The Privacy Office
Department of Homeland Security
Privacy and Technology Workshop:
Exploring Government Use of Commercial Data for Homeland Security
September 8-9, 2005

OFFICIAL WORKSHOP TRANSCRIPT

Friday, September 9, 2005
Auditorium
GSA Regional Headquarters Building
7th and D Streets, SW,
Washington, D.C., 20024

PANEL FIVE
HOW CAN WE BUILD PRIVACY PROTECTIONS INTO THE GOVERNMENT'S USE OF COMMERCIAL DATA – RECOMMENDING A ROADMAP FOR DHS?

Moderator:
Ms. Toby Milgrom Levin

Panelists:
Mr. Steve Adler
Mr. Fred Cate
Mr. Jim Dempsey
Ms. Mary DeRosa
Mr. Chris Hoofnagle
Mr. Martin Smith

MS. LEVIN: My name is Toby Levin. I'm Senior Advisor in the Privacy Office. It's my pleasure to moderate this last panel, which I challenged to provide the Department a roadmap for how we should consider the use of commercial data and what protections should be applied.

I want to start by first saying that the statutory mission of the Privacy Office is to assure that technologies sustain and do not erode privacy protections relating to the use, collection, and disclosure of personal information, and to assure that personal information that DHS uses is handled in full compliance with the Fair Information Practice Principles. I'm happy to be able to say that those are the guiding principles for the Privacy Office and for this Department.

Building on what we've learned from the prior panels, we will use this opportunity to tie together what we've discussed the last day and a half and come up with a recommended roadmap for DHS to consider. Let me start, first, by quickly introducing the panelists. Their backgrounds are extremely rich and diverse. I urge you to review their bios in your packet. To my right, we have Steve Adler, Program Director of IBM Data Governance Solutions; Fred Cate, Professor and Director of the Center for Applied Cybersecurity Research, at Indiana University; and, to my left, Mary DeRosa, Senior Fellow, the Center for Strategic and International Studies, and coming down the ramp is Jim Dempsey – who has moved to California, where things are much more laid back than on the East Coast – Jim is Executive Director, Center for Democracy and Technology; Chris Hoofnagle, Director of EPIC West -- we welcome him back to the East Coast today; and Martin Smith, who's the DHS Program Manager for IT Information. Unfortunately, Larry Ponemon could not join us today.

So, we have quite an array of expertise here, and I want to start with the first question -- Is there a framework that we can use to help build a roadmap for DHS to guide its use of commercial data? Jim, we'll throw the first softball to you.

MR. DEMPSEY: Thanks. Yes, there obviously is a framework. I mean there is nothing new under the sun. The Fair Information Principles, which were developed in the 1970s, remain the framework. They are now embodied in the Privacy Act and, to a greater or lesser degree, in the various sectoral privacy legislation that we have for the private sector. We have heard criticism, over the past day and a half of the Privacy Act, justifiable criticism. To some extent, it is outdated. The concept of system of records is no longer a technologically appropriate concept. The question of subsection M and the application of the Privacy Act to commercial databases that the government subscribes to remains somewhat of an open question and needs to be clarified. I think it can be clarified by OMB and by individual agencies, at their discretion, through their contracting. But, basically, the issues have been laid out.

Now, from time to time, the principles get mis-emphasized. Over the past three, four, five years, a lot of the privacy debate has been about notice. Fred Cate and others have written persuasively about the mis-emphasis on notice. You see a convergence of criticism from what I would very grossly call the "right" and the "left" about the inadequacy of a notice-based regime. Everybody who has gotten GLB financial privacy notices in the mail and thrown them away immediately knows the inadequacy of the undue focus upon notice. Notice remains a valid principle, but it is not the sole principle. In the past year, we have also seen an emphasis on the security principle. Custodians of data have an obligation to preserve its security. As a result of the California notice of breach law, we've seen multiple examples where custodians of data have failed in that fair

information principle, and now we see legislation pending in Congress that emphasizes the security principle.

However, I think that the other principles -- collection limitation -- collect no more information than is necessary for the purpose at hand; retention limitation; the quality principle; the accountability and enforcement principle -- these are equally important principles and need to be reflected. Now, applying them to the terrorism context poses challenges. Obviously, in a law enforcement and intelligence context you are not going to give advance notice to the bad guy that you are collecting information about him, you're not going to give him a choice about the collection of information. And the government has certain compulsory powers at its disposal. But it's, nevertheless, useful, as you go through each program and each application, to think about all of the principles. And that's what the privacy impact assessment is supposed to be about. That's what the statement of records notice process under the Privacy Act is supposed to be about. And we heard yesterday about whether people treat that seriously or whether they treat it as a checkbox simply to be gotten out of the way.

Now, I will say one final thing, and that is that the privacy advocates and the privacy officers are not Dr. No. They are not the people who say “no.” Sometimes they say no, of course. But that shouldn't be their only role. In fact, increasingly -- I think, partly because of an operational defect or operational failure on the part of operational agencies to clearly define their goals and missions -- the privacy officers play the important role of saying, "What is your goal? What is your mission? What is your objective here?" Because you can't even begin to answer the privacy questions until you have answered the operational question, "What are you trying to achieve?"

I think Secure Flight has been plagued by the inability of TSA to define, at any given time, what is the mission or goal of Secure Flight. Is it to check the terrorist watch list, is it to find unknowns, et cetera? I think you see privacy people saying, "You have not defined your mission."

Before I can start talking about a collection limitation and an accuracy principle, and a use limitation, and a retention limitation, I've got to know what you're trying to accomplish here. For example, what is the purpose of the DHS Operations Center? The DHS Operations Center published a notice of records, never really stating why they were collecting all this information. Given the fact that we have several other operations centers in the government, now including the DNI responsible for fugitive information and the NCTC. So, I think we've got to stop thinking of the concept of privacy as somehow at odds with security and then somehow at odds with achieving the mission. If you walk through these questions -- What are you collecting, for what purpose, and with whom are you sharing it ? -- those are both operational questions and privacy questions.

So, the framework that I see, the roadmap that I see is the Fair Information Principles -- a roadmap not only for privacy protection, but for operational success.

MS. LEVIN: Fred, do you want to add to that, please?

MR. CATE: First, I want to know why Toby's sitting so far away from the three of us. It's going to hurt my feelings. [Laughter.]

MR. CATE: On the whole, I think I agree with Jim, which I usually do. I do worry, though. I'm not nearly as sanguine as you are about the way in which Fair Information Practice Principles are carried out. I come at this, in part, because we have seen them typically used in very broad and vague and general ways, the same way I think the Privacy Act is being used. So, we say, "I want to collect information to fight terrorism and aid in homeland security, and that's my purpose." Well, fine, now what can I collect? Have I given any notice of any meaningful form at all?

A second reason I worry about FIPPs in this role, or even the Privacy Act, as an implementation of FIPPs. As you know, choice just doesn't play much role here. And, certainly, the experience we have with Fair Information Practice Principles has been one of notice and choice. We give notice primarily so that people can make choices about whether to engage in an activity -- whether to provide the information. You just do not have a lot of choice in the national security environment, and my guess is, that's not likely to change anytime soon.

The third reason is really the one with which you ended, and I think it's by far the most important, and that is, I really don't think you even get to the privacy issues until you answer the question of "Why are we doing this?" Why are data necessary to do this? Why are commercial data necessary to do this? Are we doing this in either the most effective or least restrictive manner possible, from a civil-rights point of view?

One of the things that I think's very troubling in these conversations -- it's troubled me throughout yesterday and this morning, and it troubles me more broadly -- is, we talk in these very broad terms about all of the things we might be able to do with commercial data. I am confident, 100 percent confident, that many of these would advance national security. The problem is, however, that I can't even begin to answer the FIPPs questions -- Is this necessary? Is this the least restrictive? Is this the least amount of data necessary? Is it retained for no longer than necessary? -- without some specific purposes on the table.

I want to just give one practical example of that. I'm sure each of you can think of dozens of others. If the primary reason, the reason stated by TSA, for wanting to collect commercial data for purposes of airline security, is to better match people to the wait list -- to the watch list -- wait list being something different in flying -- [Laughter.]

MR. CATE: -- That specific purpose certainly alters the calculus about what type of data and how much data we need. Commercial data is enormously valuable in matching. It's incredibly valuable. I think it's well demonstrated. Industry would not spend the billions it spends on those technologies if they did not enhance the accuracy of matching. Its accuracy, however, doesn't come from transferring a certain amount of data to the government; it comes from the richness of the data set, itself. You know, how many transactions have I watched this person go through over time so that I really know that this is the same individual?

Therefore, asking questions like, "How much data does the government need?" -- the government doesn't really need any data there at all. In fact, the exact opposite is what's needed. The government's data needs to go into those technologies to see "Who are these people?" On the other hand, if the government's data mainly consists of first and last names, or only last names, in the case of intercepted cell-phone calls that have suddenly resulted in someone going on a watch list, it doesn't matter how good the data are. You will never be able to match that data well. If I am just looking for Mr. Atta, just forget it. There is nothing that ChoicePoint or Acxiom or LexisNexis can sell me that's going to make it possible to match that data better. And so, if we're not clearer and more specific, not only in the privacy notices that agencies publish, but in our discussions about these things, we're really just going to be passing each other and not adding any light to all of this heat.

