the plans for the “malls” were complete, and that “our people in the other city are ready.” Later that same day, the two spoke again, referring again to the “malls” and mentioning the need to “find the theaters.”

On 23 June 2003, Khalid was speaking with another contact (unknown). This time he referred to “the mall.”

Comment: It is not known whether the speakers are using “mall” as a codeword or are actually referring to a shopping mall. Similarly, the reference to “theaters” could be a codeword. Alternatively, terrorists could be planning operations against the National Mall in Washington, DC.

According to collateral information, during military operations against the Taliban and Al Qaeda in Afghanistan, journalists found maps of the National Mall in an alleged safe house. The maps included X’s to mark storm sewers and metro stops.

No timing was given for the attack, but the persons spoke as if operations were imminent.

Distribution (by fax):
DCI
White House Situation Room
D/DIA (for SecDef) [Director, DIA]
Sec/HS (hand carry)
D/FBI

FOR EXERCISE ONLY

SECRET

FROM: DIRNSA
TO: 06292245Z

SUBJECT: Terrorists Discuss Plans

On 29 June 2003, two probable Al Qaeda members were discussing plans for terrorist attacks probably in the U.S. In the course of the conversation, the two referred to "malls" and "theaters." In another conversation, one of them referred to "the mall."

COMMENT: The probable terrorists could be using "malls" and "theaters" as code words. Alternatively, they could be referring to the National Mall. According to collateral information, during military operations against the Taliban in Afghanistan, journalists found maps of the National Mall in an alleged safe house.

FOR EXERCISE ONLY

City of Pittsburgh

BUREAU OF POLICE

REPORT OF TIP

ZONE: 6
DATE: April 25, 2003
TIME: 2:23 p.m.
SOURCE: Anonymous

A male caller to Zone 6, who would not give his name, and was calling from a pay phone on the corner of Murray and Forbes, said he wanted to report some "terrorist activity." He said he had observed "suspicious activity" at the house of one Mr. William Joseph, who resides at 2455 Hastings Street.

The caller said that a number of cars appear at the Joseph residence each Thursday night, several males come in each car, and he can hear "Arab" music coming from the house. Caller said he decided to investigate on his own. About 11 p.m. the previous evening, caller had entered the Joseph yard and peered in the window, where he saw a gathering of males, and pictures on the wall of Osama bin Laden. Also on the wall were maps of Pittsburgh marked with two large red X's. Caller said one of the X's was in the area of the Robinson Center (a shopping mall). Caller added that at least two cars had Maryland license plates.

Follow-up: Officers will locate residence and will mount surveillance on the next Thursday.

FOR EXERCISE ONLY

Federal Bureau of Investigation
Pittsburgh Field Office

May 6, 2003

LAW ENFORCEMENT INFORMATION

Letterhead Memorandum

SUBJECT: Meeting with Pittsburgh Police to Discuss Issues of Mutual Interest

Officers Smith and Brown, Pittsburgh Police Bureau, Zone 6, met with field agents of this office today to discuss a number of items of particular interest to both organizations. Some involved ongoing investigations in the area of organized crime. The subject of this memorandum concerns a possible terrorist threat brought to the attention of the police by an anonymous source.

The Pittsburgh Police Bureau received an anonymous call from an individual presumed to be a neighbor reporting "suspicious activity" possibly related to terrorism. The assumed neighbor had reported weekly gatherings (each Thursday evening) at a residence (2455 Hastings Street, Pittsburgh), involving a dozen or so males. The neighbor reported "Arab" music coming from the house, pictures of Osama bin Laden on the wall, and a map of Pittsburgh marked with two red X's. The informant had told police he thought one X was in the area of the Robinson Center, a shopping mall with many retail stores, restaurants, and several movie theaters. The owner of the house is William Joseph.

The caller said that at least two of the cars had Maryland license plates.

Police subsequently surveilled the house on the following Thursday. They corroborated the source's description of the cars and their occupants. One officer approached the side of the house to determine if he could see in a window. He reported hearing Middle Eastern music coming from the house, but was unable to see under the window shades, which were almost completely drawn.

The Pittsburgh field office will undertake to investigate Mr. Joseph's bona fides (citizenship/status, employment, contacts with known terrorists and terrorist organizations). Any subpoenas or court orders required for credit card, travel, telephone records, etc. will be handled by the FBI. Affidavits may be requested from the Pittsburgh Police Bureau in conjunction with requests for subpoenas.

The Pittsburgh Police Bureau will continue surveillance on the residence in question, and will provide license-plate traces on all cars visiting. The police will extend their surveillance time to include Thursday evenings plus other times.

Our next scheduled meeting is May 20, 2003.

FOR EXERCISE ONLY

How the information would likely be handled today

The National Security Agency transcript and report

The transcript of the intercepted conversation would go to an NSA analyst, who would produce a report. The full report would be classified "Top Secret," but a second "Secret" version might also be prepared. The fact of the NSA's access to the phones of at least one of these individuals would be considered extremely sensitive. The "Top Secret" report would go to the DHS, the TTIC, the FBI, the CIA, the Department of Defense (DoD), and the White House, and it—or at least a "Secret" version of it—might also be accessible to other cleared intelligence-community and law enforcement personnel, including JTTF members. No agency would prepare an unclassified version of the report that could be distributed to state and local entities or the private sector.

Note that the analyst identifies a "U.S. nexus," although the conversation does not say anything about the U.S. The analyst might know, however, that previous conversations between these two probable terrorists discussed operations in the U.S. Or, based on knowledge and expertise, the analyst might have concluded that they were most likely talking about the U.S.

The Pittsburgh police report and FBI memorandum

It appears from the documents in this vignette that the Pittsburgh Police Bureau Zone 6 officers have ongoing joint criminal investigations (non–terrorism related) with the FBI. Thus Zone 6 used this opportunity to convey the information about the suspicious activity, possibly terrorism-related, to the FBI. The agency's Pittsburgh office would send this information to the Pittsburgh JTTF, as well as to FBI headquarters. FBI headquarters might transfer the information to an IIR and provide it to the TTIC, although this probably would not happen until after an FBI field investigation has been completed. In any event, FBI personnel might alert the TTIC to the information by informal email. In either case, the information would be included in the TTIC analysis described below, assuming the TTIC could obtain the FBI's permission to disseminate. If FBI personnel believe dissemination would interfere with an ongoing investigation, the FBI might not give this permission.

In any case, Maryland license plates indicate a possible connection with another city (Baltimore or Washington, DC). It is possible, but far from certain, that the FBI, Pittsburgh JTTF, or Pittsburgh Police Department would pass this information to local law enforcement agencies in those cities, to the Maryland State Police, or to the Baltimore and Washington, DC, JTTFs.

TTIC analysis

The TTIC might prepare an analytical product that includes the NSA information and, assuming the TTIC received it, the information from the Pittsburgh FBI. This TTIC product would go to all of the same recipients as the NSA product, and would be placed on TTIC Online. But it would also be sent to the Pittsburgh FBI and from there to the Pittsburgh JTTF. This product would be classified "Secret" and would not go to state, local, or private sector entities.

The DHS has a mandate to provide information and warnings to the private sector. It does this for some industries that have been identified as the critical infrastructures, such as the communications sector and the airlines, but not across the board for industries that could be targets of terrorism. The DHS currently has no system for providing information such as the contents of the NSA report or the TTIC analysis, even if it were unclassified, to private sector theater- or mall-owners or to their security firms. To provide this warning, the DHS probably would work with state or local emergency staffs and might agree to have the FBI provide the information through its contacts with state and local law enforcement.

Additional sharing needed

This vignette demonstrates again that the most significant information roadblock is between the federal government and state, local, and private sector entities. The two pieces of information in this scenario—information from the NSA intercept and from the Pittsburgh police—would be available to be correlated and analyzed at the Pittsburgh JTTF, FBI headquarters, and probably the TTIC. The key issue would be a failure to produce a sanitized, unclassified report of the NSA information that could be conveyed to the state, local, and private entities.

Although the intelligence agencies have come a long way since September 11, in their recognition of the need to sanitize intelligence for use by a broader audience, they still don't see nonfederal entities as their consumers. That is, the intelligence agencies see their job as sending information up to the President and senior officials—not out to the entities that might serve as sensors to collect and contribute additional information. The federal agencies whose responsibility it is to communicate with these state, local, and private entities—the DHS and the FBI via the JTTFs—do not have the authority to declassify intelligence reports from the NSA or the TTIC. Therefore, the original classifiers must have the responsibility to produce an unclassified version of intelligence reporting at the same time that they produce the classified version. If the DHS or the JTTFs do not feel the unclassified version contains enough useful information, they should have the responsibility of going back to the originator and asking to have more details included in the declassified report.

Additional output needed in this vignette would include a version of the NSA report that is sanitized to the unclassified level. This unclassified version would not mention a source or that the information came from the NSA, but would retain more than merely a generic warning. It would read something like this:

> *Recently acquired information indicates a possible threat to malls, or possibly theaters. No specific time frame or location is indicated, but the threat did seem to imply that it would be soon.*

The DHS would convey the information from this report to private sector contacts with responsibility for security at malls, theaters, and other similar potential targets. To do this, the DHS would have to develop relationships, contacts, and reliable communication mechanisms with all relevant industries. The method of communication could be email (although email lists are hard to keep current) or some other method that pushes information to recipients. The JTTFs would also push the unclassified NSA information to state and local law enforcement agencies.

Also, information from the Pittsburgh police and FBI reports should find its way to the Maryland and Washington, DC, police because of the license-plate information that suggests a tie to those jurisdictions. The Pittsburgh police and the Pittsburgh JTTF should pass this information on to these local law enforcement entities.

Once the information from the NSA report and the Pittsburgh police is communicated to the state, local, and private sector recipients, the recipients will be sensitized to look for information relating to possible terrorist planning or activity at malls, theaters, and similar potential targets. This will inspire a second level of input to the federal government, most likely through the DHS and the JTTFs.

Vignette 3: Information-sharing between and within government agencies

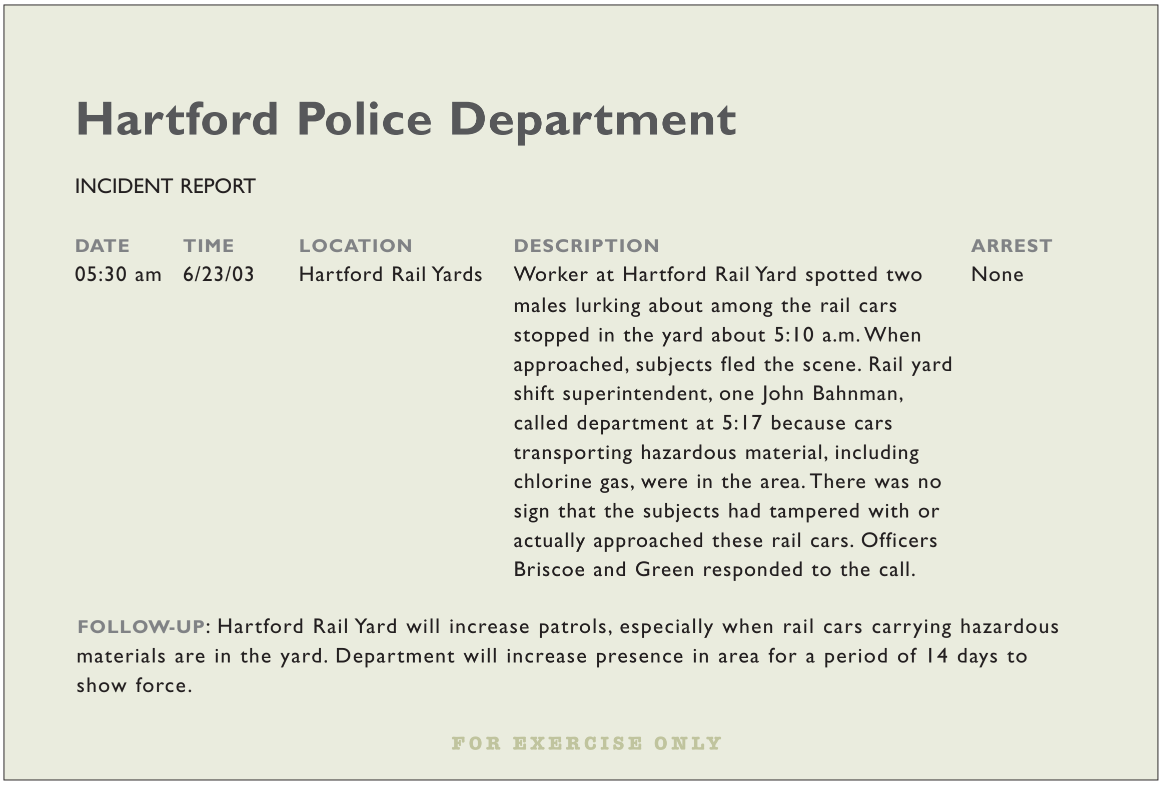
HAZARDOUS MATERIALS

On a police blotter in Hartford, CT, it says that police were called to the rail yards when a worker spotted several strangers lurking around a train. This train included tank cars carrying hazardous materials.

Meanwhile, security officials at a chemical plant in Convent, LA, that produces chlorine have noted the presence of intruders who appeared to be monitoring the loading of rail cars. The intruders ran away when they were approached.

According to a local newspaper, a zoo in Louisiana reports that several animals have died of apparent poisoning from chlorine gas. Zoo officials tell the press that a fire started in a shed where a large jug of chlorine had been placed. Zoo officials are perplexed because, although chlorine is used for cleaning out pens at the zoo, it is not ordinarily stored in the shed, which is used to store feed.

At the same time, a CIA source in Southeast Asia reports that several months ago he was present at a meeting of terrorists associated with Al Qaeda, at which the terrorists were discussing the long, unguarded rail lines and lightly monitored rail yards in the U.S. and speculating that it would be possible to use this vulnerability to stage an attack.



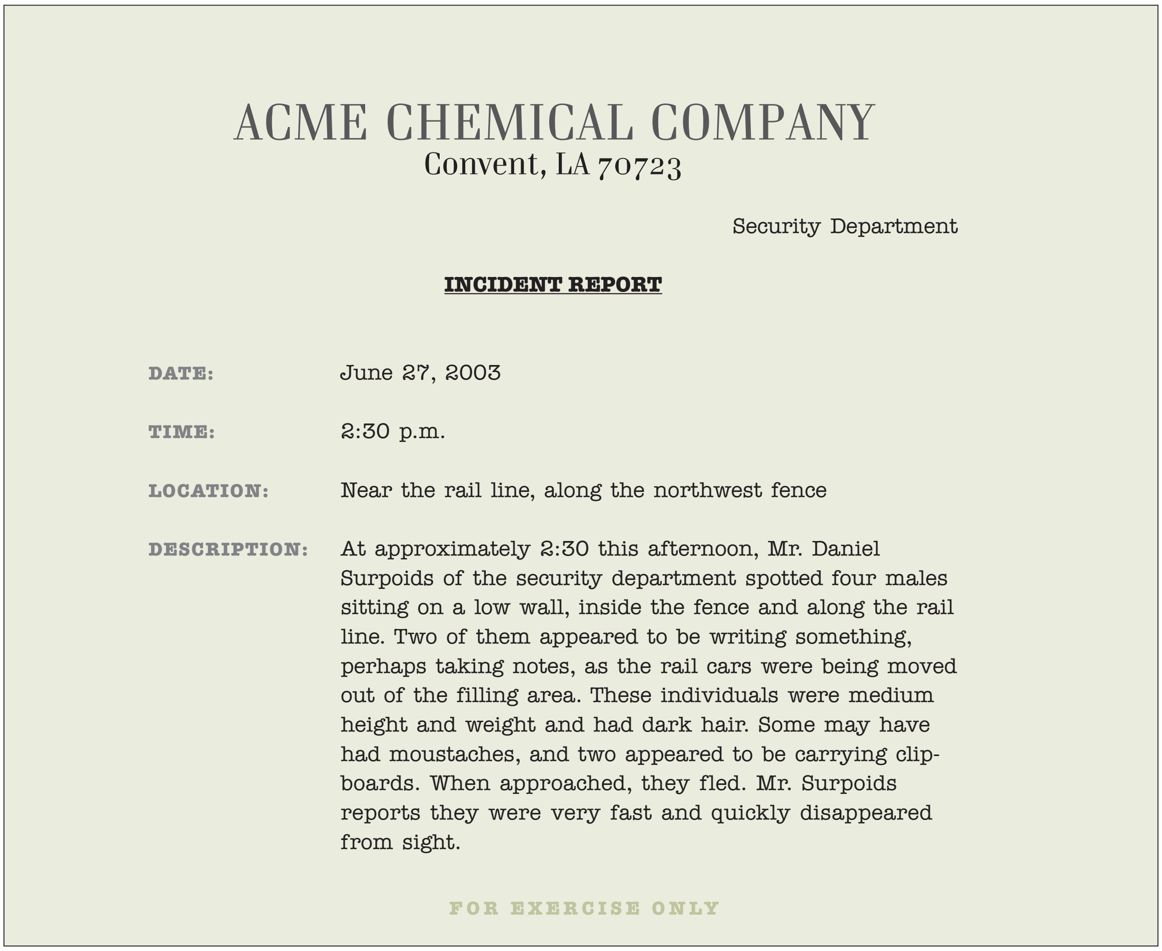
Hartford Police Department

INCIDENT REPORT

DATE	TIME	LOCATION	DESCRIPTION	ARREST
05:30 am	6/23/03	Hartford Rail Yards	Worker at Hartford Rail Yard spotted two males lurking about among the rail cars stopped in the yard about 5:10 a.m. When approached, subjects fled the scene. Rail yard shift superintendent, one John Bahnman, called department at 5:17 because cars transporting hazardous material, including chlorine gas, were in the area. There was no sign that the subjects had tampered with or actually approached these rail cars. Officers Briscoe and Green responded to the call.	None

FOLLOW-UP: Hartford Rail Yard will increase patrols, especially when rail cars carrying hazardous materials are in the yard. Department will increase presence in area for a period of 14 days to show force.

