

TESTIMONY OF JUDITH A. MILLER
BEFORE THE
NATIONAL COMMISSION ON TERRORIST ATTACKS UPON THE UNITED STATES

December 8, 2003

Mr. Chairman, Mr. Vice Chairman, and members of the Commission

Thank you for inviting me here to talk about how the government can protect individual privacy while preventing terrorism. You have asked me to discuss the use of advanced technological tools that aggregate and analyze vast data sets of personal information in the fight against terrorism. In particular, I will talk about the benefits these techniques bring to the fight against terror, the risks to individual privacy from their use, and how we can develop policy that addresses these issues.

Technology Tools and Why We Need Them

I will begin by saying that technology tools that assist with collection, sharing, and use of information have the potential to be enormously useful in the fight against terrorism. We ignore or reject them at our peril because we must exploit all advantages. As members of this Commission understand perhaps better than anyone, the challenge of protecting ourselves against new threats, like terrorism, is very different from the cold war challenge. Then, we could look – mostly overseas – for a relatively few rich sources of information on our adversary and secure our advantage by holding that information very closely. Now, we may need to look everywhere for clues to terrorist plans and behavior, including at home. Many of these clues will be in databases containing private information, some in the private sector. We cannot ignore these sources. Immigration data, watch lists, aggregations of public records like Department of Motor Vehicle records and even White Page data, criminal records, transactional data from private companies, all of these things could contain vital information that will assist in identifying terrorists or their plans or methods. Use of technology to search and analyze this data in responsible ways is a critical tool for addressing our new intelligence needs.

Techniques like data aggregation and data mining or data analysis are the tools we can use to make sense of, bring order to, and discover new information from these vast data sources. I will take a moment to explain the terminology I am using because these terms mean different things to different people.

Data aggregation or integration is standardizing data to make it useful for sharing, searching, and analysis, regardless of how the data is structured. When we talk about aggregating data, it does not necessarily mean collecting it all together in one database. In fact, “aggregated” data is often distributed in a number of databases, but identified and

accessible for searching. Aggregation or integration is necessary for data mining or data analysis, but it is not the same thing.

Data Analysis or Data Mining is using automated tools and models to discover new things from mass aggregations of data. Most technical people use the term “data mining” narrowly to refer only to the process of finding the tools and models. “Data analysis” is the more accurate, broader term. The purpose of data analysis is to turn masses of data into something usable. It can find links, uncover patterns, or even predict behaviors. Data analysis can include **subject-based** or **pattern-based** analysis.

Subject-based analysis starts with a known person, place, or thing and learns more about it and its links, direct and indirect, to other data. For example, if the government has a name of a suspect or a prospective employee, it can query its own watch lists and other databases or go to a commercial aggregator to query publicly available records to find out whether the person is linked to any known terrorists. This type of subject-based “link analysis” can be performed now and provides invaluable leads and clues that could be used for further investigation and analysis.

Pattern-based analysis is more complex. This analysis looks for patterns in data that predict certain behavior. These patterns can be discovered from data mining or from other methods such as red-teaming or intelligence. To use a simplistic example, we might know from intelligence and studying terrorist behavior that terrorists will rent a car, purchase a cell phone, buy explosives, and buy a one-way train ticket while preparing for an attack. Pattern analysis might be used to search databases for clues to people engaging in this pattern of activity. This is generally the kind of research DARPA was pursuing with its TIA project. There are many hurdles and significant research would have to be done before you could even tell whether it would work in the counter terrorism context, but the potential is obvious.

It is important that none of these techniques be seen as a complete solution to any intelligence or law enforcement problem. Data analysis alone will not identify a terrorist or uncover a plot. It can, however, be an extremely useful tool that will assist analysts and investigators to find, understand, and follow up on clues to terrorist plans and activities.

The Privacy Challenge

Having said that it would be folly to eliminate useful technological tools from consideration, I must emphasize that these tools when used to access private data have the potential for abuse and harm to privacy. We must be systematic in developing new protections for privacy that address these challenges.

There are a number of ways in which data analysis technologies can cause harm to individuals. First, there is the chance that they will not work correctly and the government will mistakenly identify innocent people as terrorists. False positives are a problem in any search, and if the results of data analysis are used only as a starting point

for additional follow-up, this might not be a significant problem. But when we talk about terrorism, any piece of information is likely to be acted on immediately, and this can mean innocent people are inconvenienced at best, and at worst have their reputations and livelihoods permanently harmed.

Second, all databases have inaccurate data. How this data is corrected in the data analysis process is a major issue. Too often, inaccurate data has a life of its own. Even when corrected in one database, it remains in others. The technology for following and correcting all occurrences of inaccurate data lags far behind the technology for collecting and analyzing the data.

Third, controls on the data may be inadequate. By putting these tools in the hands of government employees we run the risk that they will be – purposefully or not – used too often and with inadequate justification. In addition, results of searches might be retained past when they are needed or disseminated to others for improper reasons.