MS. LEVIN: Well, let me ask, as people interested in privacy, we're not saying no to the use of commercial data, but we're asking a series of questions in order to achieve a better end. So, let's start with one of the first questions -- Is collection of commercial data useful to Homeland Security's mission, and, if so, how? Does it meet accepted standards for data quality and integrity? Is it relevant, accurate, timely, and complete? Let's start with the collection principle. Chris?

MR. HOOFNAGLE: I think it's difficult to look at commercial data brokers and conclude that their data and practices are consistent with the Privacy Act. If you look at the companies, you'll see that collection limitation is generally out the window. Their request to courthouses and to other bodies is, "Give us all your data, give it on a tape, and don't redact anything, we'll sort out what we want and what we don't want." There are limits on use of the data that are affected. The commercial data broker's access to your data is not a right, it's a privilege; and it can be revoked. Not only that -- it's not clear that you're getting your whole profile, for instance, when you write to ChoicePoint and say, "Give me my file," which you can now do free, and I encourage you to do so. With ChoicePoint -- excuse me, ChoiceTrust.com, you fill out a form with your social security number, your address, and other information, and then they can do a very narrow database search that only collects data that they are pretty certain is about you. Then,

what's funny about it is, when you get that data back, it's not accurate. It's accurate when it's me, because I have a weird name. But when -- if you have a name that's common at all, you will be mixed with other people who live in your state, because companies like ChoicePoint stick data into a file based on your last name and state.

There is no data destruction when it comes to these files. Your ChoicePoint report will probably reach back to the earliest age in which you established credit. And, of course, there's little or no accountability outside of the FCRA context.

One thing that's really important to note is, the data brokers will often talk about how they are regulated, but when they are engaging in that discussion, they're only talking about certain products that fall under federal regulatory frameworks, like the Gramm-Leach-Bliley and the FCRA. They also have products that are not in these regulatory frameworks. Finally, if I could just impart two more points on this issue. I was working on a bill in California where the data brokers showed up in force to oppose it. Lexis actually sent out their corporate counsel in addition to their local lobbyists, to oppose this bill. One of the things the bill did, in its first iteration, was give you an audit trail of who was buying your report from companies like ChoicePoint. And it was actually this issue that aggravated the industry more than anything else. Industry came out with 30 different reasons why an audit trail would not be useful. "Well, we can't do it technically," they said. "Well, it wouldn't be useful to the consumer, anyway." "Well, it would confuse the consumer." There was just every excuse in the book.

If 1999 and 2000 were the year of notice, and 2001 was the year of opt-out, and 2004 was the year of security, I really hope -- and this is EPIC's focus – that future years will be about getting audit logs from these companies. I've done a lot of Freedom of Information Act work with commercial data brokers, and we have obtained thousands of pages of records from federal agencies. One of the things that's becoming clear is that there does not appear to be an effective audit trail on the very people who are using the system. When you see auditing discussed, it's often discussed in the context of not allowing people from other agencies to use your ChoicePoint account. So instead, someone from the DEA might call up the Marshall Service and say, "Hey, would you mind doing this search for me?" So, I think it's really hard to look at commercial data brokers and conclude that use of commercial data is consistent with the Privacy Act. Now, use of the data is probably useful for homeland security purposes. However, I think what's really important on that issue is to be scientific about it. You have to design studies that show that the data is effective for the purposes that you want to employ it. But, if you come out with one study that says, "Yes, commercial data is good for this purpose," that shouldn't open up the barn door to all the other uses of commercial data.

MS. LEVIN: Would anyone else like to comment? Fred?

MR. CATE: Again, I just think it's a perfect example of the type of difficulty we face, where the question was, "Is the data relevant, accurate, timely, and complete?" And the answer was, "It doesn't comply with the Privacy Act." Well, not complying with the Privacy Act has virtually nothing to do with the question of, "Is it accurate, relevant, timely, and complete?" Because to answer that question, we would have to know, "What's it going to be used for?" So, is it accurate, timely, and complete, and relevant to what purpose? So we always come back and say, "Now, what's the purpose?" I do think the point highlights the fact that the Privacy Act is not really the gold standard that we keep talking about it as. I mean, does anybody here think that government data about them is relevant, accurate, timely, and complete because of the Privacy Act? Anybody? The Privacy Act is not irrelevant, but it is certainly not the guarantor of those activities.

MS. LEVIN: Steve had his hand up, and then Jim next –

MR. ADLER: Well, you'll have to forgive me, Toby, I took a lot of notes on my BlackBerry, and I want to comment on some things that we talked about yesterday, as well as on this topic. First, I don't know if the Privacy Act is outdated or outmoded, but, in some sense, it doesn't really matter, because it's what we've got. We are data mining now, and we need to protect privacy now whether or not the Privacy Act is adequate or outdated. Maybe the guidance is outdated; maybe DHS in its statutory authority can revise the guidance for this Department and take a look at topics such as my colleague John Bliss talks about -- immutable laws and automated mechanisms for FOIA requests and disclosures. Maybe there are some modern technological methods that could be used to update the guidance under the Privacy Act to make it more relevant today without going through the long and protracted process of congressional review that really won't solve today's problems. That's my first comment.

My second comment is, as I was thinking about this yesterday while listening to the commentary and the discussions, what we're really talking about here is the use of data to infer intent. That is, human intent to commit an act, a crime, a terrorist act. In our court system, we spend months, if not years, hundreds of thousands of dollars, if not millions of dollars, to try to infer, prove, if someone committed a crime -- in the past tense. Now, you could look at that process as a data-mining exercise. We're collecting vast amounts of data, and we're analyzing it, and we're getting lots of different experts to render opinions on the data in court, in public testimony. Now, of course, maybe we have not as a society come to grips with whether we want to use the look of that kind of a system with regards to what Hollywood colloquially calls pre-crime. But still, I wonder, how many experts are we going to deploy to look at the data-mining data that we pull out? If we're going to accept the notion that data mining, itself, can compress the kind of review process that we use our court systems today to do -- that we're going to try to make it more efficient -- what are we going to do to ensure that we have the same kind of

expert review? As Vasant Dhar, yesterday, talked about an adversarial process -- that is, our court system is an adversarial process in which we pit people against each other in a destructive manner to make sure that we're looking at all sides of a case before we commit someone. We all know that we're still making mistakes in our court system. Even after months and all these people reviewing it, we're still making mistakes and sending innocent people to jail, if not to death row.

I work largely with commercial enterprises in IBM's Data Governance Council, and one of the things that we're constantly reviewing is this notion of an enterprise glossary. That is, that enterprises need a method for assuring that they have common terms and definitions across the organization so that they know what they're talking about. Because we've got lots of different people in a room, they have lots of different views. But I actually think, in the public sector, you don't want an enterprise glossary. You don't want to have that common set of meanings. You want to have different meanings. You want to have some kind of a management system that ensures that you have the right for people to disagree on what "data" means because data's dumb. Data does not have any intelligence. Data's not good; data's not bad. People are good; people are bad. People have intelligence. People analyze data. We have to make sure, if we look at roadmaps, that we've got a human management roadmap that analyzes data and enables people, who have the right to disagree in interpreting, "Is this guy really going to commit the act? How are we going to preserve privacy? What's the social consequences?"

Beyond just a technical roadmap, I've heard a lot of great ideas the last few days. I think there's a tremendous amount of value that this workshop has put together. We can all agree on an interesting technical roadmap for analyzing the data, anonymizing and protecting it, but the real challenge that I see is that human decision- making process -- the right of the people, who are sitting in a room looking at data, to disagree and to argue about it. That's, I think, the real challenge for DHS, is that decision-making process.

MS. LEVIN: Jim?

MR. DEMPSEY: I think Steve is on to something here. I think that what Steve, Fred and I are saying is that we have to stop talking about commercial data in the abstract. We have to start talking about what data, for what purpose. You were referencing a trial as a data- mining exercise. At least from a legal standpoint, I think of a trial as the exact opposite of data mining. I see a trial as a criminal justice process. As it was described yesterday by both the FBI and the Treasury speakers, where you have a known individual, or unknown subject -- nevertheless, you have a subject, you know that a crime has been committed -- or you may be investigating a criminal enterprise, including al Qaeda, the mafia. In those instances, you are trying to draw a picture of an individual's activities or an enterprise's activities, and you are collecting data for that purpose. That may culminate in a decision to invade a country. It may culminate in the

decision to arrest and incarcerate. It may culminate in other kinds of decisions, but it starts with an individual, and you're trying to collect more information about that individual. I do not call that data mining. I see that as data usage, absolutely. They are trying to put together a knowledge picture.