FOR EXERCISE ONLY



ACME CHEMICAL COMPANY

Convent, LA 70723

Security Department

INCIDENT REPORT

DATE: June 27, 2003

TIME: 2:30 p.m.

LOCATION: Near the rail line, along the northwest fence

DESCRIPTION: At approximately 2:30 this afternoon, Mr. Daniel Surpoids of the security department spotted four males sitting on a low wall, inside the fence and along the rail line. Two of them appeared to be writing something, perhaps taking notes, as the rail cars were being moved out of the filling area. These individuals were medium height and weight and had dark hair. Some may have had moustaches, and two appeared to be carrying clipboards. When approached, they fled. Mr. Surpoids reports they were very fast and quickly disappeared from sight.

FOR EXERCISE ONLY

The Jefferson Courier

ALL THE NEWS THAT FITS, WE PRINT

June 30, 2003

Several Monkeys at the Zoo Succumb to Chlorine Fumes

(JEFFERSON) Five monkeys, comprising the zoo's entire collection of capuchin and howler primates, died Sunday night under mysterious circumstances. The monkeys apparently succumbed to chlorine gas, which was emitted from several jugs of liquid chlorine when the shed in which the chlorine was stored caught fire.

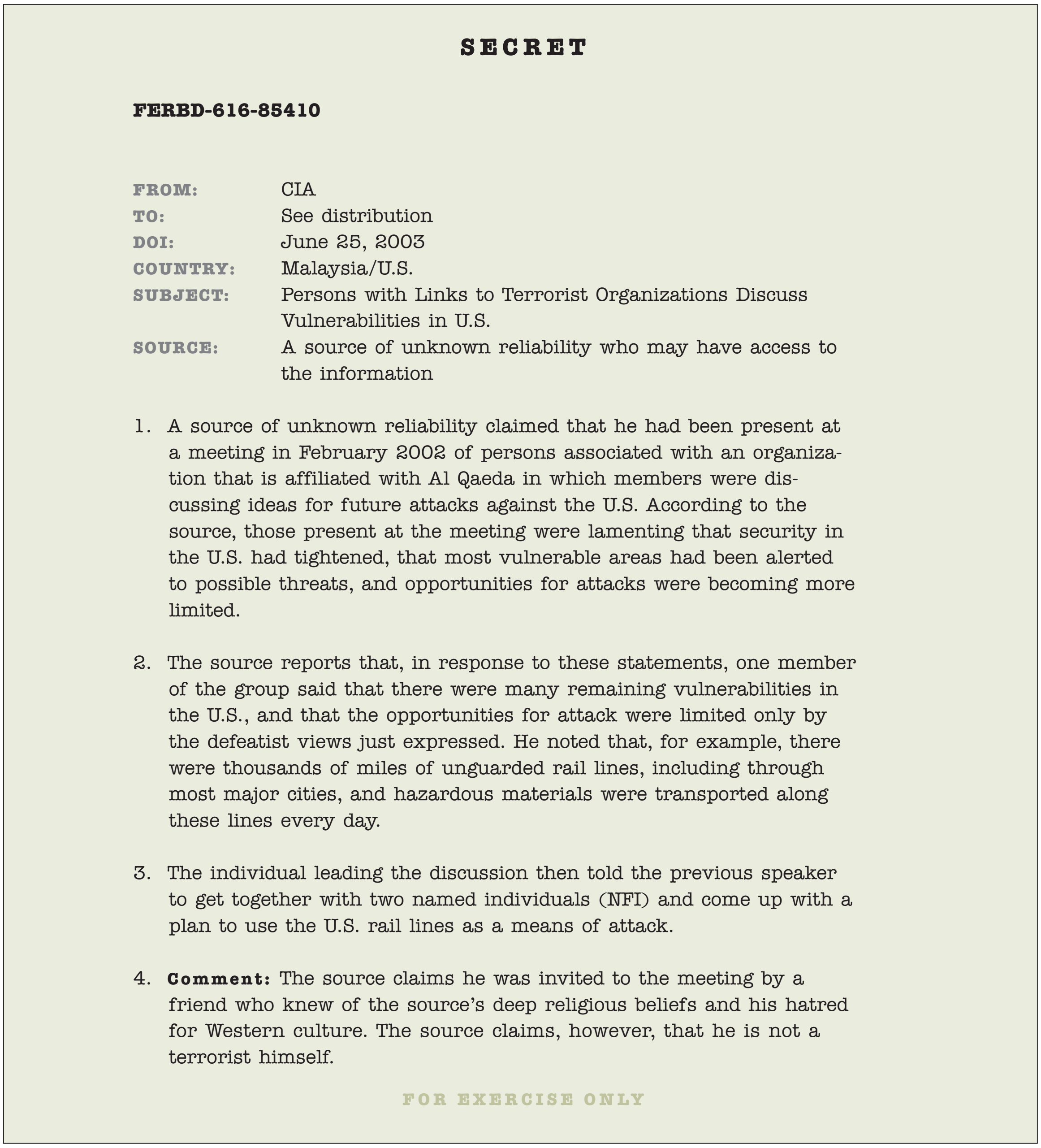
Zoo authorities are conducting an internal investigation to determine how several bottles of chlorine bleach, used to clean the animal cages at the zoo, ended up in a shed adjacent to the primate area. The shed is used to store feed for the animals.

"This is just awful. I can't understand how chlorine could even be in that shed," said Paul Le Singe. "We do use chlorine to clean out the animal cages, but it is stored really far away. All the cleaning products are."

The shed apparently burned itself out during the night. When they arrived in the morning, zoo authorities called the fire department. The firefighters who responded found the burned chlorine bottles. Toxicology analysis, which is expected to confirm that the monkeys died of chlorine gas poisoning, is expected to be completed in a few days.

"This is just awful. I can't understand how chlorine could even be in that shed."

Another question that remains unanswered is why the night watchman, Mr. John Leon, did not notice the fire.



SECRET

FERBD-616-85410

FROM: CIA
TO: See distribution
DOI: June 25, 2003
COUNTRY: Malaysia/U.S.
SUBJECT: Persons with Links to Terrorist Organizations Discuss Vulnerabilities in U.S.
SOURCE: A source of unknown reliability who may have access to the information

1. A source of unknown reliability claimed that he had been present at a meeting in February 2002 of persons associated with an organization that is affiliated with Al Qaeda in which members were discussing ideas for future attacks against the U.S. According to the source, those present at the meeting were lamenting that security in the U.S. had tightened, that most vulnerable areas had been alerted to possible threats, and opportunities for attacks were becoming more limited.

2. The source reports that, in response to these statements, one member of the group said that there were many remaining vulnerabilities in the U.S., and that the opportunities for attack were limited only by the defeatist views just expressed. He noted that, for example, there were thousands of miles of unguarded rail lines, including through most major cities, and hazardous materials were transported along these lines every day.

3. The individual leading the discussion then told the previous speaker to get together with two named individuals (NFI) and come up with a plan to use the U.S. rail lines as a means of attack.

4. **Comment:** The source claims he was invited to the meeting by a friend who knew of the source's deep religious beliefs and his hatred for Western culture. The source claims, however, that he is not a terrorist himself.

FOR EXERCISE ONLY

How the information would likely be handled today

The Hartford Rail Yard and Hartford Police Department reports

The telephone report from a worker at the Hartford Rail Yard might be reported by the rail yard to the railway industry's Information Sharing and Analysis Center (ISAC). ISACs are industry task forces that collect, analyze, and disseminate information about industry threats and vulnerabilities. The railway industry has an active and effective ISAC, which means the ISAC would likely distribute this report to its members and might also report the information to the Infrastructure Protection Directorate at the DHS. Whether the information would then be disseminated to other parts of the DHS, such as the Information Assurance Directorate, or the TTIC, is less clear. The Hartford Police Department report might be stored digitally, but it would not necessarily be easily retrievable and it is unclear how long it would be retained.

The Acme Chemical Company incident report

The Acme Chemical Company incident would be included in a daily report of incidents sent to the plant manager. The plant manager would likely direct the security department to watch aggressively for more such activity. The incident would not be reported to the Convent, LA, sheriff's office unless there was a repeat incident. In our scenario, this incident report probably is not saved digitally.

The Jefferson Zoo incident

Most likely, there would be no written report of the incident at the Jefferson Zoo other than the newspaper account. The zoo would not report the incident to the Jefferson, LA, police unless zoo officials uncovered something suspicious during their internal investigation. Someone in the Jefferson, LA, police department might notice and remember the newspaper article. In the best case, someone at the Acme Chemical Plant (not far from Jefferson, LA) who also knew about the monitoring of the rail cars, would notice the newspaper article and alert the local police or the FBI.

The CIA report

In our scenario, this a routine human-source (HUMINT) report from a source in Malaysia. This report does not contain source information so sensitive that the report would require "paper only" distribution. Instead, this report would likely be disseminated electronically to the intelligence community, including the TTIC, the FBI, and the White House Situation Room. It would be placed on TTIC Online and would be available to be pulled by cleared DHS and JTTF personnel who search that system. No information about this report would go to state or local law enforcement or to chemical companies or railroads. Those entities would be aware, generally, of warnings of risks to their industries from terrorists.

Additional sharing needed

This vignette illustrates the difficulty of separating signal from noise with information on possible terrorist activity. The initial report from Acme Chemical, and the incident at the Jefferson Zoo would, in isolation, be considered noise by those receiving the reports. Therefore, the information would not make its way into the network unless some additional information came in that highlighted its significance. The report from the Hartford Rail Yard might make its way, through the ISAC, to other rail yards and the DHS, but it is more likely that it would be seen as insignificant. The CIA report could be the additional information that would highlight the significance of these incidents, but it would have to get out to the sensors who, in turn, would be triggered to share their input with the federal government.

To get additional output, the CIA would have to create a sanitized, unclassified version of the CIA report at the time it was first prepared. This unclassified report could be disseminated to state and local entities and to the private sector. The DHS could take this information and reach out to task forces such as ISACs, in the chemical and railroad industries. These task forces would then use established communication mechanisms to get the information out to individual companies in their industries. The JTTFs, in turn, could push the sanitized information out to state and local law enforcement agencies.

Even so, when the state, local, and private sector entities received this output, they would not uncover the three incidents described in our scenario unless they went back to past records or happened to recall the incidents. Therefore, the additional output from the federal government would have to do more than provide information. To be more effective at triggering necessary responses from local and private sector entities, the output would have to include a request that these entities search for information about the specific threats mentioned in the CIA report. Some entity—probably the FBI, the DHS, or the TTIC—would have to initiate this request for additional input.

Once the state, local, and private sector entities received the sanitized CIA information and the request for input, they would be far more likely to recognize the significance of the Hartford Rail Yard, Acme Chemical Company, and Jefferson Zoo incidents—and to provide this additional input to the network. The most significant challenge at that point would be that to retrieve information about these and similar incidents, without relying solely on memories, the companies and police departments would have to search their internal records. Because of this, such records, particularly those of the law enforcement agencies, must be maintained digitally, in a manner that can be searched, and subject to some retention requirement.

The information generated by state, local, and private entities should make its way back to the TTIC, but the TTIC should not be the only place where this information is correlated, assessed, and analyzed. To be most effective, the system should also encourage communication among regional entities, within industries, at the local level, and in other decentralized ways, including among centers of expertise in government, industry, and academia. In this scenario, communication among local police departments in Louisiana about suspicious activity involving hazardous chemicals, or among JTTFs in Hartford and New Orleans about activity in rail yards where chemicals are present, could trigger additional questions, investigation, and analysis that would lead to even more information in the system.

Vignette 4: Information-sharing between the private sector and government agencies and between and within government agencies

SKYDIVERS AND MALLS

The NSA issued a report in late June that sensitive intercepted communications among known Al Qaeda leaders abroad indicate that final preparations are being made for terrorist operations against targets in the U.S. Speakers have mentioned "malls," or perhaps "The Mall," and have referred to "the other city." In one conversation they also mentioned "movie theaters."

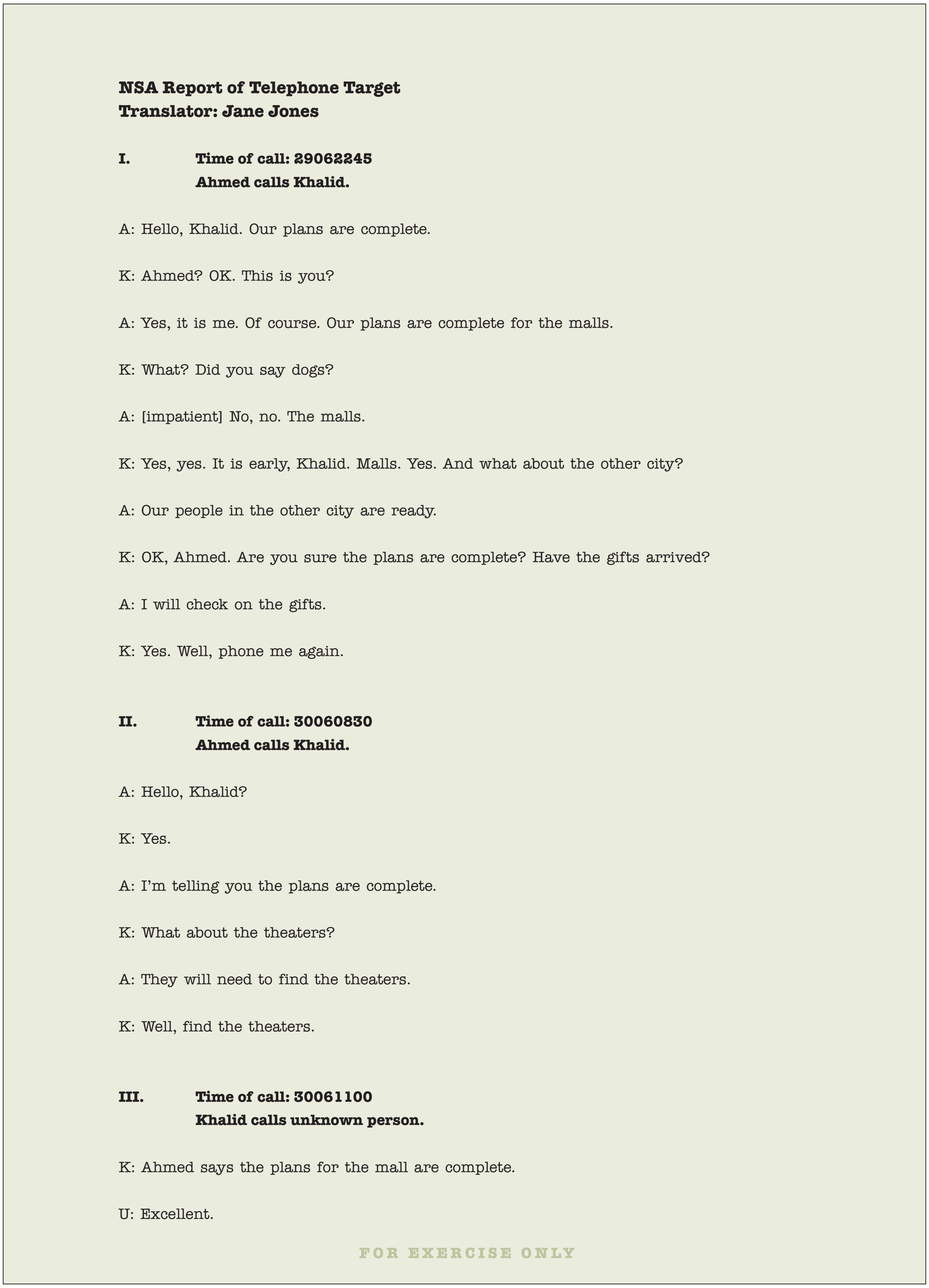
Earlier, the FBI's Chicago field office picked up some information from an informant claiming that terrorist cells in the U.S. were discussing various methods for attacks, including general aviation, scuba divers, crop dusters, and skydivers. The Urgent Report from the Chicago field office to FBI headquarters, dated March 30, 2003, indicates that the SAC thinks this is pretty low-level intelligence but is "leaning forward" on reporting.

In early August, the NSA picked up a communication in which a presumed Al Qaeda figure mentioned skydivers. The speaker has been identified, and it is known that he has visited Texas twice.

