**Judith A. Miller**
Partner, Williams & Connolly, LLP

**House Permanent Select Committee on Intelligence**

October 30, 2003

**"INTELLIGENCE COLLECTION AND CIVIL LIBERTIES: Technology and Privacy in Intelligence Collection"**

Thank you for inviting me here to discuss the very important subject of intelligence collection and civil liberties. You have asked me to focus in particular on the use of data aggregation and data mining technology and the opportunities and challenges that technology presents. This is a very important topic and one that, unfortunately, has so far inspired more emotion than rational, informed discussion.

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 them or reject them at our peril because we should be exploring all possible advantages. But we must also recognize that some of these tools, if used domestically, present challenges to the privacy and civil liberties of Americans. Our current system of privacy protections is not up to the task of protecting our privacy and liberties in the new intelligence environment. The answer, though, is not to forgo the technology; it is to put every bit as much energy, attention, and creativity into finding mechanisms to protect privacy as we do into finding technologies to protect our security.

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. The Markle Task Force will be issuing another report in a little over a month that will dig deeper into these issues, including discussing what 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. Another important source of analysis in this area is CSIS' Roundtable series on Data Mining.

**Technology Tools and Why We Need Them**

As we all know by now, 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. 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 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 to address our new intelligence needs.

I will talk briefly about some of the technology and processes that come under the "data mining" or "data analysis" label.

- ***Data Aggregation or Integration*** is making data available 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. Data aggregation is critical for homeland security because players from the federal government, state and local governments, and the private sector will all have relevant information. 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 to make sense of mass aggregations of data. "Data mining" is the term we hear most often, but its technical meaning is actually narrower than its common use. "Data analysis" is the more accurate, broader term. The purpose of data analysis is to turn masses of data into something usable. It can summarize data, find links, uncover patterns, or even predict behaviors. Types of data analysis include:

    - ***Subject-based analysis.*** That is, if you know a particular person, place, or thing, you can do analysis to learn 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 or go to a commercial aggregator to query publicly available records to find out whether the person is linked to any known terrorists. This "link analysis" will provide more information that can be used for further investigation.

    - ***Pattern-based analysis.*** This is more complex. Automated analysis can be used to find significant patterns of behavior in data and then construct models from those patterns that will predict the same behavior. This is

what the credit card companies do when ferreting out credit card fraud, and it is the technical definition of "data mining."

Another use of pattern-based searching, though, which is more likely to be useful in counter terrorism, is to search for patterns in data based on predictive models that are found elsewhere. 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, but the potential is obvious.

## The Privacy Challenge

Having said that it would be folly to eliminate useful technological tools from consideration, I cannot emphasize enough 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:

- ***False Positives***. One of the most significant concerns with data analysis using private information is that it will not work correctly and the government will end up mistakenly identifying 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.

- ***Inaccurate Data, Failure to Update.*** 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.

- ***Inadequate Controls.*** Few would argue with the proposition 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. 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.

- ***Lack of Clarity on Purpose for Use.*** A related concern is "mission creep." 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.

Because of this potential for significant harm, we cannot simply begin to use these tools without first taking steps to protect privacy when they are used. There are several urgent policy and legal steps that the government must take. Most of these are steps the Executive Branch can take on its own; but if it does not, Congress should take action.

- **First, Review Risks and Benefits Before Adoption**. The government should adopt no new use of data analysis tools without a thorough – and, to the greatest extent possible, public – examination of the potential benefits and the risks to privacy and civil liberties. The government must 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. There should be an established, government-wide process for this review. Even with research into new data analysis and related technology, the government should build into the research consideration of the privacy issues. That way, researchers can – from the start – be looking for ways to incorporate features into new systems or technology that will assist in privacy protection.

- **Second, Implement Guidelines for Use of the Technology**. 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?

- **Third, 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.

- **Fourth, Use Technology to Advance Privacy**. Technology can be an extremely important tool for protecting privacy. The government must invest in and employ technology that furthers the goals of the guidelines. This includes technology that anonymizes data; controls access to databases; and facilitates audits of database use. There is a lot of very interesting research going on right now – Teresa Lunt at the Palo Alto Research Center, Dr. Latanya Sweeney at Carnegie Mellon, and Jeff Jonas at Systems Research and Development, who has been working with the Markle Task Force – are all doing some very good work.

## CONCLUSION

This Committee has a very important role to play in seeing that there is rational, informed discussion of the use of technology in intelligence collection and the important civil liberties and privacy issues that it raises. As I have said, I do not believe all of the steps I propose require legislation, but Congress should ask the right questions and require immediate action from the Executive Branch on these issues.

I thank you very much for the opportunity to testify on these critical issues.