Finally, a related concern is “mission creep.” Many people believe, like me, that the fight against catastrophic terrorism is important enough to justify the use of new and powerful tools – even if they allow access to private information. But because these tools are justified to fight terrorism does not mean they should be used for all government activities. Once we have these tools, there will be an enormous pull to use them for purposes other than terrorism, but the balance of potential benefit to potential harm might be quite different, for example, for terrorism and bank robbery. There is a real risk that once these tools are in the door, they will be overused and privacy will suffer significantly.

Protecting Privacy While Using These Tools

Because of this potential for harm to individuals, the government must take systematic steps to protect privacy in the use of these tools. I believe these steps must come as part of an overall review and revision of our approach to privacy protection. In this connection, I’d like to draw your attention to the work of the Markle Foundation Task Force on National Security in the Information Age, which Zoe Baird and James Barksdale chair and of which I am a member. The Markle Task Force has been working for over a year and a half on issues like the one you have asked me to address today. We issued a report last October that advocated use of advanced information technology and networking tools in homeland security, but discussed the need for the government to adopt explicit guidelines on how to use these tools responsibly and in a way that protects privacy and individual liberties. Just last week we issued our second report, “Creating a Trusted Information Network for Homeland Security,” which digs deeper into these issues and discusses what specific steps must be taken to share information more effectively, what technologies we should be exploring to improve the fight against terrorism, and, very significantly, how the government can protect the privacy of U.S. persons when it uses technology to access private data in the fight against terrorism.

Implement Guidelines for the Use of the Technology. One of the principal recommendations of the Markle reports is that the government must implement guidelines for the use of private data, particularly with new technological tools. Current law and policy provide almost no guidance to workers about how and when they may collect and use private data. If the government is to use data analysis and other technology that allows access to private data, government employees must have consistent, clear guidelines on how these technologies and the information they produce can be used. These should include guidelines on:

- Relevance. For what reasons may these technologies be employed? What approval must workers obtain before using them? If terrorism is the reason we need the technology, then terrorism should be the reason to use it, not other crimes. The guidelines should also make clear what kind of showing or approval the employee needs to make or obtain in order to conduct searches of private data. In some cases approval from a court will be required, in others only approval of a supervisor. Some less sensitive uses should not require any advance approval, only after-the-fact reporting or review.
- Retention. How long should the information be retained? We should not default to retaining private information indefinitely; it should be kept only for as long as necessary to carry out the purpose for which it was collected. Indeed, there should be a preference for not retaining information at all if it comes from databases outside of the government.
- Dissemination. To whom, and for what reasons, can the data be disseminated? I believe the strong preference should be not to disseminate private information collected for counter terrorism purposes to others in the government to be used for other purposes.
- Reliability. How can information determined to be inaccurate be changed? How can a person affected by inaccurate information be certain that records are corrected?

Improve Oversight. Along with these guidelines must come reinvigorated executive branch oversight; it is the Executive Branch's responsibility to ensure that these guidelines are understood and followed. The Executive Branch must commit to rigorous training on the guidelines for all employees who might use private data. In addition, it must institute regular audit and review procedures to see that the guidelines are being followed. Oversight too often means only after-the-fact investigation of errors or abuses. It is critical for oversight to do more than this: it must ensure that government employees are on the right track, that they understand what they are supposed to do and are doing it. Periodic review and audits designed to keep employees on track will not only protect against abuse, but they will help avoid the timidity we sometimes see in employees who do not really understand the lines they are supposed to draw, but know that if they get it wrong they might be criticized, investigated, or worse.

Review Risks and Benefits Before Adoption. There must be a government-wide approach to the implementation of data analysis tools, which should include a thorough – and, to the greatest extent possible, public – examination of the potential benefits and the risks to privacy and civil liberties before implementation. I do not think there is anyone – inside or outside of the government – who could tell you about all the current uses of data analysis, or even most of them. The current approach to these tools in the executive branch is ad hoc and lacks transparency. Use of these powerful tools requires a more systematic and open discussion. Before adoption of significant data analysis programs, the executive branch should be able to demonstrate that the technology and use of private data is genuinely important to security and it will be used in a way that minimizes its impact on privacy. In fact, the current ad hoc and somewhat secretive approach to the use of these tools is much of what has caused suspicion and backlash about their use. A more open and systematic approach would give the public and the Congress more comfort that privacy issues are being considered and accommodated in the use of these techniques.

Use Technology to Advance Privacy. One very promising avenue that the government must pursue energetically is the use of technological tools to protect privacy. There is a significant amount of work going on now in the private sector and academia to develop technology that anonymizes data; controls access to databases; and facilitates audits of database use. The research must be encouraged and supported. Although I do not believe technology will ever provide the complete solution to the privacy problems, it can go a long way in furthering the goals of the guidelines. Government supported research on data mining and data analysis should always include research on how to protect privacy when using those techniques. I will note that the research DARPA was supporting with its TIA program included significant research on privacy protection. With the demise of TIA, the future of that research is uncertain, which is unfortunate.

Conclusion

I applaud the Commission for examining these important issues. Understanding of the issues and leadership on the part of the executive and legislative branches are critical to resolving the challenge of preserving privacy while fighting terrorism. I believe this Commission has a vital role to play in increasing understanding and encouraging leadership.

I thank you very much for the opportunity to testify.