Now, five individuals with names of apparent Middle Eastern origin/ethnicity have enrolled in skydiving classes in five divergent areas of the country (Texas, Pennsylvania, Rhode Island, Illinois, and Florida). All have used student identification from nearby universities.

Interest is converging on Texas, however, where one of the skydivers is asking to rent a Cessna 182 (commonly used by skydivers). Another individual, possibly with a similar ethnic origin, is trying to rent another Cessna 182 at another airfield in Texas. Both individuals want to rent the planes during Thanksgiving weekend—a big shopping weekend, and therefore a possible "mall" connection.

The skydiver in Texas is also showing an interest in explosives. He has visited a relevant website and has ordered a how-to book, using his VISA card.



NSA Report of Telephone Target
Translator: Jane Jones

I. Time of call: 29062245
Ahmed calls Khalid.

A: Hello, Khalid. Our plans are complete.

K: Ahmed? OK. This is you?

A: Yes, it is me. Of course. Our plans are complete for the malls.

K: What? Did you say dogs?

A: [impatient] No, no. The malls.

K: Yes, yes. It is early, Khalid. Malls. Yes. And what about the other city?

A: Our people in the other city are ready.

K: OK, Ahmed. Are you sure the plans are complete? Have the gifts arrived?

A: I will check on the gifts.

K: Yes. Well, phone me again.

II. Time of call: 30060830
Ahmed calls Khalid.

A: Hello, Khalid?

K: Yes.

A: I'm telling you the plans are complete.

K: What about the theaters?

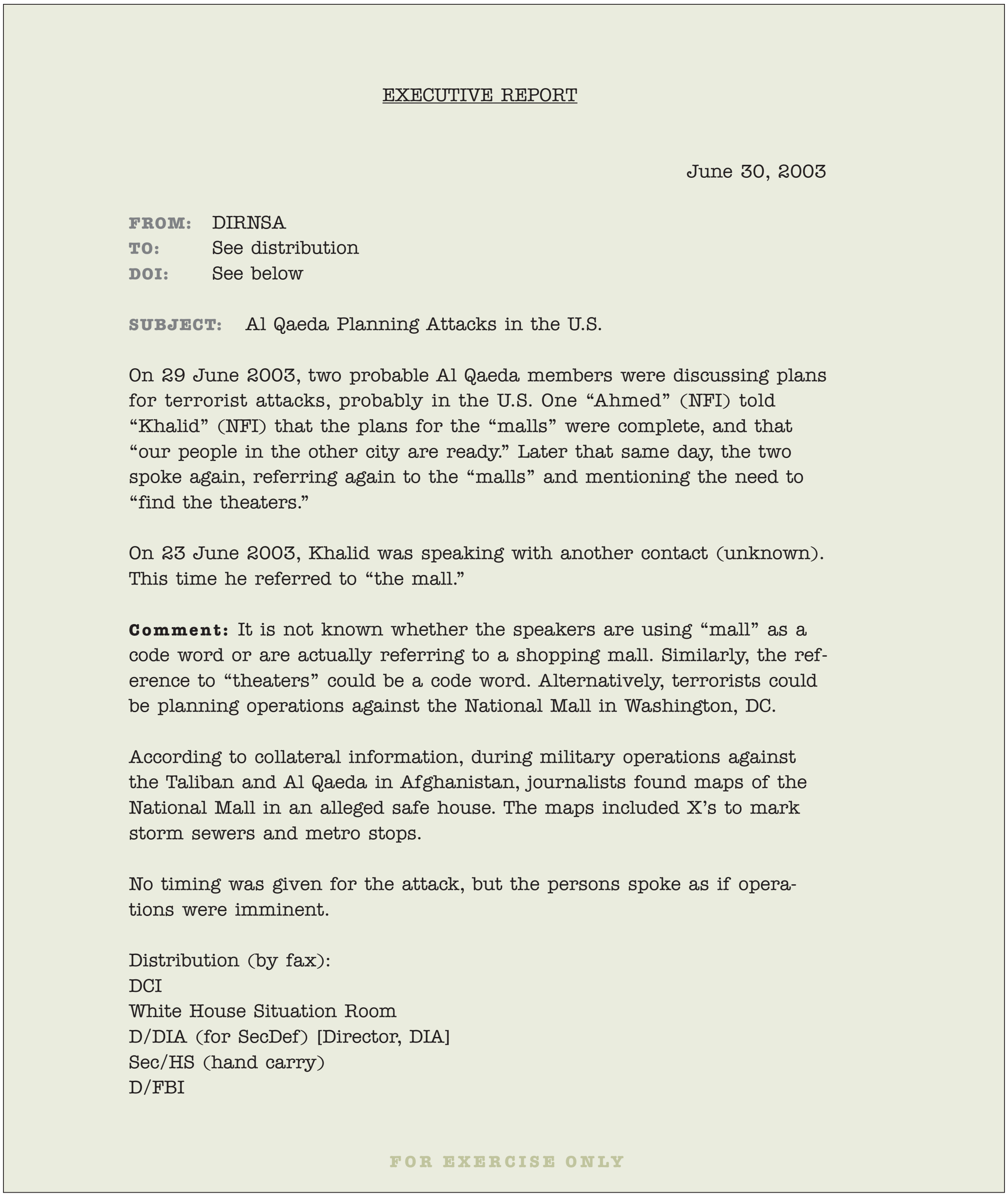
A: They will need to find the theaters.

K: Well, find the theaters.

III. Time of call: 30061100
Khalid calls unknown person.

K: Ahmed says the plans for the mall are complete.

U: Excellent.



EXECUTIVE REPORT

June 30, 2003

FROM: DIRNSA
TO: See distribution
DOI: See below

SUBJECT: Al Qaeda Planning Attacks in the U.S.

On 29 June 2003, two probable Al Qaeda members were discussing plans for terrorist attacks, probably in the U.S. One “Ahmed” (NFI) told “Khalid” (NFI) that the plans for the “malls” were complete, and that “our people in the other city are ready.” Later that same day, the two spoke again, referring again to the “malls” and mentioning the need to “find the theaters.”

On 23 June 2003, Khalid was speaking with another contact (unknown). This time he referred to “the mall.”

Comment: It is not known whether the speakers are using “mall” as a code word or are actually referring to a shopping mall. Similarly, the reference to “theaters” could be a code word. Alternatively, terrorists could be planning operations against the National Mall in Washington, DC.

According to collateral information, during military operations against the Taliban and Al Qaeda in Afghanistan, journalists found maps of the National Mall in an alleged safe house. The maps included X’s to mark storm sewers and metro stops.

No timing was given for the attack, but the persons spoke as if operations were imminent.

Distribution (by fax):
DCI
White House Situation Room
D/DIA (for SecDef) [Director, DIA]
Sec/HS (hand carry)
D/FBI

FOR EXERCISE ONLY

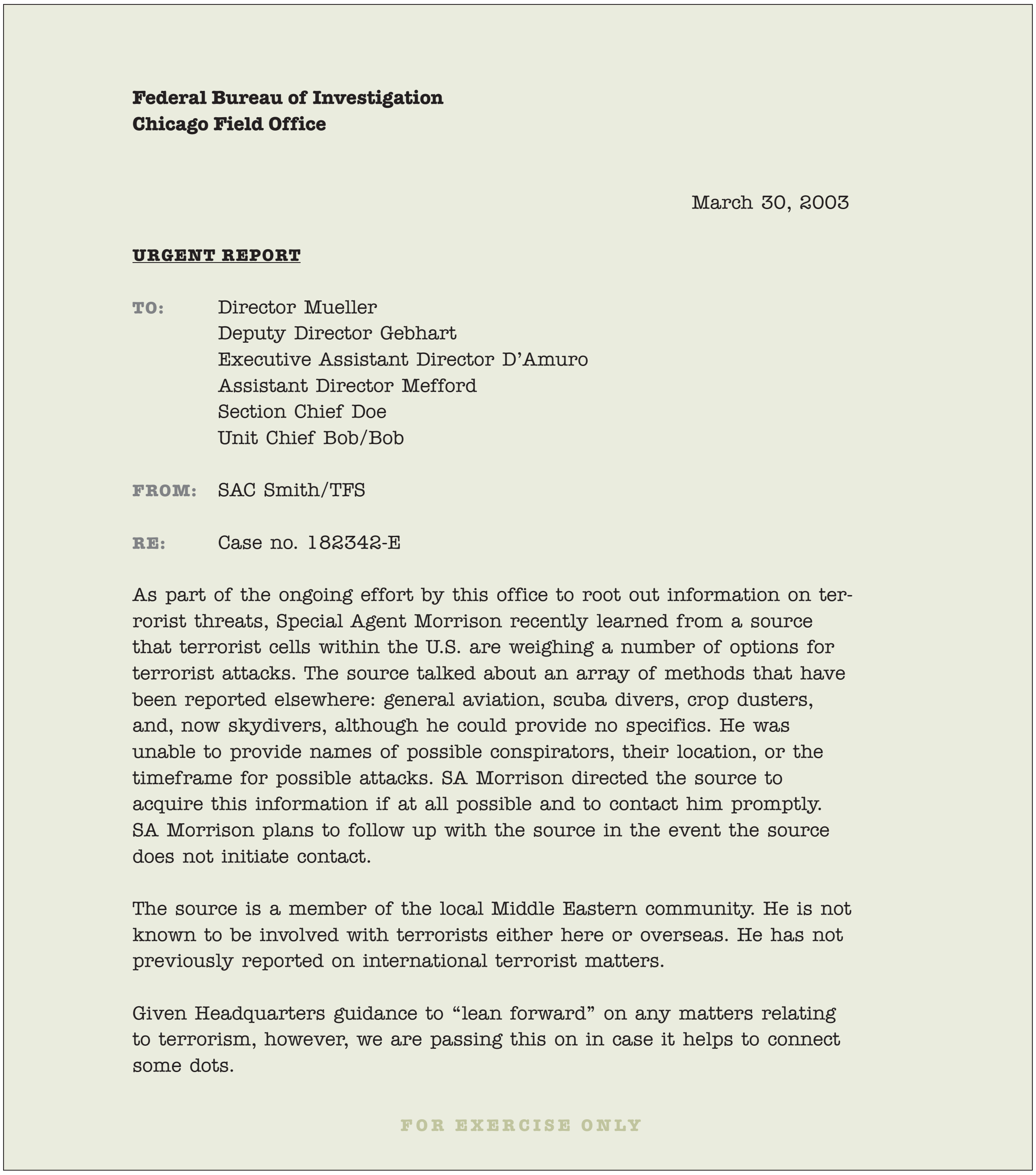
FROM: DIRNSA
TO: 06292245Z

SUBJECT: Terrorists Discuss Plans

On 29 June 2003, two probable Al Qaeda members were discussing plans for terrorist attacks, probably in the U.S. In the course of the conversation, the two referred to “malls” and “theaters.” In another conversation, one of them referred to “the mall.”

COMMENT: The probable terrorists could be using “malls” and “theaters” as code words. Alternatively, they could be referring to the National Mall in Washington, DC. According to collateral information, during military operations against the Taliban in Afghanistan, journalists found maps of the National Mall in an alleged safe house.

FOR EXERCISE ONLY



Federal Bureau of Investigation
Chicago Field Office

March 30, 2003

URGENT REPORT

TO: Director Mueller
Deputy Director Gebhart
Executive Assistant Director D'Amuro
Assistant Director Mefford
Section Chief Doe
Unit Chief Bob/Bob

FROM: SAC Smith/TFS

RE: Case no. 182342-E

As part of the ongoing effort by this office to root out information on terrorist threats, Special Agent Morrison recently learned from a source that terrorist cells within the U.S. are weighing a number of options for terrorist attacks. The source talked about an array of methods that have been reported elsewhere: general aviation, scuba divers, crop dusters, and, now skydivers, although he could provide no specifics. He was unable to provide names of possible conspirators, their location, or the timeframe for possible attacks. SA Morrison directed the source to acquire this information if at all possible and to contact him promptly. SA Morrison plans to follow up with the source in the event the source does not initiate contact.

The source is a member of the local Middle Eastern community. He is not known to be involved with terrorists either here or overseas. He has not previously reported on international terrorist matters.

Given Headquarters guidance to "lean forward" on any matters relating to terrorism, however, we are passing this on in case it helps to connect some dots.

FOR EXERCISE ONLY

FROM: DIRNSA
TO: 081545Z

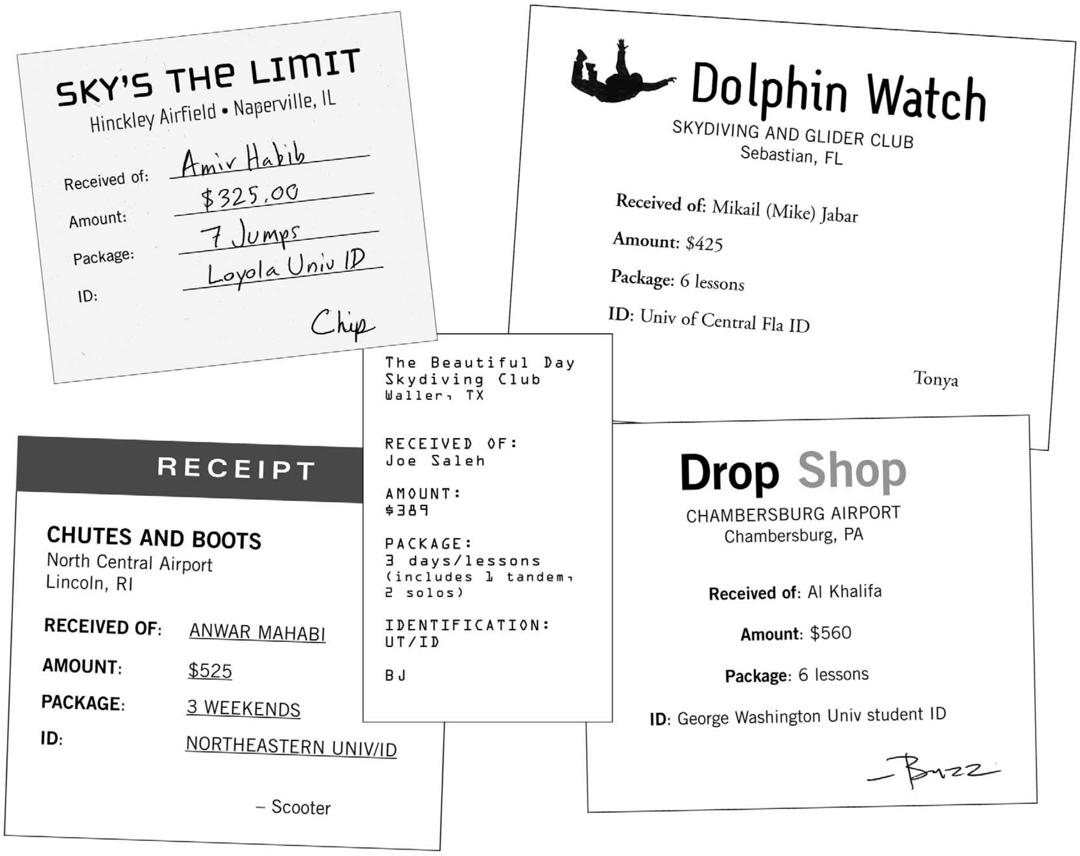
SUBJECT: Terrorist Plans

On 7 August 2003, a probable Al Qaeda figure, Ahmet Hafs, in a conversation with an unknown contact, mentioned some apparent plans for "skydivers."

COMMENT: This is the first time the intelligence community has noted mention of skydivers by a known terrorist operative.