I think that what Fred and I are saying is: define the purpose, ask what is the information you're using, then you can ask is it accurate enough for this purpose? Is it relevant to this purpose? Are we getting what we need for this purpose? If I've got a known subject -- this could be somebody whose name was on a list or on an address book taken from some suspect. This could be a phone number acquired off of a pen register or it could be a license plate. This could be a name provided by an informant. If I've got somewhere to start, then I can go out, and I can go to ChoicePoint or any other commercial data supplier and say, "This phone number, who does it belong to? This address, who has lived there? This name, where does he live? What phone has he had registered in his name? What is his bank account?"

Now, in some cases, you may need to provide something extra in order to obtain information – for example, from bank accounts; but there is a process to get that. An address is what we call "public-record information," which is somewhere in a courthouse, it is somewhere in a phone book, it is somewhere publicly available, and you are using the commercial data aggregator to create efficiency to obtain it. It is good enough for that purpose. I just think we need to set that aside, almost. Should the government use commercial data? Of course, it should. Then we heard, yesterday, this notion of, what about the guy who's living a third-class lifestyle and flying first class? Is that an indicator that this guy is a terrorist? Has anybody tested that proposition? Has anybody gone out and found that out -- can we even figure out how many people would fall into that group-- what if you come back with 100,000 hits? That is useless.

MS. LEVIN: Well, let –

MR. DEMPSEY: And, we talked yesterday about the scuba-diving case. All of these sound so tantalizing when you spend five seconds on them. In the story, it seemed to me n that nobody actually ever got all the names of the scuba-diving students. The closest they got was the names of the certified scuba-training schools. No one ever got the names of the scuba-diving students, and no one ever figured out what to do with it, then. No one ever showed that they got anything useful out of that effort. We have got to stop having discussions where people say, "Wouldn't it be neat to get a list of all the scuba-diving students, because we'll find the terrorist who's planning to blow up the Golden Gate Bridge by scuba diving under the water?" We have to start talking about these things in concrete, practical terms, where there is some evidence of efficacy before you can get to the privacy issue. That is what Fred and I are saying.

MS. LEVIN: Well, I know Mary wanted to comment. And then, Mary, after you comment, I want you to lead us into a discussion of whether there should be some limitations on what information is collected, and, if so, what guidance can we give?

MS. DeROSA: First, just a quick comment on the discussion about the Privacy Act initiatives. My reaction, I think, similar to what others have said, is that people invoke the Privacy Act in response to a lot of the questions we're dealing with. At the same time, reaction when I hear people say, "Stick to the Fourth Amendment." There is so much you can do -- I mean, "Be careful of what you wish for," because the Privacy Act does not really limit very much of the kinds of uses of information we are talking about now. So, I do wish we could get away, not completely, obviously from the Act. It is the law, and it is relevant. However, it is not a framework for resolving a lot of the issues that we're dealing with.

As to the question of when can or should the government use commercial data, and when should it not?" The answer -- I'm a lawyer -- so the answer is, it depends. I tried to think of what it depends upon, and thought of three different questions that I think are relevant and define how you answer that question. It depends, first, on the kind of analysis -- and this gets to what Jim was just talking about -- the kind of analysis that you want to do with the data. How far is it from individualized analysis – which is the most comfortable way to use it -- not attenuated analysis? I'd say the easiest use of commercial data would be if you wanted to confirm an individual's information. For example -- I have a name and I want to find out more. That's simple.

Next you have identity verification, where you have an identity and you want to get a sense of how likely it is that that person is who they say they are. That is a little more attenuated, but still a fairly easy use of commercial data. All of these are subject-based analysis. You are still basing it on an individual you know, a particular suspicion, and you're doing some analysis based on that.

The hardest, of course, is when you're just using patterns to do some sort of pattern-based risk assessment. So, when you're making decisions about whether you should use commercial data, you have to consider how you're using it and what kind of analysis you're doing. Second, how are you going to use the results? I think this is really critical. If you're using the results as input for further analysis, for further investigation, the risks of harm are less, so that's easy. If you are using the results as a sole -- a primary basis for depriving somebody of a liberty – such as an arrest, deportation or inconveniencing them, in the airport screening context, -- then you've got to consider providing protections. And, third, what kind of protections? I just think it's so important to not just assume that, you either get the data or you don't get the data. There are ways you can test who should get the data. Is there a way of determining that only authorized people with an authorized or permissible purpose are going to see the data? Are there

approvals? Should there be approvals before the data is accessed? Can it be anonymized? Will there be audits? Will it be reviewed? That's the issue of audit. I agree that audits are very, very important, because they are a critical part of the process for governing how you use the data.

I'll just take two examples, based on those principles, of controversial programs, one where I think the commercial data shouldn't be used and one where I actually am comfortable with it being used. The first is the use of pattern-based risk assessment in airline screening. I think there you've got an analysis that has a lot of potential for false positives, along with a use that results immediately as the sole or primary basis for, at the very least, inconveniencing people. I think that is too high of a hurdle. You would have to somehow show that you've got a pattern that has an extraordinarily low chance of false positives, and as far as I know -- and I'm not a technologist, so there maybe something I don't know -- but I don't think we have those patterns. So, that's a place where I would not permit the use of such analysis.

On the yes side -- and I hesitate to say this, because there's so much misunderstanding about this program -- but the MATRIX Program -- as it was at the end, before it died, was actually subject-based. It used subject-based link analysis and identify verification in a law enforcement context. It was always based on a criminal predicate, which is a protection. It was only used to assist in investigations. It was never used as the final or sole basis for some sort of action. In addition, it had protections. So, if I were creating a program like this, I would have documentation of the need, a strong audit, retention and dissemination controls. With that kind of program, I think you can do, and do analysis while protecting privacy.

MS. LEVIN: Is there any more guidance on pattern-based analysis that people would like to give about when it is okay and when it is not? Any other thoughts?

MR. CATE: This is an area in which there's actually been a fair amount of work done, which is almost consistently ignored, so we do more work so we can ignore that and then do more. [Laughter.]

MR. CATE: We have done work on this issue. We have the Pentagon Technology and Privacy Commission, which came out with a set of specific recommendations on this. There have been a number of efforts, and there are a number of efforts ongoing now. I continue to think this actually isn't really rocket science. We might disagree on the exact words to use to capture it, but it would largely follow along the lines that Mary's just laid out. We can break it down into more specific questions -- What is the goal? What do you think you're going to discover? Why do we want the names of scuba divers? Have we satisfied ourselves as to rate of the false positive and false negative, particularly false positives? It seems like that is a critical issue. It's almost an audit question. Did you take your system and you test it before you run it on millions of records to first see -- Does it

yield acceptable results? Does it yield results that are even useful? If the number it's going to yield is vastly high, if it's going to identify one out of three people in the population, it's probably not worth it.

MS. LEVIN: Does that mean that we should do pilot tests? Will pilot tests get blown out of the water when there's transparency that an agency is doing a pilot test?

MR. CATE: Well, I think it is much better to do pilot tests and have the reaction to that than do the full-blown thing and then discover it was both a waste of time and an invasion of civil liberties. So, it is better to do it in the narrow approach first. I'm reminded of the Federal Trade Commission, where Congress instructed the Commission to do a study on credit reporting accuracy. The FTC very wisely, from my point of view, first said let's argue about the methodology and let's test the methodology before we go on to actually do the full-blown study of looking across credit reports.

One of the things that we have not talking much about today is that there are a fair number of procedural protections. I hate to use the word "procedure" here, because that always sounds weak. But, for example, even in an area where you're dealing with a top secret or high level of classification, separating out who authorizes from who does it seems very important. In many of the programs that have historically gotten in trouble, someone said, "I have a bright idea, and now I'm going to run with it," and there was never any independent oversight, even if it's from another part of the agency or from a high-level official. You know, thinking about, sort of, procedural protections. For me, the one critical procedure, without which I would not even begin testing is -- What are the redress mechanisms? Frankly, TSA probably discredited itself for all time to come in our lifetimes by going forward with CAPPS II and saying, "We know we have important issues here, but we're going to resolve those later, after we've rolled it out and put it in place." It is much better to say, "I already have satisfied myself about the acceptable rate of false positives, and here's how we we're going to deal with the ones that are false positives." If you can't say that, then I wouldn't ever run a program on real data.

MS. LEVIN: Steve --

MR. ADLER: I think we're all saying the same thing -- that we want effective checks and balances on the use of data. I think we have to separate out two aspects of data mining. One is collecting -- or maybe it's three aspects -- it's the collection and aggregation of the data, its analysis, and then its use. I think what we're all concerned about here is its use.

Investigation isn't a linear process. It's an associative process. People look at data and they don't see anything. They go and they look at something else, and then they remember something they saw in a previous pattern, and go back to it. If we're going to be effective in terrorist interdiction, we're going to have to allow people to have that

associative relationship with data. Of course, there should be some kind of a bonding process. We bond tellers to deal with cash. Maybe we ought to bond database administrators to deal with data. What ultimately is more valuable, the data or the cash? It could be an interesting argument. We could take that bonding process and apply it to people who are using data.