8/10/03
Chief: Ran the traps on "Ahmet Hafs." He got a visa to visit the U.S. in 1999 and again in mid-2001. He visited Texas on both trips. – CB

FOR EXERCISE ONLY



SKY'S THE LIMIT
Hinckley Airfield • Naperville, IL

Received of: Amir Habib
Amount: $325.00
Package: 7 Jumps
ID: Loyola Univ ID

Chip

Dolphin Watch
SKYDIVING AND GLIDER CLUB
Sebastian, FL

Received of: Mikail (Mike) Jabar
Amount: $425
Package: 6 lessons
ID: Univ of Central Fla ID

Tonya

```
The Beautiful Day
Skydiving Club
Waller, TX

RECEIVED OF:
Joe Saleh

AMOUNT:
$389

PACKAGE:
3 days/lessons
(includes 1 tandem,
2 solos)

IDENTIFICATION:
UT/ID

BJ
```

RECEIPT

CHUTES AND BOOTS
North Central Airport
Lincoln, RI

RECEIVED OF: ANWAR MAHABI
AMOUNT: $525
PACKAGE: 3 WEEKENDS
ID: NORTHEASTERN UNIV/ID

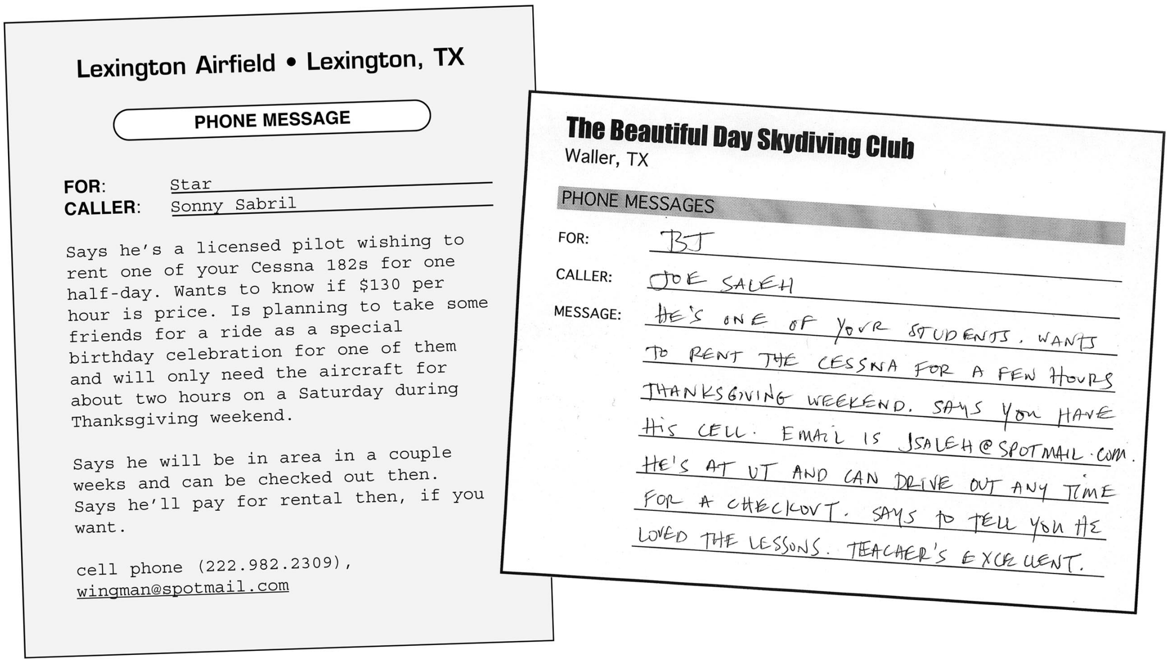
– Scooter

Drop Shop
CHAMBERSBURG AIRPORT
Chambersburg, PA

Received of: Al Khalifa
Amount: $560
Package: 6 lessons
ID: George Washington Univ student ID

–Buzz

FOR EXERCISE ONLY



Lexington Airfield • Lexington, TX

PHONE MESSAGE

FOR: Star
CALLER: Sonny Sabril

Says he's a licensed pilot wishing to rent one of your Cessna 182s for one half-day. Wants to know if $130 per hour is price. Is planning to take some friends for a ride as a special birthday celebration for one of them and will only need the aircraft for about two hours on a Saturday during Thanksgiving weekend.

Says he will be in area in a couple weeks and can be checked out then. Says he'll pay for rental then, if you want.

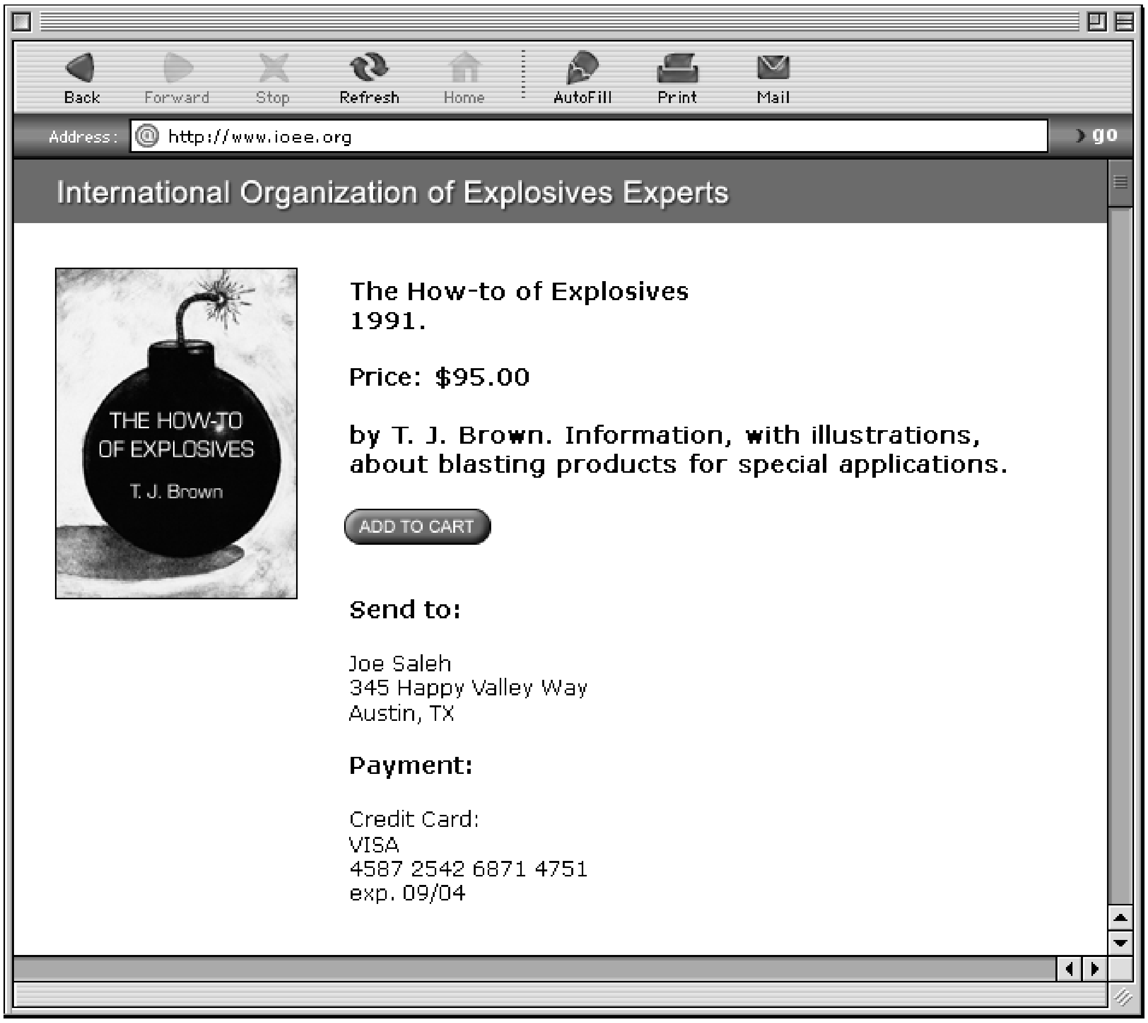
cell phone (222.982.2309),
wingman@spotmail.com

The Beautiful Day Skydiving Club
Waller, TX

PHONE MESSAGES

FOR: BJ
CALLER: JOE SALEH
MESSAGE: HE'S ONE OF YOUR STUDENTS. WANTS TO RENT THE CESSNA FOR A FEW HOURS THANKSGIVING WEEKEND. SAYS YOU HAVE HIS CELL. EMAIL IS JSALEH@SPOTMAIL.COM. HE'S AT UT AND CAN DRIVE OUT ANY TIME FOR A CHECKOUT. SAYS TO TELL YOU HE LOVED THE LESSONS. TEACHER'S EXCELLENT.

FOR EXERCISE ONLY



International Organization of Explosives Experts

The How-to of Explosives
1991.

Price: $95.00

by T. J. Brown. Information, with illustrations, about blasting products for special applications.

Send to:

Joe Saleh
345 Happy Valley Way
Austin, TX

Payment:

Credit Card:
VISA
4587 2542 6871 4751
exp. 09/04

FOR EXERCISE ONLY

Information-sharing: how this information would likely be handled today

The NSA transcript and report on malls

The transcript of the intercepted conversation related to malls would go to an NSA analyst, who would produce a report. The full report would be classified “Top Secret,” but a second “Secret” version could also be prepared. The fact of the NSA’s access to the phones of at least one of these individuals would be considered extremely sensitive. The “Top Secret” report would go to the DHS, the TTIC, the FBI, the CIA, the DoD, and the White House, in paper form. A “Secret” version of it, at least, might be accessible electronically, via TTIC Online, to other cleared intelligence community and law enforcement personnel, including at JTTFs nationwide. No agency would prepare an unclassified version of the report that could be distributed to state and local entities or the private sector.

Note that the analyst in our vignette identified “U.S. nexus,” although the conversation did not say anything about the U.S. The analyst might have known, however, that previous conversations between these two probable terrorists included discussion of operations in the U.S. Or the analyst might have concluded based on knowledge and expertise that they most likely were talking about the U.S.

FBI report on skydiving

The FBI report about possible terrorist methods, including skydiving, would be provided to the Chicago JTTF and to FBI headquarters in Washington, DC. Because the source would be considered untested, it most likely would not be turned into an IIR, although the information it contains might be conveyed informally to TTIC personnel. The information would not find its way to JTTFs around the country or to state and local law enforcement or private sector entities, such as skydiving clubs.

The NSA report on skydiving

The NSA intercept about terrorists’ interest in skydiving would be distributed to officials at the White House, the TTIC, the DHS, the FBI, the CIA, and the DoD, at least in paper form. It might be made available electronically via TTIC Online to all federal intelligence community and law enforcement organizations. Because no unclassified version would be prepared, the report’s contents would not be available to state or local law enforcement or to private sector entities. Because the NSA report has a specific connection to the Texas area, FBI field offices and JTTFs in Texas would be notified. The TTIC or the FBI may request that the NSA further sanitize the report to be handled as unclassified law-enforcement-sensitive information for distribution to state and local law enforcement authorities, possibly through the National Law Enforcement Telecommunications System (NLETS).

TTIC and other intelligence analysis

Reference to the two NSA reports on malls and skydiving would be included in the Daily Threat Matrix and perhaps in TTIC analytical products produced by other agencies, such as the Transportation Security Administration (TSA) or the Bureau of Immigration and Customs Enforcement (BICE), for their leadership. These products would go to all of the same recipients as the NSA product and would be placed on TTIC Online. These products would be classified and would not go to state, local, or private sector entities.

The DHS has a mandate to provide information and warnings to the private sector. It does this for some industries that have been identified as the critical infrastructures, such as the communications sector and the airlines, but not across the board for industries that could be targets of terrorism. The DHS currently has no system for providing such information as the contents of NSA reports or TTIC analysis, even if it were unclassified, to private sector theater or mall owners or their security firms, or to skydiving clubs. To provide this warning, the DHS would probably work through state or local emergency staffs and might agree to have the FBI provide the information through its contacts with state and local law enforcement.

Additional sharing needed

This vignette demonstrates that the most significant information roadblock is between the federal government and state, local, and private sector entities. The pieces of information in this vignette—information from the NSA intercepts and the Chicago FBI—would be available to be correlated and analyzed at the FBI headquarters, and probably the TTIC. Here, the key issue is the absence of a rapid, effective process to produce a sanitized, unclassified report of the NSA information that could be conveyed to the state, local, and private entities, particularly mall and theater owners and their security firms, but also to skydiving clubs.

Although the intelligence agencies have come a long way since September 11 in their recognition of the need to sanitize intelligence for use by a broader audience, they still do not see nonfederal entities as their consumers. That is, the intelligence agencies see their primary job as sending information up to the President and senior officials, not out to the entities that might serve as sensors who collect and contribute additional information or who need to be prepared to prevent or respond to terrorist action. The federal agencies whose responsibility it is to communicate with these state, local, and private entities—the DHS and the FBI via the JTTFs—presently do not have the authority to declassify intelligence reports from the NSA or the TTIC. Therefore, the original classifiers must have the responsibility to produce an unclassified version of intelligence reporting at the same time that they produce the classified version. If the DHS or the JTTFs do not feel that the unclassified version contains enough useful information, they should have the responsibility to go back to the originator and ask that more detail be declassified.

Additional output needed in this vignette would include a version of the NSA reports that are sanitized to the unclassified level. The unclassified versions would not mention a source or that the information came from the NSA, but they would retain more than merely a generic warning.

The DHS would convey the information from these reports to private sector contacts with responsibility for security at malls, theaters, and other similar potential targets and to skydiving clubs. To do this, the DHS would have to develop relationships, contacts, and reliable communication mechanisms with all relevant industries. The method of communication could be email (although email lists are hard to keep current) or some other method that pushes information to recipients. The JTTFs would also push the unclassified NSA information to state and local law enforcement agencies. The information should also be available to be pulled by analysts throughout the network who have the necessary permissions or authorities, in case those analysts do not realize the relevance of the information to their work until a later date, when additional information comes in.

Once the information from the NSA reports and the Chicago FBI report was communicated to the state, local, and private sector recipients, they would be sensitized to look for information relating to possible terrorist planning or activity at malls, theaters, and similar potential targets. In addition, skydiving clubs would be alert for suspicious behavior. This would inspire a second level of input to the federal government, most likely through the DHS and the JTTFs.

Use of private data: how the government would obtain and use the information today

The records of skydiving lessons

The FBI and NSA reports indicating skydiving as a possible method of terrorist attack may have prompted alert and aggressive FBI field offices to inquire about people who had taken skydiving lessons or otherwise shown an interest in skydiving. To inquire about this, the officers would contact skydiving clubs in their areas. But without more specific search parameters, the numbers of students or inquirers would have been too high to permit follow-up on all of the names. FBI field-office agents would likely attempt to reduce the number to a manageable volume, perhaps by first looking at recent training records of immigrants from select countries of concern or of people with Arab-sounding names.[3] They might also investigate students who were deemed "suspicious" in a report from the skydiving instructor or club owner.

The NSA intercept concerning a suspected terrorist who had traveled to Texas probably would have caused Texas FBI field offices to conduct a more thorough investigation. In addition to questioning personnel in local skydiving clubs about suspicious activity, they would likely have asked for lists of people who had taken lessons in the past few months or year. Skydiving clubs, for the most part, do not keep records in a searchable form. (There might be digital records of people who have skydiving certification, but such records would not be the only relevant documents—terrorists probably would not see a need to become certified to carry out their plans.) Therefore, the FBI personnel would have to prepare their own lists of names based on conversations with skydiving-club personnel. The lists would be long, and the challenge for the FBI at that point would be to reduce them to a manageable number of people who could be investigated further.

One first step would be to compare the names with government databases to obtain more information. However, local field offices are not connected directly to most relevant databases. A field agent would probably need to submit the list to others at FBI headquarters to conduct the searches. The field office would probably prioritize its submissions to FBI headquarters based on information from the skydiving-club personnel about suspicious activity, apparent Arab origin of names, and other factors that lead them to have some level of concern. Prioritization would also be based on access by the field office to FBI-wide case and watch list databases. Besides submitting the names to headquarters, the field-office agents might—with the requisite legal basis—open preliminary investigations (or threat assessments, a more preliminary step under the new Attorney General's Guidelines on National Security Investigations and Foreign Intelligence Collection, effective October 31, 2003) on the higher-priority individuals and reach out for assistance through their local JTTF.

At FBI headquarters, the names would be checked against other local online database and hard-copy records at the FBI as well as with other federal agencies. One new step would be to check the names with the BICE at the DHS to determine whether any of the people were not U.S. citizens, had outstayed their visas, or had entered the country recently. The BICE has a variety of data-bases with information on immigrants and visitors to the U.S. These include the Student and Exchange Visitor Information System (SEVIS), which manages and maintains data about foreign students and exchange visitors; the National Security Entry-Exit Registration System (NSEERS), which contains detailed registration information about foreign visitors of "elevated national security risk"—primarily nationals of certain "high-risk" countries; and the United States Visitor and Immigration Status Indication Technology (US VISIT) system, a new system that will manage data—including biometric identifiers and entry, exit, and status information—on all visitors to the U.S. The field agent would probably submit a query to the BICE's Law Enforcement Support Center (LESC), a national enforcement-operations center located in Vermont. The LESC gathers information from eight DHS databases, including SEVIS, NSEERS, US VISIT, and other former INS, Customs Service, or Federal Protective Service databases.

From searches of the watch list and BICE records, discussions with skydiving-club personnel, and other initial investigation, the FBI might conclude that some smaller number of names—even as many as 100—merited additional investigation. At that point, the investigators could search aggregated public-records data—from a commercial data aggregator—to determine whether there were other reasons for suspicion. For example, searches of these records could reveal false identities, or show that someone lived with or associated with people on the watch lists, or with others on the lists of skydivers. Any of these could be a reason to keep a person on the list. If that search narrowed the suspicious names down to a very small number, link analysis could be done on those names to determine their association with others around the country. That would give FBI personnel in other jurisdictions something more concrete to check against lists of skydivers in their areas.

[3] Despite concerns about ethnic profiling, it is highly likely that Arab-sounding names—and other indicators of possible Middle Eastern background—are taken into account by agents trying to make gross determinations about what individuals in a large pool warrant further scrutiny—at least when that scrutiny does not involve intrusive investigative measures that would require third-party approval, such as a court order.

Assuming the FBI field office in the Houston area took the first step of going to local skydiving clubs, it would have found the record, among others, for Joe Saleh. It is possible that the FBI would consider Mr. Saleh's request to rent a plane, in addition to his apparent Arab ethnicity, to be enough to warrant searching his name against the watch list and BICE records. Mr. Saleh's record also indicated that he had University of Texas student identification. Therefore, the LESC search, which includes the SEVIS database of foreign-student records, might have produced some information on Mr. Saleh. If these avenues revealed something suspicious, the FBI would do additional checks or surveillance on Mr. Saleh.

The requests for airplane rentals

If the FBI interviewed "BJ" from the Beautiful Day Skydiving Club, it would find that Mr. Saleh was attempting to rent a plane on the Saturday after Thanksgiving. This might cause the FBI to focus more attention on Mr. Saleh. They might also investigate other local companies that rent small airplanes to determine whether there were other plans for small-plane rentals at the same time that were suspicious. This could have led them to Mr. Sabril.

The credit card records

Assuming previous investigation turned up enough information to suspect Mr. Saleh, they might have obtained his credit card records, by subpoena or National Security Letter. Those records could have led the investigators to the purchase from the International Organization of Explosives Experts website of a book on explosives.

The email records and online activity

If the FBI found Mr. Saleh's letter requesting the airplane rental, it would have had an email address, which could have led it to an ISP. Still, with a generic email address, the ISP that Mr. Saleh was using would have been difficult to identify. The FBI may have been able to begin with the university server. Assuming the investigation to date had revealed sufficient suspicious behavior, the FBI could obtain a court order for email transaction records and records of online activity. These records would reveal Mr. Saleh's visits to websites about explosives.

How the government could use this information more effectively

One of the greatest challenges in using private data to assist in an investigation is narrowing the search to something that can produce a meaningful result. Traditional investigation techniques—making phone calls, asking questions—will always be critical to this process. These will turn up the facts that give investigators some information with which to query the private databases. The more data an investigator can access to do this narrowing, the more accurate the narrowing is likely to be and the less reliant on hunches, stereotypes, and ethnic profiling. Some steps to improve the ability to search government and private sector databases would assist in this narrowing process.

First, it would increase the ease with which field agents could conduct searches of government databases if there were real-time interfaces between those offices and key systems like those of the TSC and the BICE. Along with these interfaces would have to come protections for data security, guidelines for appropriate searching, and auditing technology to assist with oversight.

In addition, there might be government databases other than those of the TSC and the BICE that contain data useful for the narrowing process in this vignette. Creating locator directories of government databases would make it easier to find this data. Locator directories contain searchable pointers—like card files in a library—to where data can be found. Some private-data holders could also operate their own locator directories and make them available for searching by the government under certain circumstances. In this vignette, the FBI field office could have determined, using private sector locator directories, whether, for example, explosives manufacturers had any recent records on any of the names on the list of skydivers. A "yes" answer would have been a reason to keep a skydiver on the list for additional investigation.

As this vignette demonstrates, searches of aggregated public records are a powerful tool for narrowing the scope of an investigation. For effectiveness and privacy reasons, however, these searches should not be the first step investigators take to narrow a massive list of names. In addition, investigators should conduct these searches consistent with clear guidelines on, among other things, when searches may be conducted, how their results may be used, and how private data may be retained and disseminated.

Finally, there are some improvements to data availability that are too costly to recommend. For example, it would be helpful to the FBI in this vignette if records of skydiving clubs were maintained digitally in a standard format so that they can be searched. It is extremely important that state and local law enforcement and many private sector entities maintain their records digitally so those records are available for searching. There are some industries, though—and skydiving clubs are most likely among them—for which this will be far too costly. The marginal benefit to law enforcement and intelligence is unlikely to be enough to recommend federal funding support for digitizing skydiving-club data.

Appendix E

The Four Key Questions of Detection and Prevention: Who? How? Where? and When?

by Jeff Jonas and Gilman Louie

Who?

In many cases we know a "who." Our federal, state, and local government know of individuals who are not to be permitted into the U.S., who are not to be permitted on planes, who are wanted by law enforcement, etc. Once a "who" is known, the goal becomes finding him before he engages in a "how," "where," or "when."

The first order of business in protecting U.S. assets is to implement a process by which watch lists can be applied to U.S. transactional data (for example, visas or driver's licenses) in search of these individuals. Additionally, it becomes prudent to locate not only the watch list individuals, but also those closely associated with them, such as roommates, etc.

Discovering the whereabouts of a "who" allows law enforcement to determine a course of action (for example, whether to pick individuals up for questioning or surveil them). In either case the objective is to preempt a "how," "where," or "when."

One of the challenges of discovering a "who" is for the U.S. government to determine the correct name of the entity it is searching for. If we plan to make watch lists more effective, we need to have more data on an individual than just his or her name. Identity resolution is a technology that combines many different data points on an individual to determine if there is a match. This technology exists today and is being used by private sector industries.

There are more than a dozen watch lists managed by various agencies. Many of these watch lists are not available for dissemination due to security concerns. In addition, it is difficult for an organization to make sure that all of the disseminated watch lists are current, coordinated, and integrated into all of the appropriate government agencies as well as commercial databases and real-time transactional systems. We need more appropriate methods to manage, update, and unify our watch lists.

Link analysis is a set of tools that helps an analyst understand the relationships between individuals (individuals who, for example, may be related to one another through a common set of associates, may have trained at a common flight school, or lived in the same apartment). The government needs not only the tools but also the data to research and investigate potential linkages. Once again, these technologies and the required data sets exist, but the data sets must be accessible with the appropriate tools.

New technologies, similar to those being pioneered for digital rights management in the entertainment industry, are being developed. These will provide for tighter control and strong audit of the data. In addition, technologies are being developed for anonymization of the data that would enable the enforcement of privacy policies without encumbering intelligence analysis. These capabilities should be ready for deployment within 18 months.

Technologies are already being implemented to help analyze potential risk associated with individuals in near real-time. Government systems such as Computer Assisted Passenger Prescreening (CAPPS) and CAPPS II and commercial credit-analysis systems use rule-based scoring systems to assess risk. The performance of these systems is dependent on the quality of the assumptions used to build the rules, the ability of the system to resolve identities, and the quality of the underlying data. More work needs to be done in measuring the quality of data of any source as well as in the development of technologies that will improve the quality of the rules to reduce the number of false positives and false negatives.

How?

Sometimes intelligence uncovers a potential "how." When a potential "how" is known in conjunction with a "who," very specific sets of data often become of interest. Consider, for example, a threat potentially related to scuba divers. From the "detect and preempt" point of view, gaining access to a specific data set, like scuba diver licenses, can be invaluable. Identity resolution of scuba divers against watch list entities (for example, matching the "whos") can provide excellent insight. Discovering watch list entities who are scuba divers or who are connected to scuba divers opens the door to preempting a "how," "where," or "when."

Sometimes intelligence uncovers a "how" without a "who." What clues are available then? Tactics to unravel such a plot may involve performing identity resolution against several specific data sets for the purpose of generating a "persons of interest" list. For example, if a "how" is believed to involve a passenger cruise liner, a scuba diver, and hazardous materials, it makes sense to correlate future passenger reservations, cruise line employees, scuba diver lists, hazardous materials permit holders, and government watch lists. Identity-resolution and link-analysis outcomes from this step will yield candidates.

There should be a national, or perhaps worldwide, terrorist-methods database (or databases that could be simultaneously searched) that an analyst could employ to determine strategies to prevent an attack as well as to support ongoing investigations. This methods database should include a catalog of potential threats, methods for detection, inventory of necessary components, necessary expertise of individuals to exploit threats, lists of individuals and organizations who may have access to the methods or materials, a database of known devices that currently exist or those that might have been used in the past, and a database of previous threats and attempts. The U.S. should collaborate with other friendly intelligence services to create a worldwide methods repository. There are no technology barriers to developing such a repository.

In addition to a methods database, we should create standard operating procedures for each method that can be employed by national, state, local, and commercial assets, as well as by first responders, to assist in preparedness, detection, prevention, and consequence management. To support the development of standard operating procedures, we should collect lessons learned from national, state, and local simulations of attacks, and create a national test plan. We should also deploy digital-simulation technologies to support the development of scenarios and potential responses to scenarios. These simulation tools can also be deployed for training and rehearsal. The underlying technologies for the creation of these tools exist in both defense applications and computer-gaming applications, such as the massive multiplayer games *SimCity* and *The Sims*. The application of digital-simulation technologies to support homeland defense could be developed in less than 24 months.

Where?

Knowing a "where" and a "who" or a "where" and a "how" can provide enough clues to solve the remaining plot condition. For example, a known "where" and "who" provide enough focus to select very specific data sets for analysis. If the "where" is a hotel, then comparing past guests, future reservations, employees, vendors, government watch lists, and the known "who" may very well uncover the links needed to crack the case. A broader view might include testing data from surrounding hotels and regional transportation records.

The U.S. government should have a geospatial repository, or a network of geospatial databases that should include all major structures, critical infrastructure, and any potential terrorist targets. This database should support analysts attempting to answer the questions "What's there?" and "How is it vulnerable?" Much of the data to build this repository lies not with the federal government but with state, local, and commercial repositories. We need to be able to either collect the data or access it. The good news is that there are both commercial standards and emerging open standards for geospatial data interchange. The challenge is to make the data accessible and searchable with appropriate access controls. It is also important for the government to perform a risk assessment of major commercial as well as federal, state, and local infrastructures, and moderate- to high-risk targets.

There should also be a sensor network of chemical, biological, and radiological sensors networked with traditional physical-security systems monitoring critical infrastructure and potential government and commercial targets. These sensors should be monitored by one or more network-awareness centers (NACs). We could have a national sensor grid with initial capability in less than 24 months.

These NACs should also be monitoring ongoing message traffic (both radio and data messaging) of police, fire, and medical personnel, and should have appropriate technologies to analyze these collections for patterns as well as for early warning. The technologies required to support this effort have already been built for signal intelligence and collections.

We should also be able to track most potential delivery systems (for example, aircraft, ships, large trucks, containers) by using existing commercial GPS tracking systems and fleet data-management systems. These systems already exist for commercial-transportation management.

Similar to identity-link analysis, we need to develop technologies that support geospatial link analysis. For example, starting with a location, with or without a time reference, an analyst should be able to determine all of the high-threat individuals within an area or who have passed through a given area. The analyst should be able to determine all of the potential targets or potential areas of interest, given a profile. She should be able to identify a group of individuals and see if they have ever been physically together.

When?

Having intelligence or a predictive notion of a "when" helps significantly to reveal a plot. Each of the above examples is further scoped and focused when a specific time constraint can be added. To effectively detect a plot with a known "when," another element—whether that be a "who," "how," or "where"—is required.

There should be a national calendar that lists all of the major events with locations and audience. We should have the ability to track the travel of key individuals whom we are trying to protect as well as those who are considered high potential threats. In order to do so, we need to be able to correlate reservation, travel, and lodging data. We should also be tracking the transportation schedules of hazardous materials. Appropriate temporal and geospatial analysis and visualization tools could assist the analyst in answering the "when" question.

Appendix F

Technology Challenges for the Near Future

by Stewart Baker and Jeff Jonas

Introduction

Recent headlines about the government's technology capabilities in fighting terrorism have suggested that agencies are deploying cutting-edge data-mining capabilities that seek to identify terrorists by knowing everything about everyone. These suggestions are unfortunate in several respects. First, they grossly overestimate the government's technical capabilities, both now and in any plausible immediate future. The most ambitious effort, the Terrorism Information Awareness initiative at the Defense Advanced Research Projects Agency (DARPA), includes one project designed to explore the ability of investigators and computers to identify terrorist activity in advance by processing transactional and other information. This is highly speculative research, and there is no guarantee that it will produce useful results. Second, these suggestions distract us from understanding the government's very real lack of current capabilities and actually undercut responsible efforts to improve those capabilities. In an effort to focus attention on capabilities that the government should have—and that the government could have if it used existing data-processing technology—this paper seeks to identify a set of concrete challenges for the near future.

Finding terrorists before they strike is not unlike a high-stakes game of *Clue*. To be sure of success, the government is likely to need information about the identities of the terrorists, the weapons they plan to use, and the location they intend to strike—who, how, and where. We have identified a series of capabilities that seeks to improve the government's ability to answer each of these questions. These are capabilities 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 industry. Using them in an integrated fashion could enhance our safety in a manner consistent with current law while also taking into account concerns about privacy and civil liberties. Privacy concerns that go beyond the protections of current law should be addressed not by denying the government the ability to use technology or by imposing new legal restrictions on government investigations of terrorism, but by using technology to enforce accountability and reduce or eliminate access to data unrelated to terrorism. Proposals for such a use of technology are being prepared by other task forces.

Who?

By far the most productive approach to preventing terrorism is identifying terrorists before they strike. Therefore, the greatest number of challenges focuses on this problem, which can be subdivided into two categories: locating suspected terrorists and detecting when a suspected terrorist is operating under a false identity.

Challenge 1: Finding known terrorists and associates operating in the U.S.

When a counterterrorism agency knows the identity of a suspected terrorist, it should be able to determine whether the terrorist is in the country. Data searches need to be conducted in an attempt to locate that person on an ongoing basis, using phone listings (published and unpublished), DMV records, basic financial indicators (as already used by database marketers), Immigration and Naturalization Service (INS) visitor and immigration information, academic enrollment, special licenses, and travel records. Within 30 seconds, the counterterrorism agency should also be able to access U.S. and international financial records associated with the suspect.

Counterterrorism officers should be able to identify known associates of the terrorist suspect within 30 seconds, using shared addresses, records of phone calls to and from the suspect's phone, emails to and from the suspect's accounts, financial transactions, travel history and reservations, and common memberships in organizations, including (with appropriate safeguards) religious and expressive organizations.

Rationale

On August 26, 2001, two weeks before the hijackings, the Federal Bureau of Investigation (FBI) received unequivocal word that two of the hijackers were in the country and were associated with a "major-league killer" in Al Qaeda. Despite having two weeks to find them and their associates, the FBI failed. There were two principal reasons for this failure. The first was an unwillingness to use law enforcement resources in the search, due to the wall between law enforcement and intelligence, both inside and outside the agency. The second was an inadequate technical capability that made tracking the two hijackers difficult, despite the fact that they were living under their own names, were listed in the phone book, had driver's licenses, and shared a variety of travel information with their air carriers. They were, in short, eminently findable. And once found, a search of other private databases (for example, airline systems) would have turned up links to many of the other hijackers. Done promptly, such searches might have stopped the attacks. It may be inappropriate to blame the government for not having in place a system for finding a conspiracy with such an unexpected goal. But Al Qaeda's goals are no longer unexpected, and the next time they attack we may not have two weeks. The government must implement procedures that will at a minimum prevent failures such as those in the past. This is not enough, but it is the first thing that must be done.

The technology to meet these challenges is already in existence. Indeed, versions of the technology are already in use in some industries, such as the gambling industry (see Appendix G). The technical challenge, which cannot be underestimated, is to bring the capabilities of counterterrorism agencies up to the capabilities of private industry so that American lives receive the same protection as the business interests of the private sector.

The challenge includes a requirement that investigators be able to use information about membership in organizations, including religious and expressive organizations, when examining a known subject. Denying investigators access to such information is not the answer to civil-liberties concerns. The American commitment to equality is not violated by observing that many of the 1993 World Trade Center bombers were linked through a common religious leader. Nor is it a violation of civil liberties to notice that those who belong to an organization advocating "Death to America" are more likely to be planning the deaths of particular Americans than members of an organization devoted to highway beautification. At the same time, it is possible to misuse such information. Safeguards should be designed against improper access to such information—"pretext" searches and the like. Safeguards should also be designed to discourage improper use of the information. These safeguards may include careful authentication of users, audits of the data accessed, and scrutiny of unusual search patterns by users of the system.

Required technologies and infrastructure

1. Active watch list program
2. Connectivity between key data holders
3. Data-sharing guidelines, policies, and procedures
4. Identity recognition
5. Immutable audit
6. Link analysis
7. Locator directories

Challenge 2: Foreign-student accountability

The government should be able to search, in real time, records showing the status and locations of foreign students, including prospective and former students, research assistants, and teachers in programs that raise terrorism concerns.

Rationale

Many of the hijackers of September 11 came to the U.S. on student visas. But many student-visa holders do not show up, or soon abandon their studies, or overstay their visas. Of equal concern are the students who are here to learn skills that will be used to kill Americans.

Required technologies and infrastructure

1. Connectivity between key data holders
2. Data-sharing guidelines, policies, and procedures
3. Identity recognition
4. Immutable audit
5. Locator directories

Challenge 3: Enabling local law enforcement with watch lists

Local police checking driver's licenses or license plates should, in most cases, be automatically alerted when they run the documentation of a terrorist suspect. However, the watch list database should not be easily reconstructed by local police agencies, and the alert should be tailored to the circumstances of the suspect and the stop.

Rationale

Local law enforcement is an essential element of antiterrorist strategy, but integrating local agencies into federal data capabilities is a complex matter. Local police had several interactions with the September 11 hijackers while they were in the U.S. Assuming that we are doing a better job of sharing information about terrorist suspects now believed to be operating in the U.S., local police are the most likely to encounter the suspects. Integrating identity checks performed by local police with federal suspect lists is thus a priority.