The real checks and balances should come into play after the analysis has been completed and someone's ready to render a conclusion and use the data. I think that's where we want to have an adversarial process. As Americans, we want to have a really diverse group of people at the table who are looking at that data to determine its use, because everybody looks at data differently, depending on your perspective. If you're in law enforcement and you're seeing criminals every day, every piece of data is going to be a criminal. And if you're a privacy advocate and you're looking at data every day, everybody is going to be innocent. It's not to say that one view is better than the other, but I certainly want both those guys at the table when they're determining the use of my data.

MS. LEVIN: Well, let me ask how should DHS access the data? Should it collect it, take it in-house, leave it in the hands of the information providers, set up intermediary organizations? How should the data be accessed?

MR. DEMPSEY: Toby, before –

MS. LEVIN: Yes, Jim –

MR. DEMPSEY: Quickly on the pattern-based analysis. You have Able Danger out there, and it has an odor of weirdness about it. But somebody needs to get to the bottom of that. Senator Specter has said he'll hold hearings on the Pentagon program that claims it is doing data mining. Some people -- maybe you have been associated with it -- but I can say the whole thing has a little bit of the odor of weirdness about it. But they claim that they had been doing data mining and had identified Mohammad Atta pre-9/11. You know, wow! I mean, is this proof of the concept? And it all gets tied up with the fact that, it's the military, and should they have been doing it domestically? And if they had it and didn't share it with the FBI, or maybe they didn't have it at all. The whole thing is a little bit weird. However, I can't believe that somebody is claiming to have done this and isn't willing to defend it. Somebody needs to get to the bottom of that. And, so far, I have not seen the proof of concept, in terms of finding the unknowns.

MS. LEVIN: All right. Again, how should DHS access data? Should we use an intermediary? Should we take it in-house? Is it more secure, more privacy protected if the Department applies its principles to the data?

MR. SMITH: Just to remind people, I have a feeling most of my colleagues here are attorneys, judging from the discussion, and I'm not. I did hang around with economists a lot -- [Laughter.]

MR. SMITH: -- and economists refer to the adversarial -- to dispute settlement processes, in general, as deadweight loss. [Laughter.]

MR. SMITH: Of course, people say that economists know the price of everything and the value of nothing. So, insults go around. In any case, I think there's a lot of discussion about doing the technology slightly differently. And I think in a lot of those instances, it doesn't matter, you could do it either way and still the core thing is, Do you have accountability for what happens?

I'm very pleased to hear people hammer on the audit capability as a core requirement, because I agree that it is a core requirement. In some ways, that's relatively easy from a technology point of view to do, because, first of all, it's something you can do after the fact, and, in general, the motivations around privacy violations are not the same as around suicide bombers who don't care if they get caught. So, ex-post audit is probably a pretty effective tool.

Secondly, although people say, "Well, gee, that's an awful lot of data if you audit every transaction back to the individual." Keep in mind the kind of underlying data. We're getting, four-hour movies and the audit record on that is still only, like, three characters long. So, I'm not too concerned, frankly, about the volumes of data associated with that. Plus, if you have an audit record, I think of it as in accounting. For example, you only keep your general ledger for a certain amount of time. You keep some of your records after that, so you have a big window -- how much of a window you need to have -- for people to notice that some crime has been committed, and then go and investigate it. So, I think audit's great. I would say in terms of your quick fix, you have it. It doesn't matter, as long as I have accountability for what happens.

If I could also just make one observation, I would strongly agree with what several people have said, that it's highly desirable to know what the heck you're doing before you do anything. And this doesn't just apply to the privacy rules, this applies to security as well. It's a human-nature issue, to a large extent. If you are in charge of something, and, of course, you're in a hurry, and you're under a congressional mandate, and there's a lot of pressure to get some results in the next 15 minutes, then your instinct is to say, "Give me as much money as possible, as much authority as possible, as little oversight as possible, as few restrictions as possible, and as little constraint as possible." That's a human-nature issue, and it applies to all kinds of things. It's like, "Give me a bigger PC." "Why?" VOICE: Do you have a demonstrated need for it?

MR. SMITH: However, I would say, on the other side of the equation, let's not lose sight of the principal value – the mission value – which involves, (a) getting things done quickly, because, you can't move faster than the other guy if you don't get things done quickly. So, you do have to be mindful of the amount of process you put in the way of, not only executing operations on a day-to-day basis, but of changing your operations, and then let's go back and check the basis for this, and let's go change the MOU and that sort of thing. It takes a heck of a lot of time, and I think that it doesn't serve the privacy community well to be tagged as the obstacle to implementing and achieving the mission. I mean, that's something you definitely want to avoid.

MS. LEVIN: I want to move into talking about better governance, but, before we do, I want to make sure that we address the process issue. What do we do about addressing individual choice and access? On the last panel, Michael Daconta suggested exploring automated FOIA. Can we get it to the point where technology makes it automatic, so that people just have automatic FOIA capability? What do we do about access and correction and redress regarding the use of commercial data? Those are all important protections as well. Chris?

MR. HOOFNAGLE: I think the Department of Homeland Security can do a lot to ensure better accountability of data brokers. Mr. Adler's idea of bonding the data holders is a good idea. But consumers do not understand how their data are acquired by these companies. These companies have thrived, and they've collected data so successfully because they have flown under the radar. They are very vague about the ways they use data. They all can say, "Well, we use this data for homeland security and law enforcement," and while really the majority use appears to be marketing. They will buy data from government records and public records.

It seems to me that, as a condition of contract, Department of Homeland Security could just mandate that these companies bring all their activities into compliance with the Fair Credit Reporting Act. They could give better notice of when they're collecting personal information, how they're collecting it, the purposes for which they're using it, and give better access. This is all within the power of the agency, should it choose to do it. What's happening right now is that a lot of these companies are using means that border on deception to collect personal information from people. I have looked at these databases and ads for personal information very carefully. You will find that among the things they do is get information from customer service centers. So, if you call up a customer service center, you reveal your subscriber name, address, and phone number through a system called ANI. That data is picked up, and it's sold, and they'll resell your number, if it's unpublished or unlisted. They will then ask you questions. And, of course, then the company that you're calling has an incentive to ask you more and more

questions, because they can turn around and sell that data to a commercial data broker. I think a lot of these systems of information collection could be reformed. There could be a lot more transparency here.

MS. LEVIN: Anything else? What about due process?

MR. DEMPSEY: Well, Chris, I wouldn't go quite so far as to say that all of the data aggregators' products or services or activities should be made subject to the Fair Credit Reporting Act. I would say, just on the question of access, that I think that if you focused on what is it that the government is accessing or using, subscribing to, you -- it's worth considering whether those databases should be subject to an access requirement. I think this is an example of where you can use the access principle even in a national security context. Obviously, you're not going to tell Osama bin Laden, "This is what we have about you." So, Osama bin Laden is not going to be able to write to the Department of Homeland Security and see what the Department of Homeland Security has about him. On the other hand, I don't think you're giving anything away to Osama bin Laden if you tell him and the rest of the world that we subscribe to certain ChoicePoint services, and that you can go to ChoicePoint and find out what ChoicePoint has about you. I mean, I would like to see Osama bin Laden come in and then try to challenge the accuracy of the ChoicePoint data. [Laughter.]

MR. DEMPSEY: So, I think that you can have the notice principle and the access principle, which then allow you to implement the accuracy principle. You can have, through the contracting power, and I think a number of companies, of course, have come forward, to one degree or another, and opened their data systems and given various forms of access. The simple way to do it is to say, "Whatever you're selling to the government, provide to the individual." I see that as a win-win, for accuracy, for government effectiveness, for individual rights.

MR. CATE: I'm just going to keep saying about how important it is to be specific here, because -- and not just because I didn't have time to prepare. So I only wrote this one line down, "Be specific." [Laughter.]

MR. CATE: But because, again, this is a perfect area of where we run the risk of confusing what the real issues are -- or what the issues that concern consumers, as opposed to those of us who love to debate privacy issues. If you're denied boarding on an airplane because your name didn't match with the address you had given, you probably don't really care a lot about the underlying data. You know, maybe 500 transactions went into generating that data set, listing former addresses gotten from half a dozen different vendors. Going back and correcting exactly when I moved 14 years ago from Washington, my guess is not foremost on your mind. What you do want to know is, "Why were you denied boarding on that plane?" So, the question is the use, not the underlying data. What you do want to do is correct it if it is wrong for any future use.

Maybe the inference was wrong. In other words, the data may have been entirely accurate. But if the inference was wrong, and, as a result, you're denied boarding the next day, as well, you're no better off than you were before. So, focusing on the data, in that example, runs the risk of completely mis-focusing where the real need for redress is. The redress is -- "What did the government do with that data? What was the –

MR. DEMPSEY: But if you correct the data, how are you ever going to stop the government from saying, "Your address and your name don't match"?