While providing access to counterterrorism databases is a challenge, it is not technically demanding. Here, the more difficult problem will be to build the safeguards. A single database that can be accessed by every law enforcement agency in the country will not likely be secure and thus will not likely contain the most important and sensitive information. An effective system must, therefore, include strong safeguards to ensure accountability, audits, pattern reviews of searches, and similar protections. The good news about this challenge is that the same technical capabilities that must be developed to meet the challenge can also be used to prevent other forms of misuse, including abuses of civil liberties and privacy.

Required technologies and infrastructure

1. Active watch list program
2. Anonymized data, sharing, and analytic correlation
3. Connectivity between key data holders
4. Data-sharing guidelines, policies, and procedures
5. Identity recognition
6. Immutable audit
7. Link analysis

Challenge 4: Creation of a consolidated watch list

The government should have a consolidated list of terrorism suspects that includes the different lists that have been assembled by different agencies for different purposes.

Rationale

Once again, the most difficult challenge here may turn out to be the problem of maintaining a highly sensitive list without having its contents end up on bulletin boards in every Customs back office. The safeguards designed to make sure the list is not accessed directly or improperly may also serve privacy interests.

Other challenges concern the problem of how to avoid being swamped with false positives. These can call the system into disrepute while also weakening security. For example: Simply placing "David Nelson" on a watch list of people who are banned from flying causes every David Nelson in the country to be stopped at the airport. Safeguard mechanisms are likely to include an ability to immediately recognize that the 71-year-old David Nelson from Oregon is not the "no fly" David Nelson, so that the same list of questions and background checks are not needed before every flight to conclude that the Oregon David Nelson is free to fly.

Required technologies and infrastructure

1. Active watch list program
2. Anonymized data, sharing, and analytic correlation
3. Connectivity between key data holders
4. Data-sharing guidelines, policies, and procedures
5. Identity recognition
6. Immutable audit
7. Link analysis
8. Locator directories

Challenge 5: Watch list development and sharing policies

Watch lists should be updated in an accountable fashion on a real-time basis.

Rationale

Watch lists must conform to a standard process that clarifies how names get on and off these lists. Then, as list holders make changes, these same changes must be instantly transferred to the centralized watch list. In turn, the updates to the centralized watch list must be disseminated to watch list subscribers.

Required technologies and infrastructure

1. Active watch list program
2. Data-sharing guidelines, policies, and procedures
3. Immutable audit

Challenge 6: Detecting false and stolen identities

Both the government and the private sector should be able to identify false identities in real time when vetting employees or preparing to engage in a material transaction—opening a bank account, making a cruise-ship reservation, providing a pilot's license, etc. This necessitates, for example, checking identities against death records for individuals (usually children who have died young enough to avoid acquiring a social security number) whose identities might be used to generate a false identity and flagging improbable identities, such as that of a 35-year-old with unusually few public records (for example, no phone book records, no credit-header files, no driver's license).

Rationale

Our most effective systems for investigating and protecting against possible terrorists depend on knowing the identities of suspects. But if identities are easy to forge, the government is forced quickly back into the position of treating everyone as a suspect, with unfortunate consequences for civil liberties. Thus, it is important to find ways to reduce opportunities for false identities. The capabilities identified in this challenge have been available to Western European governments for many years, and it is embarrassing that the U.S. hasn't yet automated them, despite the use by several September 11 hijackers of false identity papers.

Required technologies and infrastructure

1. Connectivity between key data holders
2. Identity recognition
3. Immutable audit
4. Locator directories
5. Real-person validation

How and where?

Sometimes we will not know the identities of possible terrorists but will have some idea of their plans, locations, and activities. Perhaps the most demanding challenge is finding ways to identify terrorists based on knowing little more than a potential target or threat.

Challenge 7: Accessing data about people in response to a credible methodology threat

When the government develops a credible new concern about a possible terrorist methodology—the intent to use a hazmat tanker in suicide attacks, for example, or scuba attacks against a specific port—it should be able to selectively request and receive data sets of specific interest associated with the threat. For example, it should be able to compare a list of persons with hazmat or scuba licenses against watch lists and other data sets that may give rise to concerns, such as travel, origin, or communications with foreign countries that are sources of terrorism; association with other terrorism suspects; and the like.

Rationale

As with the first challenge, this need grows out of the circumstances of September 11. An FBI agent in Phoenix raised the possibility that terrorist suspects were disproportionately enrolling in flight schools. No search was performed of flight-school records, perhaps for fear of charges of ethnic or religious profiling, but largely because of the difficulty of conducting rapid, efficient searches to test hypotheses about possible terrorist plots. While such a hypothesis is not a basis for assembling files on every scuba diver in the country, a review that located and flagged scuba divers who have overstayed a Yemeni visa and have bank accounts that are replenished regularly from foreign sources is an important capability that should be available on a decentralized basis so as to allow decentralized hypothesis-testing by agents in the field. The ability to conduct a "virtual background investigation" on individuals—most of whom will have nothing to do with terrorism—also requires safeguards. In addition to accountability safeguards of the sort identified above, it would be prudent to design systems that maintain practical anonymity for the subjects of such reviews. That is, it should be possible to conduct a background investigation of hazmat-license holders without maintaining or even allowing human review of the information unless the investigation turns up other indicia of concern such as the factors described above. Identifying the indicia of concern is not a simple or a one-time matter. Extensive contacts with Middle Eastern countries, an attachment to Islamic fundamentalism, and foreign travel to countries associated with terrorism are all indicia of concern today and for the foreseeable future, but as Al Qaeda steps up its nontraditional recruiting to avoid these indicia, others may have to be added. The Padilla case suggests that prison time, particularly prison time combined with conversion to Islam, should be an indicium of concern. Other indicia may need to be kept confidential. It may be appropriate to develop scoring mechanisms that do not identify the particular indicia that contributed to the score but that simply order the data to identify the people who should be examined more closely first.

Required technologies and infrastructure

1. Active watch list program
2. Anonymized data, sharing, and analytic correlation
3. Connectivity between key data holders
4. Data-sharing guidelines, policies, and procedures
5. Identity recognition
6. Immutable audit
7. Link analysis
8. Locator directories

Challenge 8: Accessing resource and infrastructure data specific to a credible methodology threat

The government should be able to respond to reports of a particular mode of attack (for example, a plan to use chlorine tanker trucks to attack office buildings in several cities) by gaining access within four hours to private sector data relating to the status of that mode (for example, to obtain available information from industry sources about the location, status, drivers, and contact information for chlorine tankers).

Rationale

This challenge assumes that counterterrorism agencies will have to guard against a specific threat without knowing who will carry out the threat. In many cases, it will be possible to locate all sources of threat information much more quickly than within four hours. Presumably, if the government had been aware that a suicide hijacking was planned for the immediate future, the Federal Aviation Administration (FAA) would have been able to identify in less than four hours all flights scheduled for departure on September 11, 2001. But not all industries are as regulated or centralized as the airline industry, and elaborate information-sharing mechanisms are not likely to be cost-effective in many circumstances. Instead, the government needs standby mechanisms for rapidly gaining access to such information when a particular threat is identified. This means tools, links, organizational contacts, and knowledge about the kinds of data maintained by chemical companies, nuclear plants, truckers, petroleum companies, railroads, and the like. The government also needs a mechanism for keeping these tools, links, and facts up to date, a task that is achievable—but only if the government makes the effort to identify the data of particular importance in an emergency and limits its data requirements to only that data.

Required technologies and infrastructure

1. Connectivity between key data holders
2. Critical-infrastructure risk assessment
3. Data-sharing guidelines, policies, and procedures
4. Distributed environmental-sensor web
5. Geospatial and event query support
6. Major-events calendar
7. Terrorist-methodologies database with threat profiling
8. Threat-scenario simulation
9. What/where recognition

Challenge 9: Alerts of suspect international cargo containers

The U.S. should be able to determine the past history—cargo and itinerary—of containers bound for its ports, and should be able to identify suspicious patterns before those containers reach American waters.

Rationale

Containerization has revolutionized world shipping. But its ubiquity could also become a serious hazard in an age of weapons of mass destruction. In fact, substantial information about containers is gathered at every stage of the container's progress, but this information has not been stored in an accessible fashion or transmitted across national boundaries. Concerted U.S. leadership could end this gap in our data capabilities and also provide a new tool for identifying potential weapons before they reach our shores.

Required technologies and infrastructure

1. Connectivity between key data holders
2. Critical-infrastructure risk assessment
3. Data-sharing guidelines, policies, and procedures
4. Distributed environmental-sensor web
5. Geospatial and event query support
6. Identity recognition
7. Link analysis
8. Terrorist-methodologies database with threat profiling
9. Threat-scenario simulation
10. What/where recognition

Challenge 10: Detection of terrorist-sponsored money-laundering activities

Financial institutions conducting anti–money-laundering reviews should be able to identify account holders whose finances reflect such indicia of concern as irregular deposits from overseas. It should also be possible to review the background of such account holders on a rapid basis for other indicia of concern.

Rationale

U.S. law now requires extensive information-gathering and -processing by financial institutions of anti–money-laundering data designed to locate terrorist financing. But there is considerable uncertainty among financial institutions about how to identify financial patterns associated with terrorism. And in many cases, a review of financial information can only begin the analysis; it will be necessary to review data from other sources to confirm or rebut suspicions raised by anti–money-laundering scrutiny. While terrorist financing is a potential source of effective counterterrorism action, it must be focused and integrated with other data to succeed.

Required technologies and infrastructure

1. Active watch list program
2. Connectivity between key data holders
3. Data-sharing guidelines, policies, and procedures
4. Identity recognition
5. Immutable audit
6. Link analysis
7. Locator directories
8. Terrorist-methodologies database with threat profiling
9. Transactional-pattern analysis

Challenge 11: Accessing geographic data specific to a credible location threat

Sometimes intelligence may only be able to produce a "where"-related threat (for example, a scenario in which a major sporting event at a specific stadium is believed to be a target). In this case, data must be accessible that enables analysts to rapidly assess the threat and hunt for other corroborating evidence or activity, including potential relationships to any watch list entities.

Required technologies and infrastructure

1. Active watch list screening
2. Connectivity between key data holders
3. Critical-infrastructure risk assessment
4. Data-sharing guidelines, policies, and procedures
5. Geospatial and event query support
6. Identity recognition
7. Immutable audit
8. Link analysis
9. Locator directories
10. Major-events calendar
11. Terrorist-methodologies database with threat profiling
12. Transactional-pattern analysis
13. What/where recognition

Challenge 12: Prompt response to actual incident

The government should have the ability to locate critical infrastructure nodes in the vicinity of an attack within five minutes—pipelines, power-generation plants and transmission lines, communications facilities, transportation, and the like. The impact of a major attack could include much more than the immediate casualties if the responding agencies are not able to respond with full knowledge of nearby facilities that may pose a threat or provide resources. These facilities should be identified once—not multiple times by multiple agencies at the federal, state, and local levels—and their identities made available to first responders in a method that does not expose the information to public (and therefore possible terrorist) access.

Required technologies and infrastructure

1. Connectivity between key data holders
2. Convergence of emergency communication systems
3. Critical-infrastructure risk assessment
4. Data-sharing guidelines, policies, and procedures
5. Distributed environmental-sensor web
6. Geospatial and event query support
7. Major-events calendar
8. Terrorist-methodologies database with threat profiling
9. Threat-scenario simulation
10. What/where recognition

Appendix G

Technologies Required to Meet the Challenges

by Jeff Jonas and Gilman Louie

Introduction

In "Technology Challenges for the Near Future" (Appendix F), we present 12 scenarios to illustrate technologies, infrastructures, and related data issues that will be instrumental in enhancing our national security. The overall requirements are reduced here to a finite number of specific enabling technical capabilities. In the chart below, these capabilities are presented in alphabetical order, so as to enable the reader of "Technology Challenges for the Near Future" to find the description, availability, and best-case time frame for implementation of each capability. The capabilities are then organized into three prioritized phases of implementation.

This document supports the optimistic position that many requirements to improve national security can be met by existing technologies, and that it is possible to implement these technologies in very short order. In reality, the challenges will essentially be tied to changing culture and consensus regarding guidelines and policies.

Required technology and infrastructure

How to read this chart

Capability: This value is for the technical capabilities we described in Appendix F as necessary to enhance our national security.

Availability: This value is the amount of time it might take in a perfect world, and with appropriate funding, to make the technology useable for the intended mission. As can be seen, the majority of these capabilities are already available.

Best-case time frame for implementation: This value provides a time frame for actual implementation as measured in months or years. These time frames are for implementation in a limited fashion in the most relevant areas and for delivery of some immediate enhancement to national security.

CAPABILITY	AVAILABILITY	BEST-CASE TIME FRAME FOR IMPLEMENTATION
Active watch list program Ability to aggregate various federal, state, and local watch lists into a single repository (the centralized watch list must be kept current with source systems, and the data on it must be able to be securely queried and securely disseminated); access audits and policies on how names get on and off each type of watch list required	Available	6 to 12 months
Anonymized data-sharing and analytic correlation Ability to convert actual data values to anonymous values before data is shared between parties; recipients of anonymized data must be able to perform analytical processing against anonymized data	3 to 9 months	6 to 18 months

CAPABILITY	AVAILABILITY	BEST-CASE TIME FRAME FOR IMPLEMENTATION
Connectivity between key data holders Including the ability to sustain real-time interfaces between key systems at federal and state offices (for example, the Bureau of Citizenship and Immigration Services and the FBI); ability to connect government systems with data aggregators for real-time information requests and responses; ability to sustain real-time interfaces with enterprise-class organizations, which have highly automated, high-volume transactional systems; ability to support at least batch interfaces with highly autonomous, independent, noncentralized organizations engaging in transactional activity	Aggregator connectivity: Available Government connectivity: 6 to 12 months Enterprise-class connectivity: Available Independent-organization connectivity: Available	Aggregator connectivity: 3 to 36 months (challenges include data security, policy, and culture) Government connectivity: 12 to 24 months Enterprise-class connectivity: 6 to 12 months Independent-organization connectivity: 12 to 24 months (only practical on a very selective basis)
Convergence of emergency communication systems Ability for first responders, operators, and facility managers of critical infrastructure and high-risk targets to communicate via integrated and interoperable platforms	Available	1 to 5 years
Critical-infrastructure risk assessment Including the creation of reporting standards for critical-infrastructure locations, inventory, vulnerabilities, practices, and anomaly reporting; creation of a central repository containing up-to-date critical-infrastructure reports enabling vulnerability assessments, analytics, and hypothesis exploration; and the ability to assess and rank critical-infrastructure risks based on a centralized critical-infrastructure repository, terrorist-methodologies database, and threat-scenario simulations—all of which must include the related visualization tools to interact with the data	Technology: Available Reporting standards: 3 to 5 years (very difficult)	Limited coverage: 1 to 5 years or more
Data-sharing guidelines, policies, and procedures Ability to develop agreements between data creators and data users concerning policies and procedures for sharing highly protected intellectual property; must include policy and standards for digital-rights management, encryption, anonymization, reporting, currency, connectivity, synchronization, and precedence rules	Available	1 to 10 years
Distributed environmental-sensor web Ability to deploy and integrate network-robust environmental sensors for weather, wind, biological, chemical, and nuclear information	Available (with the exception of biological)	3 to 5 years

CAPABILITY	AVAILABILITY	BEST-CASE TIME FRAME FOR IMPLEMENTATION
Entity extraction from unstructured data Ability to locate and extract "who," "what," "where," and "when" values from unstructured text (for example, letters and newspapers) with a reasonable level of accuracy and little to no human intervention	Available (at a reasonable level of accuracy)	In progress
Geospatial and event query support Ability to query critical infrastructure databases, resource databases, threat databases, terrorists, suspects, etc. on the basis of a geospatial area; must include the related visualization tools for interacting with the data	Available	12 to 24 months (dependent upon collection of appropriate data)
Identity resolution Ability to recognize when two individuals or organizations are the same across data sources, despite disparity (for example, poor data quality or obfuscation); required to raise the fidelity of watch list data and transactional data, which in turn reduces false positives and false negatives	Available	In progress
Immutable audit Ability to maintain detailed audit logs that are highly tamper-resistant (including data authors, maintenance changes, system queries, and query responses) and that are stored for after-the-fact analysis and integration of real-time trip wires	3 to 9 months	12 to 36 months
Link analysis Ability to connect people or organizations based on common attributes (for example, address or phone number) to watch list identities at one or more degrees of separation; must include the related visualization tools for interacting with the data	Available	In limited use
Locator directories Ability to create locator directories (locator directories contain pointers to where data can be found)	Available	12 to 18 months
Major-events calendar Ability to create a centralized collection of major events (for example, holidays, sporting events, and concerts), which could provide analysts with critical information to correlate with threat intelligence and input for threat-scenario simulations	Available	6 to 36 months

CAPABILITY	AVAILABILITY	BEST-CASE TIME FRAME FOR IMPLEMENTATION
Real-person validation Ability to confirm that individuals presenting themselves are who they say they are; required to prevent access to secure areas by those using false or stolen identities	False identities: Available Stolen identities: 6 to 12 months (true solution requires biometrics)	False identities: 6 to 12 months Stolen identities (with biometrics): 3 to 10 years
Terrorist-methodologies database with threat profiling Ability to develop a terrorist-methods database (including required resources, expertise, known targets, known terrorist skills, etc.) that will support behavioral profiling of terrorists; and ability to develop a model or signature (for example, a large purchase of ammonia nitrate and rental of a moving truck) that might suggest future intent; must include the related visualization tools for interacting with the data	Some available, but further research needed	5 or more years
Threat-scenario simulations Ability to use digital-simulation technologies for training, test plans, simulated outcomes, rehearsal, etc., in efforts to avert an event; must include the related visualization tools for interacting with the data	12 to 24 months	18 to 36 months
Transactional-pattern analysis Ability to integrate geospatial, temporal, and event data that can be used to generate alerts and enable analysts to query for specific hypotheses; must include the related visualization tools for interacting with the data	Somewhat available (lack training patterns for terrorism)	2 to 5 years
What/where resolution Ability to resolve disparate data that describes the same object (what) or place (where)[1]	"What" resolution: Very limited availability based on subject area "Where" resolution: Moderate availability	"What" resolution: 3 to 5 years "Where" resolution: 12 to 18 months

[1] While identity resolution solves the challenging problem of understanding when two people are the same, similar capabilities are required to resolve object/what or place/where data. For example, chlorine might appear in scientific data in any one of these forms: (1.) Name_Chlorine; (2.) Atomic Number: 17; (3.) Atomic Symbol: Cl; (4.) Atomic Weight: 35.453; or (5.) Electron Configuration: [Ne]$3s^2 3p^5$. "Where" resolution would need to establish, for example, that the following locations are one and the same: (1.) the Stafford Building; (2.) 1104 48th Street; (3.) the S.E. corner of Stafford and Vine; and (4.) Starbucks location #246.

Priorities for technology implementation

We believe all of these capabilities are urgently needed. However, knowing there must be some prioritization for focus, we have organized the capabilities into three phases. These priorities were developed with consideration of the following general characteristics: criticality to national security, criticality to privacy and civil liberties, and potential for timely implementation.

PHASE 1: OPERATIONAL WATCH LISTS (WHO)

1. Anonymized data-sharing and analytic correlation
2. Active watch list program
3. Connectivity between key data holders
4. Data-sharing guidelines, policies, and procedures
5. Identity recognition
6. Link analysis
7. Real-person validation

PHASE 2: ENHANCED ANALYTICS (WHAT, WHERE, WHEN)

1. Entity extraction from unstructured data
2. Geospatial and event query support
3. Immutable audit
4. Locator directories
5. Major-events calendar
6. Transactional-pattern analysis
7. What/where resolution

PHASE 3: INFRASTRUCTURE MONITORING, DISASTER RECOVERY, AND SIMULATION

1. Convergence of emergency communication systems
2. Critical-infrastructure risk assessment
3. Distributed environmental-sensor web
4. Terrorist-methodologies database with threat profiling
5. Threat-scenario simulations

These capabilities should be worked on now so that they can be ready in coming years.

Appendix H

The Landscape of Available Data

by Jeff Jonas

Introduction

The purpose of this appendix is to present the types of data that exist as a byproduct of our digital economy. This chart below should not be viewed as a comprehensive reference work, but rather as a sketch of existing digital data. While much of this data is not generally shared by its holders, its existence reveals the fact that the landscape of available data is rather rich. It is hoped that as guidelines and policies are considered, this chart will be of assistance by presenting the big picture of existing data and usage.

Data available in the U.S.

How to read this chart

Data source: This value represents general life events, actions, industries, etc., from which data is generated. We included 26 data sources.

Record: This value represents types of documents that are generated from the corresponding data source.

Domain: This value represents the availability of the corresponding record as follows: "free" (available via the Internet or made available upon request to its collector); "public" (government data that is considered public record and is generally available without restriction); "for purchase"; "for limited use" (subject to use restrictions consistent with state or federal law); and "private" (generally not for sale under any circumstance, unless a person gives express consent—for example, an authorization to pull a credit report during a loan application).

Class: This value defines the nature of the record as follows: "PII" (data that contains personally identifiable information, such as a name, address, or phone number); and "transactional" (data acquired by means of a transaction, such as the purchase of flying lessons). PII data often spells out an action ("has subscribed to a magazine") or status ("has pilot's license"). For the purposes of our chart, the term "transactional" implies that each transaction is associable to a specific person.

Some organizations with centralized access: This value is for organizations that possess either aggregated data from the corresponding record or connectivity to such data. Where this field is blank, it should not be inferred that no such organization exists. In such cases, we were simply unable to find such an organization. Also, while not indicated on our chart, there are differing degrees of latency in the data at various aggregation points. For example, bad debt, initially documented by a credit grantor, is often reported months after the fact to credit-reporting agencies. Those agencies, in turn, provide the information to other aggregators on a weekly, monthly, or, even quarterly basis (the frequency depends on the aggregator's relationship with its credit-bureau supplier).

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Births, deaths, and marriages	Birth certificate	Private	PII	VitalChek
	Death certificate	Public	PII	Social Security Administration
	Divorce papers	Public, or for limited use (varies by state)	PII	VitalChek
	Marriage certificate	Public, or for limited use (varies by state)	PII	VitalChek

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Communications	Calling-card log	Private	Transactional (often without PII reference)	
	Cellular geo-positional locator	Private	Transactional	Cellular carriers
	Internet chat-room dialogue	Private	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Email-account directory	For purchase	PII	411.com
	Email	Private	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Express-mail form	Private	PII and Transactional	USPS, FedEx, UPS, and Airborne Express
	Instant message	Private	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	ISP subscriber list	For limited use	PII	
	Page or text message	Private	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Phone call	Private	Transactional	Phone carriers
	Prepaid phone card	For limited use	PII	Amdocs (toll calls), phone carriers
Corporations	Business license	Public	PII	ChoicePoint
	Corporate officer and director lists	Public	PII	Edgar
	SEC filing (for example, a 10Q or 10K)	Public	PII	Edgar
Courts, county recorders, and secretaries of state	Bankruptcy records	Public	PII	Banko, ChoicePoint, TransUnion, Equifax, and Experian
	Eviction notice	Public	PII	ChoicePoint
	Lien	Public	PII	Banko, NDR, TransUnion, Equifax, and Experian
	Pleading, motion, complaint, judgment, order, and other civil recordings or filings	Public	PII	Westlaw[1]
	UCC filing	Public	PII	ChoicePoint
Credit and banking industries	Credit card application	Private	PII	
	Documentation of credit card issuance	For limited use	PII	VISA, MasterCard, and American Express

[1] U.S. Supreme Court, Circuit Court, Court of Appeals decisions, and reported district cases from State Supreme and Appellate Court decisions can be purchased from Westlaw. Generally, courts are quite far behind in records automation. Some data aggregators have some PII data related to certain county court abstracts.

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Credit and banking industries (cont.)	Credit card transaction report	For limited use	Transactional	VISA, MasterCard, American Express, and First Data Corp.
	Credit report derogatory line	Private	Transactional	TransUnion, Equifax, and Experian
	Credit report header	For limited use	PII	Equifax, Experian, ChoicePoint, and Lexis-Nexis
	Credit report inquiry line	Private	Transactional	TransUnion, Equifax, and Experian
	Credit report trade line	Private	Transactional	TransUnion, Equifax, and Experian
	Debit card transaction report	For limited use	Transactional	
	Fraud-protection registry (self-enrolled)	For limited use	Transactional	TransUnion, Equifax, and Experian
	Loan application	For limited use	PII	
	Loan-issuance documentation	For limited use	PII	TransUnion, Equifax, and Experian
Education	Academic-institution records	For purchase	PII	List brokers
	Educator records	For purchase	PII	List brokers
	Enrollment records	For limited use	PII	
	Alumni list	For limited use	PII	Classmates.com
Entries and exits (U.S.)	Border entry and exit records	For limited use	PII and transactional	BCIS I-94s
	Passport	For limited use	PII	BCIS
	U.S. visa application	For limited use	PII and transactional	DOS's Consolidated Consular Database
	Visa	For limited use	PII and transactional	BCIS
Import and export	Container shipment record	For purchase	PII and transactional	PIERS
	Crew registration	For limited use	PII	
	Ship registration	For purchase	PII	List broker
Insurance	Claim	For limited use	Transactional	ChoicePoint

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Insurance (cont.)	Policy application	For limited use	PII	ChoicePoint
	Policy	For limited use	PII	ChoicePoint
Internet	File downloads	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	File postings	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Online purchases	For limited use	Transactional	eBay, Amazon, and Travelscape
	Website search history	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, Google, AltaVista, MapQuest, and eBay
	Web-page-hits record	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Domain-name registrations	Free	PII	Atriks and Internic
Licensing	Aircraft owner documentation	Free	PII	List brokers and ChoicePoint
	Automobile registration	For limited use	PII	Experian and ChoicePoint
	Flight-instructor license	Free	PII	List brokers
	Scuba-diving certification	For limited use	PII	Scuba-certification organizations PADI, NAUI, SSI/NASDS, SDI, and YMCA
	Concealed-weapons permit	Public or Limited (by state)	PII	ChoicePoint
	Commercial or noncommercial driver's license	For limited use	PII	ChoicePoint
	Driving record	For limited use	Transactional	DMV
	Fishing license	Public or for limited use (by state)	PII	ChoicePoint
	Gun background check	For limited use	PII	ATF
	Hazardous-material license	Public	PII	ChoicePoint
	Hunting license	Public or for limited use (by state)	PII	ChoicePoint
	Pilot's license	Public or for limited use (by state)	PII	List brokers and ChoicePoint

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Licensing (cont.)	Trucking permit	Public or for limited use (by state)	PII	PermitVision
	Weapons permit	Public or for limited use (by state)	PII	ChoicePoint
Lifestyle interest	Cable-viewing history	Private	Transactional	
	Library-materials user records	Private	Transactional	
	Magazine or newspaper subscription	For purchase	PII and transactional	Acxiom
	Online-group records	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Internet opt-in news sources	For limited use	Transactional	AOL, MSN, Yahoo, CompuServe, and EarthLink
	Product activation	For purchase	PII	
	Product purchase or registration warranty card	For purchase	PII	Acxiom
	Video rental	Private	Transactional	
Loyalty and affinity rewards programs	Grocery store loyalty-program record	For limited use	PII	
	Loyalty-based transaction record (cash-only, etc.)	For limited use	Transactional	
	Travel loyalty-program record (airline, rental car, hotel, train, etc.)	For limited use	PII	Global distribution systems Galileo, Sabre, WorldSpan, and Amadeus; and central reservation systems Airline Automation Inc., and Cendant
Marketing	Cluster-code flag	For purchase	Transactional	Marketing data aggregators Acxiom and MITI
	Income-indicator flag	For purchase	Transactional	Acxiom and MITI
	Presence-of-children flag	For purchase	Transactional	Acxiom and MITI
	Purchasing-power flag	For purchase	Transactional	Acxiom and MITI
Medical	Drug prescription	For limited use	PII	IMS
	Infectious-disease record	For limited use	PII	InfoUSA
	Laboratory results	For limited use	PII	
Memberships	Labor association records	For limited use	PII	

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Memberships (cont.)	Political organization records	For limited use	PII	
	Recreational club records	For limited use	PII	
	Religious or expressive organization records	For limited use	PII	
	Technical association records	For limited use	PII	
	Trade association records	For purchase	PII	List brokers
Open-forum meeting	Conference attendee list	For purchase	PII	Reed Elsevier
	List of conference speakers	For purchase	PII	Reed Elsevier
Open source	News story	Free	Transactional	Lexis-Nexis
	Press release	Free	Transactional	Lexis-Nexis
	Published research paper	Free	Transactional	Lexis-Nexis
People	Competition record	Free	PII	Online competition results by association or club
	List of distinguished persons	For purchase	PII	Who's Who Registers
	Lists of executives	For purchase	PII	
	Professionals lists	For purchase	PII	
Politics	Political contributions	Public	PII	FECInfo/tray.com's Political Moneyline
	List of politicians	For purchase	PII	List brokers
	Voter registration	For limited use	PII	Aristotle's Voter-ListsOnline.com
Postal	National Change of Address (NCOA)	For limited use	Transactional	USPS and Group 1
	Post-office and mail-drop box owners	Private	PII	USPS
Preemployment	Job applications	For limited use	PII	Monster.com
	Employment-history records	For limited use	Transactional	TALX (25% of U.S. work force), ChoicePoint
	Drug-test results	Private	Transactional	
Real property	Property deed	Public	PII	ChoicePoint

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Real property (cont.)	Property ownerships	Public	PII	ChoicePoint
Travel and transportation	Air travel itineraries	Private	PII and transactional	Galileo, Sabre, WorldSpan, and Amadeus
	Airport parking license plates	For limited use	Transactional	
	Buses	For limited use	PII and transactional	
	Cab pick-up requests	For limited use	PII and transactional	
	Car rentals	For limited use	PII and transactional	Galileo, Sabre, WorldSpan, and Amadeus
	Cruise ship reservations	For limited use	PII and transactional	Galileo, Sabre, WorldSpan, and Amadeus
	Hotel reservations, check-ins, and folios	For limited use	PII and transactional	
	Intersection traffic vehicles	For limited use	Transactional	
	Parking privileges	For limited use	Transactional	
	Toll road auto-pay enrollments	For limited use	PII	
	Toll road auto-pay transactions	For limited use	Transactional	Galileo, Sabre, WorldSpan, and Amadeus
	Train itineraries	For limited use	PII and transactional	
Utilities	Beeper/pager subscribers	For limited use	PII	
	Cable customers	For limited use	PII	
	Garbage-collection customers	For limited use	PII	InfoUSA
	Phone books	For purchase	PII	Qsent
	Phone subscribers	For limited use	PII	
	Power customers	For limited use	PII	
	Water and sewer connections	For limited use	PII	List brokers
Work force	Airframe/power plant mechanics	Free	PII	TSA
	Airline employees	For limited use	PII	

DATA SOURCE	RECORD	DOMAIN	CLASS	SOME ORGANIZATIONS WITH CENTRALIZED ACCESS
Work force (cont.)	Airport workers	For limited use	PII	TSA
	Bridge workers	For limited use	PII	
	Dam workers	For limited use	PII	
	Defense decision-makers	For purchase	PII	List brokers
	Benefits records	For limited use	Transactional	
	W-4s	For limited use	PII	IRS
	Government officials	For purchase	PII	List brokers
	Power-plant workers	For limited use	PII	
	Public servants	For purchase	PII	List brokers
	Shipping owners, operators, and managers	For purchase	PII	List brokers
	Port workers	For limited use	PII	TSA

Appendix I

Government Requests for Private Sector Data: An Informal Survey

by Mary DeRosa

Introduction

We conducted interviews with personnel from a variety of companies from which the federal government seeks information to fight terrorism. The purpose of this informal survey was to get a sense of what kinds of customer data the government currently seeks for national security reasons, how it seeks that information, and some of the issues the private sector sees with government use of its data.

We spoke primarily with chief security officers (CSOs) and legal personnel from large corporations. The people we interviewed were very helpful, but, for a variety of reasons, most requested that their company not be identified here. Therefore, this document identifies companies only by their industry. Among the companies—all of which are major players in their industries—were the following: (1.) a credit, debit, and other payment-card company; (2.) a bank; (3.) a manufacturer and significant government contractor in the national security area; (4.) an internet service provider (ISP); (5.) a producer of agricultural products; (6.) an insurance and financial-services corporation; (7.) a chemical company; (8.) a pharmaceutical company; (9.) a telecommunications provider; (10.) a transportation and consumer-service company; and (11.) a data aggregation company. We also interviewed a person familiar with the information the Transportation Security Administration (TSA) seeks from commercial airlines.

Six significant observations

1. Requests from the government for company data are narrow and specific, usually by the name of an individual

With the exception of the data aggregation company (discussed below), the company representatives reported that government requests for information are almost always for discrete record checks. The government does not request broader, pattern-based data inquiries. All company representatives reported that the government sometimes asks for checks of employee records for specific names. The personnel from the credit card company, telecommunications provider, bank, insurance company, and ISP all reported that requests for customer data are almost invariably to provide particular names or accounts and request information about account status, account activity, or transactions. The consumer service company is asked to track to whom it provided a specific service and what that service was.

The requests are somewhat different for companies that do not have individuals as customers, but they are still narrow. The agricultural-products company, for example, receives requests for shipment information: where a shipment is going, to whom, and what is in it. The chemical company is asked whether it has sold specific products to named companies or to customers in certain locations.

One departure from this pattern of narrow requests is with the government contractor. That CSO reported a somewhat broader request from a Department of Justice–sponsored review being conducted by the Defense Criminal Investigative Service (DCIS). For this review, the DCIS has asked the contractor to review its visitor records, which include information about vendor employees and others who have visited the facility, to determine whether all visitors are legitimate and authorized to be in an area where classified work is being conducted. The visitor-record information includes social security numbers or passport and nationality information. If the review turns up any unauthorized individuals, the contractor provides that information to the DCIS.