MR. CATE: Precisely. Because, in fact, the way in which data goes into making those calculations, there may be a hundred disparate data points, and identifying which one triggered the algorithm to yield that result. What you care about is the result. You want to go in and say, "This is my address." Now, how do I verify that, as opposed to verifying that a credit card application I filed 12 years ago, on which I listed my address correctly but they couldn't read my name right, so they transposed the digits in my address and wrote it down incorrectly, and now I want to go back and use the force of federal law to correct that data in a private data set. That just strikes me as a waste of time and of no interest to the public.

MS. LEVIN: It may be –

MR. DEMPSEY: How do you ever prevent the recurrence of a false inference? I mean, if it's a true inference based upon false data, how do you ever prevent the recurrence of that even after you correct the data?

MR. CATE: That brings us to the next step, which is to say, is the government bringing this data in-house? In which case, it is not bringing in all that background data, it's bringing in only the conclusion. "These are the addresses we show for this person." You allow correction of that.

MS. LEVIN: But you have –

MR. CATE: -- commercial –

MS. LEVIN: -- You have to distinguish between misidentification, where I'm not the right Toby Levin that they're looking for. But what if I am the Toby Levin that's on their list? Then what? And I want to assert that their intelligence is wrong.

MR. CATE: I would seek correction of that conclusion that's reflected in government data, that you are the Toby Levin they're looking for, rather than all of the underlying data on which that was based. It strikes me we are just largely missing the point to do it any other way. There are many other types of data situations that we might imagine where you have data that we're going to look for because we're looking for patterns of data. We're not just verifying identity, we're not just verifying address. That's where we might need to talk about -- What is the use? What's triggered by this? right

back to Mary's questions -- what is going to be the effect of having been identified on this list? Depending upon the potential severity of that effect or the other opportunities for redress, it may be necessary to go back to the data. It may not. In other words, if the effect is so minimal, and it's merely to just move us along one of a hundred steps of an investigatory chain, I'm much more worried about the use than the data.

MS. LEVIN: All right. Steve?

MR. ADLER: I can only comment on the ability to create a system like this. I would say it's very difficult. There are lots of different processes involved, there are lots of different steps involved. I'm tempted as a privacy advocate to say, when we go banking online, we should get a bank statement of all of our transactions. That's wonderful. Similarly, I'd love to see a transactional record of what DHS does with my data. But the banking statement is a statement of fact. It's not a statement of inference. The bank is not disclosing the inferential decisions they made based on the factual data they obtained. They're only telling us what's in our bank account.

MR. DEMPSEY: That's not entirely true –

MS. LEVIN: We have to use the microphones, please. Jim –

MR. ADLER: Jim, let me finish. What I think is really difficult to do -- first of all, there's a lot of data. And we're talking about the government's use of commercial data. But, there's a lot of data that in the future is going to be mixed together. The government's going to be collecting a lot of data. You can see video cameras in this room, and we're being recorded right now, and, you know, this data's going to be aggregated, right? They're going combine our voice with our video, and at some point, with a transcript. Now, it may not be broadcast on the Web, but, you know, I've been to conferences before where my presentation was recorded and it was posted as a video. I'm sure everybody else on the panel has, as well. Is that government data, or is it private data? Well, the government was recording it. Will that be combined with private data? It might be. Where's the delineation?

I think the challenge that we face is that we're going to be recording terabytes of data about people, vast amounts. Even if you were to automate the recording -- and I'm in favor of immutable audits and lots of recording -- what are you recording? Are you recording facts, decisions, inference? What are you making available? Even if you automate the availability of it, you're also inviting, a lot of people who don't have access to computers. They're not going to access it online. Is it going to be via voice response? Who's going to man the phones to answer the questions that people are inevitably going to have? It's a vast undertaking. I'm just saying that this is not a simple endeavor to do, from a technology perspective, and even less so from a management perspective.

MS. LEVIN: Well –

MR. ADLER: It's really a vast undertaking.

MS. LEVIN: Accessing inferential data brings us to a whole other level than factual. But, Jim -- do you want to pass the microphone down to Jim, please?

MR. DEMPSEY: Steve, I a hundred-percent take your point, in terms of the massiveness of the undertaking. I will just say that when the bank charges you a higher loan rate because they have drawn an inference from a fact presented to them, the fact that you did not repay an earlier loan. They have inferred from that that you will not pay your future loan and you're a greater risk, and they, therefore, charge you a higher interest rate. At the same time, they are required to report that to you and to give you an opportunity to challenge that. And you go back, and you can say to them, "You drew a false inference," or, "The facts were wrong." So, we do, in fact, have an obligation upon the bank and every lender in the country, every employer in the country, every insurance company in the country is required, when they have drawn an inference about you that is adverse to your interest, to tell you that they have done so, and to tell you why they have done so, and to give you the opportunity to respond to them. We have a whole industry created to respond to those claims. And, by the way, almost half of those claims that people are making are false. That is, people are saying, "That wasn't my loan," when, in fact, it was. And the banks even deal with that.

MS. LEVIN: But, Jim, we're operating now, though, in terms of Homeland Security, in an environment where we're talking about national security, talking about law enforcement exemptions, where the community does not feel that they can give you access to the intelligence, or access to information –

MR. DEMPSEY: Well –

MS. LEVIN: -- underlying why the decision to put you on a list.

MR. DEMPSEY: Sometimes yes, sometimes no. If you're keeping Osama bin Laden off of an airplane because he gave a false address, I think you can tell him, "Your address doesn't match." If you're keeping him off the plane because he's a terrorist, and he doesn't know you know, then that's a different thing.

MR. ADLER: He's not an American citizen. Is he entitled to –

MS. LEVIN: Use the microphone, please.

MR. ADLER: -- a FOIA request?

MR. DEMPSEY: Yes, like FOIA. FOIA is -- he is not going to get -- [Laughter.]

MS. LEVIN: Okay, let's move now to something that may be a little bit more manageable. And I think -- [Laughter.]

MS. LEVIN: Thanks to some of the work that's being done at DHS, I'm much more confident it can be more manageable. How can DHS better govern when and how commercial data is used? What can we do, in the way of process in policy and technology? Steve, I'm going to let you start, and then, Martin, I want you to tell us what we're doing at DHS.

MR. ADLER: It's not an easy answer to that, either. I think you have to look at the data supply chain that you have here, because there is an extended supply chain in both the origin of the data, which may come from the commercial sector, maybe from the people we heard yesterday, or others, that are supplying commercial data that the government is co-mingling with its own data. Beyond that, you have other extended supply chains, because it's not just that the Department is leveraging the data and analyzing and determining its use; it's also sharing that data with law enforcement authorities throughout the country. And those law enforcement authorities throughout the country have varying degrees of ability to protect that data appropriately. And all those factors need to be taken into account.

You've got to take a look at concentric rings of privacy protection, as well as data protection. I think they're equally important in looking at a roadmap, as well as the governance model. As I said earlier, it's difficult to do, but all these issues are really complicated, and you have a small group of people -- and even though the Privacy Office has a somewhat larger group, it's still fairly small, given the challenge that you have in DHS. But, you guys aren't enough to look at all the complicated issues here. You need a collaborative team of people, of data-management experts, of lawyers, of privacy people, of security people, of operational- risk people, of systems engineers and architects. You need to put together, collaborative teams that are looking at these issues and looking at the data supply chain to map out the business processes to understand far more specifically, how we are going specifically to apply policy to these complex challenges.

I almost want to take the human governance approach. I don't remember the gentleman's name who was talking about separating the rules into middleware from the application design. I think somebody, yesterday, talked about how, if you hard code your rules into your applications, and your rules change or something evolves, you're stuck with an outdated application. Well, of course, most companies and most governments throughout the world are stuck with tons of outdated applications that were hard-coded with rules, and we're trying to figure out what to do with those now. I think the same goes true for the decision- making process, itself.

If you want to find a flexible structure, a flexible governance model in which you can analyze how policies specifically apply to a business process, how are we going to model that, but, at the same time, collaborate across disciplines, break down those functional stovepipes. If we're talking about breaking down the stovepipes that separate

data, which I think one guy talked about earlier, stovepipes have largely preserved our privacy, from a government perspective, as the inability of government to share data has been the single factor in preserving our privacy for the last few decades that we've been collecting data. If we're going to break down those stovepipes and aggregate this stuff, we, at the same time, have to start breaking down the stovepipes that separate people from working together, break down specialization stovepipes and get more people to collaborate effectively. All of these different perspectives that we're sharing here at the table are equally valid in analyzing appropriate use for data. I see building that governance model as the most important part of the task we have ahead of us.

MS. LEVIN: Well, that's why Martin's shop is so extremely important at DHS.

MR. SMITH: Okay, thank you.

MS. LEVIN: Martin, take the microphone out. That's better.

MR. SMITH: I'm not going to talk about the whole solution, because that's as you say, a pretty big problem. However, I wanted to just put out some information for people to begin to think about -- How can I manipulate the processes that are in place in order to achieve the policy goals I have?