Another significant exception to the practice of narrow government requests for customer information is with the commercial airlines. Currently, the airlines and global distribution services (GDSs)—clearinghouses for travel records—conduct searches of passenger records and determine a risk score for each passenger, based on a calculation that the government has provided. The airlines and GDSs do not share passenger information with the government in this process. If the TSA implements the second Computer Assisted Passenger Prescreening (CAPPS II) program, however, the TSA will obtain passenger name–record data from the airlines and GDSs. The

TSA will provide some of this data to data aggregators, which will authenticate the passenger identities and provide a score to the TSA that indicates the degree of certainty about the identity. The TSA will then screen the passenger against government databases and will assign a risk level to each passenger. At the end of the process, both the TSA and the data aggregation company will discard the passenger information.

2. Companies provide most information to the government pursuant to a subpoena or other legal process

Most of the companies reported that they demand a subpoena or other legal document before they will provide private or customer information to the government, even when the information is requested for counterterrorism reasons. The telecommunications company, for example, always demands a subpoena before turning over customer information, and a court order for a wiretap. Similarly, the ISP demands a subpoena for member-identity and account information, a court order for transactional data (such as information about with whom a customer is communicating or the customer's online activities), and a Title III court order for the content of communications. According to the representative we interviewed, the only time the ISP will provide information voluntarily is if the government informs the provider of exigent circumstances, such as when lives are in danger.

Some of the companies expressed a willingness to provide information voluntarily on terrorism-related inquiries. The chemical company's CSO, for example, said the company has decided to be a "good corporate citizen" and provide information voluntarily for national security reasons as long as the request is "legal, ethical, and moral." In practice, that typically means that the company will answer questions voluntarily, but if documentation is requested, it will ask for a subpoena. The consumer-service company has made a decision that it will voluntarily provide data—including customer-database information—for homeland security investigations. For normal criminal matters, however, the company demands a subpoena. The government contractor generally provides information voluntarily about visitors, employees, or suspicious activities. It is rarely asked for customer information, but requires a National Security Letter (NSL)—an administrative subpoena that the FBI can use in national security matters—before providing it or when asked to conduct a covert search of employee property. Some CSOs conceded that there could be more information provided informally to law enforcement by security personnel in local offices, who often have law enforcement backgrounds and close relationships with law enforcement personnel.

The situation for the financial companies is somewhat different. They are required by law and regulation to provide a significant amount of customer information, such as Suspicious Activity Reports (SARs), automatically to the Treasury Department's Financial Crimes Enforcement Network (FinCEN). In addition to that information, the CSO of the bank reported that the government frequently makes 314(a) requests for information, which seek information about any listed individuals. These search requests are based on section 314(a) of the USA PATRIOT Act, which provides that law enforcement agencies may, through the Treasury Department, obtain information from financial institutions about identified individuals or entities suspected of terrorism. The bank will search its records and provide the government a "yes" or "no" answer voluntarily, but will require a subpoena or NSL before turning over any documents. The insurance and financial-services company will provide broker-dealer information to the government voluntarily in terrorism cases, but requires a subpoena for any credit or debit card information. The credit card company provides information voluntarily about whether a card is good and about the bank that issued it. For private customer information, the company will require a subpoena.

With the exception of the employee of the data aggregator (discussed below) and the person familiar with airline practices, none of the people we interviewed was aware of special arrangements to protect the privacy or accuracy of information they provide to the government.

3. Some companies conduct their own internal programs to detect terrorist activity

A few of the people we interviewed described programs their companies have implemented or are implementing to conduct broader, pattern-based searches designed to uncover terrorist activity. Two of the financial companies described the activity as related to the Know Your Customer rules that the USA PATRIOT Act and subsequent regulations have required them to implement. Know Your Customer rules require financial institutions to adopt customer-identification programs that verify customer identities and can check them against those of known or suspected terrorists. The credit card company is looking at refining the data mining that it conducts for fraud detection in order to assist with Know Your Customer compliance.

The bank described a sophisticated terrorist-financing detection program that it has created to look for indicia lof terrorist activity in its accounts. This program's software receives information generated from the Know Your Customer rules and from SARs and does data mining to look for patterns of terrorist-financing activity. The motivation for the program is the strong desire of the bank to be disassociated from terrorist groups. The bank views the program as an outgrowth of the Know Your Customer rules; it simply goes one step further to do something about the information it collects pursuant to those rules.

The government-contractor representative also described an internal surveillance detection program the company instituted to detect possible terrorist surveillance of its facilities. The company keeps records of suspicious activities in and around its facilities in a database. It does pattern analysis of this database to find any correlations that could suggest terrorist surveillance. For example, if one security official at a plant sees people in a blue Ford van taking video footage, he will enter that information in the database. The detection program will then look for incidents with similarities, such as other blue Ford vans, the same license plates, or the same behavior. If the detection program finds suspicious correlations, the contractor will provide the information to law enforcement.

4. Companies generally do not find terrorism-related requests from the government to be burdensome

We asked the companies whether government requests for their information for counterterrorism purposes pose a burden. They all answered that the requests are not burdensome. Each company representative described a significant upswing in requests for name checks from the government immediately after September 11, when the FBI was investigating the September 11 attacks. Companies that do not have individuals as customers, like the government contractor, the chemical company, the pharmaceutical company, and the agricultural-products company, were asked to search employee databases for the names. Other companies were asked to search customer records as well. The credit card company CSO said that because of the intense interest in credit card information immediately after September 11, it set up a special coordinating group to facilitate provision of information from issuing banks to the FBI. The pharmaceutical company, in addition to having received requests for employee information, had a great deal of interaction with the FBI in late 2001 about the anthrax investigation.

After this post–September 11 escalation, all companies reported that there was a decline in government information requests related to terrorism. For some, these requests have returned almost to the pre–September 11 level. For most, they remain somewhat higher but are not a burden. The ISP, for instance, still sees an increased volume of requests, and more requests for real-time information about communications, than before September 11.

The financial companies we looked at have increased reporting requirements since passage of the USA PATRIOT Act. This is especially true of the insurance and financial services company, whose brokers and dealers were not required to submit SARs to FinCEN before the PATRIOT Act. The company did submit some reports voluntarily, but its overall reporting has increased tenfold.

5. Companies complain of inadequate information-sharing and a lack of coordination of government requests

The company representatives we spoke with almost universally noted a failure on the part of the federal government to provide information to the private sector. They complained about what they called one-way-street exchanges, in which they provide information in response to requests, but hear only the most general information about the reasons for the requests and, more importantly, receive no follow-up information. The financial companies, in particular, expressed frustration with FinCEN and the FBI. The bank and financial-services companies both noted that they would find information about trends, patterns, and red flags in terrorist financing to be extremely helpful to their efforts to look for suspicious activity. They argue that because they know their business better than the federal government does, they could be helpful to the government if they were brought into the process a little more. The bank CSO noted that the FBI's terrorism-financing section began having meetings with the banking industry to improve information-sharing, but these have fallen off.

Threat information, in particular, is criticized as vague and generally of little use. As one CSO put it, "If you're just going to tell me there is a 'threat to your sector in the U.S.,' don't bother. I can't do anything with that." These officials are told that they cannot receive more specific warnings from the government in part because the information is classified. Therefore, several of the company officials focused on what they consider to be a need for more personnel with security clearances. They all noted that the federal government is unwilling to support an increase in security clearances for private sector personnel. One CSO who has clearance and access to classified information, and who is a former federal government official, noted that much of the classified information he sees is, in his view, "not that sensitive" and should not be classified.

Another common view of those we interviewed is that the federal government needs to coordinate better its approach to the private sector. Several people sense an increase in competition among federal agencies since the Department of Homeland Security (DHS) opened its doors, particularly in the financial area. Several bemoaned a lack of a single point of contact in the federal government, or even just a few. The agricultural-products company is receiving overlapping requests willy nilly from an increasing number of agencies. The chemical company CSO "would like not to be asked the same question by five different agencies." Several CSOs believe the DHS is the appropriate solution to the coordination problem. But they do not see the DHS taking on this role. As the representative from the pharmaceutical company commented, the DHS has been "slow coming out of the gate" and has not even taken the first step of developing a list of contacts in key industries.

6. The federal government is using data aggregation companies to perform broader, national security–related searches

Although the federal government is not making broad requests for searches of the databases of the corporations discussed above, it is doing that kind of search using the services of data aggregation companies. Data aggregation companies collect information that is, for the most part, publicly available. What they do is bring together information from thousands of sources and make it searchable. The types of information collected by the company we looked at, according to the company's representative, are listed below.

TYPES OF INFORMATION CONTAINED IN DATA AGGREGATION DATABASES

1. Public records, such as courthouse records, real estate records, tax liens, judgments, business-related information from secretaries of state, professional licenses, and some Department of Motor Vehicles (DMV) records

   Some states do not allow sale of, or access to, DMV records, but approximately 28 states do allow at least limited access. For example, the states will allow access to the records for law enforcement purposes. The data aggregator can then search these records only for clients that have legitimate access.

2. Other publicly available information, such as White Pages information on the Internet

3. Nonpublicly available information, such as "credit header" data

   Data aggregators do not have access to entire credit reports, but they can collect the header information, which usually includes name, social security number, address, and phone number. Access to this information is restricted under the Gramm-Leach-Bliley Act, but that law allows some market access for specific purposes, such as fraud detection and law enforcement. Again, the data aggregation company must determine the purpose for the search before it can be conducted.

Prior to September 11, most government queries to data aggregation companies were discrete: The government would request searches on specific names or address. Since September 11, there has been a significant increase in interest from many national security agencies in providing large quantities of information to data aggregation companies and quickly having that information analyzed for links to other relevant information. One national security agency, for example, is providing a stream of data on possible terrorists to the data aggregation company we looked at. The company uses Extensible Markup Language (XML) technology to identify those individuals and provide links to other people and organizations. The information returned is the raw data with the links that the agency can integrate into its existing system for further analysis. Another agency has requested that the company conduct link analysis on subjects and notify the agency of noteworthy links. One other example involves an agency that provides a long list of names to the aggregation company and asks the company to check regularly on the status of those people and alert the agency of changes of address, etc.

The data aggregation representative we interviewed identified the filtering of false positives as one of the biggest challenges in conducting these searches for the government.

There are some errors in the public data or credit-bureau data, such as incorrect addresses or transposed digits in social security numbers. The aggregation company has a number of software-based methods for identifying these errors. When an aggregator identifies a possible false positive, for example, he or she flags it for the government and can discard it from the output the company provides. The aggregator does not, however, correct the original data.

Appendix J

Data Analytics Practices of the Private Sector

by James X. Dempsey and Lara Flint

Introduction

In considering how the government could make better use of information technology for counterterrorism purposes, our Task Force thought it would be useful to consider how the private sector uses data for identity verification, risk assessment, and related purposes. To this end, the Center for Democracy and Technology (CDT) held discussions with representatives of companies and government agencies involved in data analytics. Specifically, CDT consulted with representatives from four leading companies—Acxiom, IDAnalytics, JP Morgan Chase, and SRD—to find out how their companies make use of data analytics. Those representatives explained their companies' technologies as follows: (1.) Acxiom uses a data-matching and identity-verification methodology; (2.) IDAnalytics is developing a fraud-identification system in order to identify fraudulent credit applications before they are accepted; (3.) JP Morgan Chase uses data analytics to identify fraudulent transactions within its customer base; and (4.) SRD uses data analytics to identify relationships among individuals.

Based on the presentations by the four companies' representatives and the ensuing dialogue, we came up with some preliminary conclusions and questions that might help inform the debate surrounding government use of data-analytics techniques on large databases—public and private—for law enforcement and intelligence purposes.

Conclusions about the use of data analytics in the private sector

1. **Data matching or retrieval on a name-only basis is very difficult—even worthless in many contexts.**
   Data matching or retrieval on a name-only basis is difficult because commercial data analysis is usually done on the basis of two identifiers (name and address, at a minimum), or on the basis of other identifiers.

2. **Effective commercial data-analytic techniques rely on the establishment—and maintenance—of a set of good, or true, identifiers.**
   There are effective techniques used in the private sector for recognizing that two sets of information pertain to the same individual. Their success depends on the accuracy of the information in the databases against which a particular set of personal identifiers is matched (for example, a name and address). Private companies determined several years ago that the most effective way to match data was to build verified reference tables. These tables consist of personally identifying data that the company is (nearly) certain is accurate. New information can be compared with the repository of verified information to determine whether they match. Thus, for example, if a new name and address combination is presented for review, it is possible to evaluate how close the representation of the name is to names already known to be associated with that address. This methodology results in a more accurate match than does matching two sets of data without initially evaluating the accuracy of at least one of the data sets. In short, it is not effective to run one set of data against another to look for matches without first evaluating the accuracy and completeness of at least one of the sets of data.

3. **Identity theft has complicated the process of data analytics.**
   Individuals are no longer the only holders of complete and accurate data on themselves. The ever-growing problem of identity fraud makes the usefulness of a repository of verified "known" information uncertain because fraudsters are now able to access the full panoply of identifiers about other people. As a result, in the attempt to identify fraudulent applications for credit or insurance, it is increasingly less reliable to compare the information provided by an applicant with known "true" information because the fraudulent applicant will have the same accurate information about the individual whose identity he or she is assuming as the private company seeking to verify that identity.

4. **Methodologies are being developed to detect fraudulent credit applications that contain accurate information and those for which there is no match to a known fraudulent record.**
These methodologies are based on millions of case examples from financial institutions, cell phone companies, etc., at which it has been possible to track the record of applications to find which ones are fraudulent. The ability to track these records has allowed analysts to determine the predictive patterns that the data analytic technology must find. The development of sophisticated methodologies such as these depends on a company's ability to accumulate information from fairly controlled environments and from many transactions. The technology is consortium-based (it requires that a variety of industries provide information), so that patterns can be identified. In the context of credit card fraud, this approach to predicting bad behavior requires a very large sample of other similar bad behaviors against which an individual's current behavior can be compared.

5. **One of the best technologies for identifying known fraudsters is based on voluntarily provided information.**
An effective system for protecting businesses from individuals with known undesirable backgrounds relies in large part on information that is voluntarily provided to the employees who conduct screening processes. For example, a representative of a technology company told us about a method of detecting fraud at a casino. The casino collects data from vendors, employees, hotel guests, and others who voluntarily provide that data in the course of filling out a job application or hotel registration. The casino's data-analytics technology then helps to root out employees, vendors, and guests who have connections to known fraudsters—in this case, people on the gaming commission's blacklist of persons with a record of fraud—by finding relationships among the data.

6. **In attempting to identify individuals who pose risks, it is more effective to identify those who have relationships with known fraudsters than to try to predict patterns of suspicious behavior.**
The tracking system described above uses information about vendors, employees, hotel guests, and others, to determine who might pose a risk or have a relationship with gaming felons. Such a system could also help to determine individuals who might have relationships with known terrorists. Moreover, this relationship-awareness approach can be more effective in identifying risks than that of attempting to predict suspicious patterns of behavior. For example, one company's research in pattern-recognition technology showed that attempting to identify risks through pattern analysis resulted in such an overload of statistically interesting leads that would need to be investigated, that it became impossible to prioritize them, much less investigate them. Essentially, the overload of information that results from looking for patterns renders the analysis useless. This suggests that attempting to locate patterns of behavior indicating the planning of a terrorist attack would result in huge numbers of false positives and false negatives and would not be useful.

7. **As a practical matter, watch list fidelity (its accuracy and completeness) is one of the biggest challenges faced when attempting to identify risks.**
If a watch list contains inaccurate or incomplete data, it will be very difficult to compare data against that list. In particular, as stated earlier, name-only matches are meaningless because more information than simply a name is necessary to determine whether an individual is, in fact, the person listed. In terms of the government's use of data, this suggests that watch lists need to be verified to ensure they are accurate, complete, and up-to-date. This is particularly important if watch lists are to become the centerpiece of a system that seeks to identify those who have relationships with known terrorists.

8. **Anonymization of watch list data may be possible for purposes of comparing a list to private sector information.**
Technology is being developed that would allow the government to provide an essentially unreadable version of a watch list to commercial entities, who could check the watch list against their information without actually learning what information is on the watch list. If a match is found, the government would be notified that the commercial entity has some relevant information. The government, in turn, could obtain that information directly from the commercial entity (using appropriate legal processes). During the anonymized matching process, the commercial entity would not know whether there had been a hit and ultimately would not necessarily need to know whether the government was seeking further information on an individual as a result of the particular search.

Questions raised by our conclusions

The data-analytics practices of the private sector seem to depend heavily on matching name and address (plus other identifiers in some contexts). Some questions revolving around this issue and the conclusions above are as follows:

1. What is the quality of the name and address information available to the government?
2. Is this information available on short notice to government agencies?
3. Is it possible to determine the most useful data for detecting fraud?
4. How are broader categories of data used (for example, purchasing records or travel information)? And would these broader sets of data prove useful for fraud detection? And for risk analysis?

MARKLE FOUNDATION

Markle Foundation
10 Rockefeller Plaza
16th Floor
New York, NY 10020-1903
Phone 212.713.7600
Fax 212.765.9690
www.markle.org