A large part of what DHS is doing is implementing OMB, and I'm talking about, basically, the creation of new systems and the upgrading of new IT systems. OMB guidance includes figuring out what you're trying to build the system for and how you'll measure whether you have success or not. That said, the implementation record across government is pretty primitive in these areas.

Although we're doing, in general, what you could go and read in A-130 and other kinds of OMB guidance, it might be useful to have some of the specifics of what we're doing. In particular, we conduct a review of any major IT system that is funded in the Department. That review point, which occurs in the capital planning process and the Enterprise Architecture process, provides a tollgate that anybody with an issue, whether it's a privacy issue or a security issue or an accessibility issue or an efficiency issue or how-do-we-risk-management issue, can intercept these things as they come through and get their issues considered in a formal way. This is distinct from chasing these guys around the landscape.

This review process will make better use of the limited resources of folks like the Privacy Office, who don't have a large police force to go out and find out what people are doing off site somewhere. My perception is the Privacy Office has actually focused in on this considerably more lately. I say that, because I've seen Privacy Office people over in my office talking to the people who do this. So, that's the inference.

The Enterprise Architecture process, for those of you who are not around IT at all, is not necessarily just about IT, although it commonly tends to focus on it, and that's unfortunate. But it can be used as a process for a structured review of any investment proposal and it doesn't even have to be to build a system. It can actually be a review of a program that, in principle, doesn't have to involve systems at all, although there aren't many of those anymore. It's more efficient to get these issues focused on early.

In the capital planning process, in order to spend large amounts of money on a new system, you have a very early stage review, which is supposed to be the kind of high-level review asking -- What is this for, what's the value proposition of this investment, and what am I getting out of it? And, again, it's commonplace to say, with policy issues and all these other things, the earlier you get involved, the more effective you can be and the cheaper it is compared to cleaning up later. That's an excellent place for people to get involved. Let me just analogize briefly, if I may, to the security process. The security folks have been working on a similar set of concerns -- very similar, in some cases, including desire for things like audit, for accountability reasons -- for quite a while, and NIST has released some guidance, as has OMB. But, I think the key lesson is that the security guys have caught on to the idea of trying to formalize, as much as possible, the risk they're addressing and the standard ways of covering those risks. I would hope that we can do that more in the privacy area.

It's not an infinite situation. Most of them -- 80 percent of them -- aren't open-ended situations. They're standard sets of concerns that need to be addressed, and, of course, those are codified to some extent. But, beyond that, a standard set of solutions, a standard set of possible tools or mitigations that you apply to get through that. That has all kinds of good effects -- (a) it works better than making up solutions individually; (b) it moves the process along and makes this seem like less of a burden to the people who have a mission role that they're trying to achieve. There's a pre-built solution, they can plug it in and go, and that's all they care about. They just want to get down the road on the project; and (c) it promotes consistent application. Consistency is a huge issue, obviously, for any legal requirement, or any kind of a requirement -- consistency and transparency.

So, I think the investment review process is a good way for us to plug in these considerations. However, my perception as a citizen here and what I read in the papers and what I hear around, is that we're having a little trouble gelling on exactly what rights it is we're trying to enforce, and what the threat scenarios that we really think are realistic are. And, as was demonstrated a few minutes ago, what the heck is data mining, anyhow? Until you get some kind of better definition, consensus on what's allowed and what's not allowed, it's very difficult to routinize the application of these things.

Let me just give myself a hook for some future discussion, I hope, here, maybe. One thing that we are trying to do on the technology side is to build some tools that will

help mitigate certain kinds of problems that -- coincidentally, in the security and privacy areas -- but which actually help us do better information sharing. I do think it is more successful to focus on how your process is going to get your mission done than it is to always style yourself as a -- you know, a constraint on the process. I'd like to see some research in this area. What percent of an ideal distribution of a particular document is actually happening now? Are we giving the document of record to 2 percent of the people who have a legitimate right and need to know it? Eighty percent? I don't know what the answer to that is. And that's a hugely important question that we don't have good data on. Another one, concerns human review -- by the way, human review is what corporation are trying to eliminate in the business world. In other words, if I have to make a million decisions a day, I do not need human review on many of them. I would also assert that there is at least the possibility of improving the consistency and quality of decisions by eliminating human review, because a lot of time human review is no good unless the humans are trained, and training a lot of people and changing policy, you know, three times a month, and retraining them, is actually not a very reliable process.

MS. LEVIN: Well, you've covered a lot of points. Unfortunately, we won't be able to go into all of them. I would like for Mary to talk about when we should share information within the Department and outside the Department? Should there be any limitations, in terms of how the data is shared? Martin was saying it would be a good thing.

MS. DeROSA: There's so much I want to talk about, because I also wanted to talk about the governance points. But I'll the question you asked me -- what should DHS do with information that it has, how should it share it, and should it share it? As we talked about before, should they treat information -- sharing information inside DHS the same as they treat sharing it outside? My view is that I think, absolutely, you've got to share the same way within and outside the DHS. I think the agency boundaries, at this point, really have no relevance to need for information. You have to get to the point where they're not relevant to how you share information. I think the agency boundaries are irrelevant to the Privacy Act. I don't think that the Privacy Act actually presents a barrier, but that's something that should be considered. When we're talking about why you need to share information, and with whom, the DHS boundaries have no relevance.

Now, that said, I think I need to be very clear about what I mean by "sharing," because -- and this is something the Markle Foundation Task Force has spent a lot of time talking about -- when we say "sharing," we don't mean taking a database and plopping it over to the other people who need it. When you actually transfer an entire database or data, what you're doing is freezing that data in time, so it ends up becoming inaccurate, because the original database is continuing to be updated, but the transferred data is not updated. So, when I'm talking about sharing data, I mean allowing access to data. The

owner or the entity that originally had the database has to act as steward to update it, to allow access when the person or entity has the correct permission, to allow access consistent with the rules -- and there have to be rules, and there have to be policies, and audits of all access to databases.

Martin, you and I had a conversation, along with a colleague of mine who has a lot more technical knowledge that I did, and I understood just a little bit of the conversation. I know that there are challenges to being able to access databases in this way, and setting things up so that another agency or another entity can access the data. But my understanding from that conversation was that there are challenges, but it can be done. I think that the way you need to look at sharing is the same basic approach that we discussed about accessing commercial data -- that DHS shouldn't be acquiring and bringing in-house commercial data unless they absolutely have to, and that it's more important to access it, or however you say it, and leave the data where it is. I think that should be the preference for really all data.

MS. LEVIN: Okay. Let me say, before we move into your closing points that I've only been at DHS now since April. I came to the Department after being at the Federal Trade Commission for 20 years, working on privacy issues for much of the time-- I guess about a dozen of those years. And I say this very sincerely. I've never found an organization to be as focused on privacy as is this Department. And, in terms of thinking about privacy and asking all the right questions, folks at DHS are asking all the right questions. We're in the process of building a culture of privacy. It takes time, but there is a great deal of receptiveness by the components within the Department and from top down. I think this is an opportunity for this Department to be a leader in implementing -- first, in terms of figuring out how to do privacy with a great deal of specificity, moving away from just principles, but also to serve as the leader, government-wide. So, I'm very excited about being at DHS during this formative time.

To close the panel, I thought I would continue on the roadmap theme and ask each of our panelists to come up with a red light or green light to wrap up their guidance. We could go on for another hour. And, actually, if we don't have a lot of questions, I do have some more that I'm going to pose. So, if there are people who have questions for this panel, please line up behind the microphone. In the meantime, let's start with -- Martin, with you. Is there a red light or a green light that you would add to our roadmap?

MR. SMITH: Well, I've got one of each.

MS. LEVIN: Okay.

MR. SMITH: I actually broke your rules, and I did this as two stops.

MS. LEVIN: Okay.

MR. SMITH: -- I'll just rephrase one as -- a la Jeopardy. I've got: Technology can optimize mission performance against almost any set of rules, but we can't do it against a lack of rules. So, we really need to get at the business owners -- and, in this case, the Privacy Office, but in cooperation with the mission -- to try to formalize what we're doing and what we're not doing. If there's controversy on something, make a call, take the heat, and let us, operate on whatever that basis is. But the idea that, last week it wasn't data mining, now it's data mining, you can't do that. So, do formalize the rule sets as much as we can. And I think it's a real open question how successful that can be, but I sure would like to find out.

My second stop sign is to look for solutions that avoid direct conflict between privacy and the mission. I think there are lots of win-wins here. We need to go find them and stay out of the, "Stop what you're doing until you satisfy my list of things." Because generally speaking, when a big enough disaster happens, you're going to lose.

MS. LEVIN: Okay. Chris?

MR. HOOFNAGLE: There's a joke or argument that's often made in this community, that privacy advocates are Luddites. [Laughter.]

MR. HOOFNAGLE: I like hearing that sometimes, because if you go to the offices of the Center for Democracy and Technology, or a group like Electronic Frontier Foundation, you'll find more technology than anywhere else. However, I think what you'll also find is that we're skeptical of technology. I think it's good to ask lots of questions. And to ask whether or not a lot of the goals can be done through procedure rather than through technologies.

The second idea I think I would leave you with is to be skeptical of identification as a means for security. Identification has a lot of meanings attached to it that maybe in this room we don't always get. I'll just give you one example. You showed a driver's license to come into this building, but if you walked just seven blocks down the road, you could walk into any of the House office buildings or any of the Senate office buildings and not show an ID. You would go through the magnetometer, but you would not show an ID. Why is that? It's a political reason, right? Identification is political. It's about power. We should figure out how we can meet security goals without abusing them.

MS. DeROSA: I guess I'd go back to something I think Steve said -- that the end of stovepiping or the relevance of agency boundaries has taken away one of the great protections we had -- privacy protections that we had as a tension of just having restrictions on sharing information and not for policy or other reasons sharing it. So, I

think we need to look -- it's extraordinarily important -- to the governance discussion to have new ways to put that tension in. And the one that I might suggest as a green light, is to really focus on audit. When I say "audit," I mean both the logging of the information and what you do with the information -- how you, in a real-time practical way, keep up and control what is happening in the system and have periodic reviews. I mean audit logs that are being checked in a way that we have not traditionally done, much more real-time compliance.

MS. LEVIN: Jim?

MR. DEMPSEY: Okay, before I give my red and green light, I do want to say that you've asked for a roadmap – and I think Mary gave a roadmap, so I don't want people to miss or forget what Mary said. I think a lot of it's spelled out in a paper she wrote at CSIS. It's consistent with an article that Lara Flint and I wrote in the George Washington Law Review, but Mary spelled it out. So, I want to say that we've got a roadmap, right there. I guess I'm almost narrow-minded in my red- light/green-lights. My red light, I guess, is, don't abuse the Privacy Act exemptions. And I have to say that, particularly with DHS, where recently, I've seen Privacy Act exemptions evoked with abandon. And I think that's a mistake. I think we need to find a way to deal with the relevancy clause of the Privacy Act rather than simply exempting entire databases from the requirement to collect only relevant information. And I'll just leave it at that.

MR. CATE: I would return to Martin's point about the need for rules, and echo this, for a number of reasons. I think rules are incredibly important for individuals, in protecting individual rights. I think they're incredibly important for the Department and for the people who work for it. And I think you don't have to watch too many bright people come up with ideas that then get slammed in the press, and, you know, Congress comes out and says, "You cannot do that," to say, "Wait a minute. I'm not going to deal with these issues. I mean, why would anyone in their right mind get involved in this area? It's nothing but a death trap for promising careers."

I think we also need rules for, frankly, better security, because if we have a clearer sense -- and I think we've talked a lot about the ways in which privacy might enhance security by focusing on rules and objectives and how reasonable this method is to achieve that goal or objective. The more that rules help formulate that process so that you couldn't spend $100 million on a computer system without saying, "Here's what we're trying to do. This is why we're acquiring data. This is why we're going to process it."

But I guess the place I would focus -- and this is both a red light and a green light -- and, frankly, one of the reasons this Department has such a rich and well-respected Privacy Office is because Congress created the Privacy Office and also has provided funding for it that's unparalleled in Congress's history with privacy offices. That's not to

take anything away from the Privacy Office or the people in it. They deserve enormous credit.

But, it is also to say that if Congress said to the Secretary and to the Assistant Secretaries and everybody else, "We think this matters." If Congress, which has, on the whole, been just about invisible in this debate, other than to just slap the hands of departments that do things they don't like, were, instead, to step up to the plate and say, "Here -- yeah, this is what we think the rules are for doing this; and, by the way, we write federal laws, so now this is what you think the rules are, too." Such action would go a lot farther towards putting those rules in place, making them more consistent across agencies so that the problem wasn't replicated by having a different set of privacy rules that applied in different environments, by giving some predictability to people who work in this Department and others about what they could do, and elevating the protection of civil rights and security to federal law, rather than the good intentions of individuals or for regulation from within a department.

MR. ADLER: The last word. I want to agree with what Martin and Mary said, specifically, because I think we have a balance challenge or a challenge in balance -- and they're mutually inclusive goals. We want to find terrorists before they act, because we want to prevent terrorist or criminal activity. But, by the same token, we also want to prevent privacy violations before they occur. And I think those goals are mutually complementary. I think that there is a way for us to find a middle road that allows us to prevent the incidence of both terrorist acts and privacy incidents.

It comes down to something Martin said. It's a quality-control issue. That is, ultimately, data governance is about quality control of the data-supply chain from the origin of where you find the data to how you clean it up, to how you analyze it, to who decides how to use it, to where you disclose it. There's a process involved. And we have to get better, both in government and in business, at figuring out how rules and policy apply to that business process, and knowing, in a much more enlightened way, how to make appropriate decisions along the way, and document them.

The audit log shouldn't just be for access to data. The audit log should be, more importantly, about the human decision-making process. We ought to be capturing what people say in meetings, how they decide to remediate things, and what decisions were made -- as importantly as the audit logs. And I want to say one more thing about the audit logs, because it's a real challenge. Audit logs, on a technical basis, are synchronous, but understanding them is a totally asynchronous event. One big challenge we have is provisioning the information we glean from the information usage process, so that we're talking about auditing how information is used. But we have another challenge, and that is provisioning back the information we capture from the audit logs, both from the digital use of information, as well as the human use of information, so that managers in lines of

business or in IT departments can review what's going on in the organization and make appropriate decisions to prevent privacy incidents from occurring. Those are really big challenges. It's a quality control issue at base.

MS. LEVIN: Well, I see the makings of many more workshops, as I close my part by saying, we see privacy as good information risk management. It's not about saying no. It's about good information risk management. And we're all about working out solutions to help the mission. So, the first question?

MR. VON BREICHENRUCHARDT: Thank you. Hi, I'm Dane von Breichenruchardt. I'm with the U.S. Bill of Rights Foundation. I guess I'd like to say, a question was brought up in the earlier panel -- "Against whom are we most trying to protect our privacy?" And I think that answer is borne out of the idea that we recognize that it's the -- it's the entity that can do the most damage to us -- not that it has a desire to do the damage, but the one that has the capacity to do the most harm. And, in this case, it is the government. It's really the government is the one, because, as it was brought out earlier, you know, it's very easy -- not easy, but it's a lot easier to combat private enterprise if they get a hold of our name and call us or sent us unwanted advertising, or somebody tries to steal identity. As tough as it is, we can deal with it.

But it's the government that has the capacity to really do harm, severe harm, and much tougher to penetrate. And so, it seems to me that -- my question goes to -- don't we have conflicting premises here? The first premise of this country was that we ceded the authority to use force to the government, appropriately, but not to instigate force, to use force against force. And it -- and I don't mean "force," like knocking people over the head with billy sticks and carrying them off to jail, because there are all kinds of forces, including the force of investigation, requiring people to answer questions, data mining, and this sort of thing. And the principle has always been, before 9/11 and data mining and this sort -- to the level that we're talking about -- it's been that you got warrants, you had to show probable cause, you had to have a suspect, and you had to move along. Now we're talking about going in and doing this pattern stuff.

So, now we have a meeting. We've -- this whole conference, and particularly this panel, is premised on -- that government is going to use commercial data. So, I'm not trying to be iconoclastic here. I want to back up and think outside the box. Shouldn't we be challenging that premise and back up and re-ask the question, Should we be allowing government to be using these private databases in the manner that we're talking about? Not that they can't use them at all, but they -- for the purposes and the way that we're talking about them using them -- at all. We ought to go back and challenge that premise. And it seems to me that, since it is about government that we're the most concerned about our privacy, to be in a government building discussing with a government agency, where it says "to build privacy protections into the government's use of commercial data" -- is

very much like a bunch of chickens sitting around in a fox's den discussing with the fox implementations or protections to keep him from eating so many chickens.

MS. LEVIN: I would start -- I think we –

MR. VON BREICHENRUCHARDT: Well, let me make my last point and I'll sit down.

MS. LEVIN: Okay.

MR. BON BREICHENRUCHARDT: And that is, the roadmap. And my suggestion is, shouldn't we think in terms of telling DHS and other members of government -- not because they're bad people; they're hardworking decent people -- that they already have the roadmap, and they should follow it, the one that got them to where they are now, and that they should use it to go back where they started from, and let's start the dialogue over again and rethink this premise about them using this kind of data in the way they way to use it. Thank you.

MS. LEVIN: I think there is agreement across the board, that the first question that anyone would ask in any program is," Is the commercial data necessary?" And we want to start with that question. Who wants to respond?

MR. DEMPSEY: I just don't see how you can say that the government shouldn't use commercial data. How is the government agent going to find somebody's telephone number? The phone book -- the phone book is a commercial database. That's where you start. There are some commercial data that is very useful, that is accurate enough for the purpose at hand. And I think, yes, the premise of this panel was, "the government will use commercial data." The question is, "Which data, for what purpose?"

MS. LEVIN: Any other responses? Mary?

MS. DeROSA: I guess I'd just say that the premise of this panel was now because we were told "This is what you have to start with," but because at least for myself and most of the people on the panel, we have asked the first question, "Is commercial data necessary?" And we've concluded that it is an enormous resource for the government to protect security that you can't just ignore. Now, do you have to use it right; do you have to use it with protections; do you have to use it only when you need it? Absolutely. But I think you assume that you hear a premise, but we've asked that question of ourselves. I won't speak for everyone, but at least I have answered it yes, we do need them -- we do need to use that data.

MR. ADLER: Well, you know -- you know, the –

MS. LEVIN: Steve, please use the microphone.

MR. ADLER: The premise, of course, is based upon the presumption that, the old methods of protecting ourselves -- that is, perimeter defense -- no longer apply and that we can't simply hold off our borders, put up thick walls, a national line, or submarines off the coast to protect ourselves from terrorists, that there is a threat from within, and those old models don't work. But I think, as a democracy, we ought to revisit that assumption on a regular basis.

MS. LEVIN: Thank you for your question.

MS. BENOWITZ: Hello, I'm Brittany Benowitz, from the Center for National Security Studies. I wanted to thank the panelists for a really fascinating discussion. And I'm going to beg everyone's forgiveness, but I'm going to return to this question of pattern-based analysis. I was really struck, yesterday, that the assistant director of the FBI Criminal Division said that they do not engage in data mining, that all of their investigations are based on predication.

It echoed to me the statement -- I believe it was of Deputy Attorney General Kome during the Patriot Act hearings, that they always start with known facts and known individuals. And so, it seems to that if I were engaged as an employee, it would give me pause before I'd go down the road of any sort of pattern-based analysis. But the question today is, "What would the guidance be for good versus bad pattern-based analysis?" So, I think it's worth reviewing what is the real danger if you go down that road?

I would suggest that there's a central civil rights issue that comes up whenever you do this sort of pattern-based analysis, and I've never heard an answer to the question of what we're going to do about the problem. The problem is this. If you take, for example, the group of people who are flying first class, but they have third-class lifestyles. Let's say we run that pattern analysis and we find out that there are 100,000 people. Well, that's not a helpful group. So, then we're going to narrow it down. And in this age of counterterrorism, I think we need to be honest that, when you're going to narrow that down, you're going to start asking questions like, "Which one of those individuals has flown to Saudi Arabia in the last five years? Which one of those individuals has flown to Pakistan in the last five years?" And you're mining commercial data in order to do that. Then you get it narrowed down to 400 people. The next logical step might be to compare them with the FBI's name index. Are any of these people associated with known terrorists? And then let's say the answer is no. But then you end up with 400 people who are suspicious based on a pattern developed possibly by a red team entirely within the Department of Homeland Security based on a hypothetical threat. And you have these 400 people. What are you going to do? Are you just going to leave them out there?

I read with interest the Washington Post articles, about two months ago, about the 400 successful terrorism prosecutions post-9/11. The Post reported that the vast majority of them were for fraud or immigration charges. So what I posit to you is, if you have a

DHS official with 400 people who seem suspicious, but they don't really have any concrete known fact that they're associated with any sort of terrorist plot, what's going to keep the Department from seeing if those people have overstayed their visas, seeing if those people have ever committed some sort of petty fraud against the government, and then prosecuting them on those grounds? And if you do that, it has both civil rights implications and a national security implication, because you've then alienated the key community that we need to be in contact with in our fight against terrorism. So, my question is, before the Department of Homeland Security embarks down that road, what's the answer to that question?

MS. LEVIN: Particularly when, as Mary suggested, we want to eliminate boundaries for information-sharing within a department? How does that play out, in terms of how information is used in making decisions about people? Fred?

MR. CATE: First of all, I'm not going to answer the question. [Laughter.]

MR. CATE: It's a great question, and it's worthy of an answer. I'm just going to talk until Jim figures out what the answer is, then he'll tell you what it is. [Laughter.]

MR. CATE: I think this really focuses back to the really critical issue about use of data, as opposed to where the data came from, or access to data, or whatever. What we really care about is what is being done with it. In this case, we're talking about -- I think the term we generally use for this is "mission creep." We say we're going to collect this data because we want to keep terrorists off airplanes. We find we don't get anybody, but we've spent a lot of money on it, and a lot of time, so what are we going to do? We deport them for traffic tickets, instead. Because we had the data and now we knew this. This why you've got to come back to having those rules baked into the system. And by "baked in," I don't mean necessarily baked into the technology. This is where you come back -- if your reason for doing this, is clearly stated at the beginning when you first got authorization to try to prevent terrorism, you then really ought to be stopped. That's exactly the point about it.

Now, here's where the question, I think, comes down to – and this is why, I suppose, you never do get an answer to this question – "Who do we trust to do that?" In other words, who do we trust enough in the intricacies of political debate to say, "Well, I have spent a lot of money on this, and I didn't catch anybody, but I do have these people who -- " -- who do we, in that instance, trust ? And that's where I think you come back to oversight. And we've been talking about oversight. It was not just within the unit, and it was not the guy who came up with the idea saying, "Trust me, I'm going to really follow these rules," but oversight within the Department, oversight from outside of the Department, and oversight from the Congress.

MS. LEVIN: Before –

MR. HOOFNAGLE: I'd like to respond to that question. One of the reasons why I've been talking --

MS. LEVIN: Could you speak into the microphone?

MR. HOOFNAGLE: I'm sorry – One of the reasons I've been talking about the Privacy Act and the Fair Credit Reporting Act is that these laws create rules that help produce accuracy in databases. And when you start using databases that don't follow these type of rules, and when you can put something in my ChoicePoint file, because someone with my last name lived in my state and had an arrest record, also puts errors on the record. So, there's serious accuracy problems with ChoicePoint data-broker reports, and all of the independent studies confirm that. The studies that ChoicePoint runs on its own databases, I think, only look at FCRA databases. If they say that there were less than 1 -- it's 1 percent inaccurate. Okay? So -- inaccuracy.

And then there's another issue. And I think this is what amplifies your question. I don't know how to answer it. In addition to inaccuracy, you have the problem of not having a file. And who hasn't been filed? These databases are designed to measure commerce. So, people who are not involved in commerce have no file. And I think if you look at the people within the files, the people who are going to appear suspicious because they don't fit within this commercial mean that we live in, frankly, I think you're going to see that they look a lot like the people in New Orleans. And I think, in the oversight process, one thing we have to look back on is how much racial discrimination these databases are going to pose, because they measure commerce, and not everyone is active in commerce, for one reason or another.

MS. LEVIN: On that very sober note –

MR. DEMPSEY: Okay, if I could just -- Fred had said I'd come up with the answer, and people talked long enough. And I haven't. [Laughter.]

MS. LEVIN: We were waiting.

MR. DEMPSEY: I think that the bottom line of Brittany's question is that there is no answer, partly because one of the things we have baked into the system is a certain unfairness, which is, if the government is looking for something serious, and doesn't find it, but finds something minor, they can still ruin your life. And race and ethnicity and national origin, because of the nature of the terrorism problem today, have factored into that equation. All I would say is -- and I think it was the premise of Brittany's question -- that's why we have to be cautious with these pattern-based searches, because I think that you're going to turn up -- so, I'll just say that. I'll leave it at that.

MR. ADLER: From a technology point of view, we can protect DHS's ability to scan lots of data without ever knowing who they're looking at. You know, as John Bliss

talked about earlier, we can anonymize all that data so that DHS is just pulling anonymous de- identified information. That is, we can protect the facts. We can't protect the inferences. We can't protect against the use. But, there are lots of technical solutions for de-identifying and anonymizing the data so that DHS is, indeed, just looking for needles and not looking through every piece of hay. But it's still going to be a governance and control and oversight model that enables the government to make sure that the inference -- that the use is going to be as protected as the fact.

MS. LEVIN: Our Chief Privacy Officer, Nuala O'Connor Kelly, is going to provide closing remarks, but I think we have one more question –

MS. ACKERMAN: I think that Steve just answered it, really, the importance of oversight, based on part of Fred's response to Brittany's question about this kind of problem. You could look at –

MS. LEVIN: Could you identify yourself –

MS. ACKERMAN: Sorry, Linda Ackerman, Privacy Activism. You can take TSA over the evolution of CAPPS II and Secure Flight as a poster child for the inappropriate use of commercial data, probably for one of the reasons that Jim gave -- that they've never fully articulated a plan for the use of that data. But, you know, it's too late now to start to address the importance of oversight, or the means of oversight, but at least to acknowledge the necessity of oversight in any scenario for privacy protection with the use of commercial data. So, a statement, ultimately, not a question, because Steve just really answered it.

MS. LEVIN: Well, a very important statement. Thank you very much.

nology and our people can bring together.

I want to thank you very much for coming today and thank the panelists.

[Applause.]