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

Thursday, September 8, 2005
Auditorium
GSA Regional Headquarters Building
7th and D Streets, SW,
Washington, D.C., 20024

PANEL ONE
HOW ARE GOVERNMENT AGENCIES USING COMMERCIAL DATA TO AID IN HOMELAND SECURITY?

Moderator:
Ms. Anna Slomovic

Panelists:
Ms. Jennifer Barrett
Ms. Carol DiBattiste
Ms. Grace Mastalli
Mr. Jeff Ross
Mr. Chris Swecker

MS. SLOMOVIC: Good morning, everyone, my name is Anna Slomovic, and I'm moderating this morning's first panel. The purpose of this panel is to build our knowledge base about the commercial data the government uses, and how it uses such data. We're going to start with a little bit of definitions, something about -- what do we mean when we say commercial data? Do we mean transactional data, do we mean data from public records, all of the above, some of the above, what is the government actually buying? And what are companies actually selling? I think that will be important for setting the stage for the rest of the discussion.

We will also look at what benefits are received from the use of this data, and how these benefits are measured. We will talk about whether or not there are different benefits that can be derived from different types of data, or different types of applications. We will address data quality issues, how the quality can be measured, and how this quality is

tied to the benefits. We will also learn something about the contractual relationships between buyers and sellers, and in particular, about the privacy and security provisions in these contractual relationships. If I have to frame the main question for this panel, the question is, "what is?" Subsequent panels will look at what could be, or what should be.

So, before we get started, let me also introduce a note of caution, there may be contractual provisions or specific discussions about data types or data uses that our panelists may not be able to address, either because they're bound by contract, or because there are national security considerations. So, let me put that issue out, up front, and ask you to please be sensitive to that.

With this, let me introduce our panelists, you have their bios in your packets, each panelist has had a distinguished and varied career, and is in a position to talk about first-hand experience with the use of commercial data in government applications. We have asked each panelist to prepare about five minutes or so of introductory remarks that will set the stage, and tell us about the organization, what data they either buy or sell. We will spend the bulk of this panel's time discussing questions related to both the buying and selling and using commercial data, and then we will have a significant amount of time for questions from the audience, and I hope you will ask a lot of good and interesting questions.

With this, let me turn to the panel. We have Ms. Jennifer Barrett, who is the Chief Privacy Officer for Acxiom; we have Ms. Carol DiBattiste, who is the Chief Credential and Compliance and Privacy Officer at ChoicePoint; we have Ms. Grace Mastalli, who is the Principal Deputy Director for the Information Sharing and Collaboration Program in the Department of Homeland Security; we have Mr. Jeff Ross, who is a senior advisor in the area of money laundering and terrorist financing, in the Office of Terrorist Financing and Financial Crime at the Department of Treasury; and we will have a late arrival, Mr. Chris Swecker, who is the Assistant Director of the Criminal Investigative Division for the FBI. Hurricane Katrina re-arranged the schedules of a number of government officials, and I'm afraid Mr. Swecker is one of those officials. He will be with us as soon as he can get out of his morning meeting.

So, with this, let me introduce Ms. Jennifer Barrett.

MS. BARRETT: Thank you, I'm delighted to be here, I want to thank Nuala, Toby and others for putting this conference together, I think it's a timely topic and one that bears a lot of discussion. You probably have seen most of the information up here right now, about these issues and how commercial information is being used by the government. Acxiom is a little bit of an unusual company, in that we both are a data provider to commercial and to the government sector, as well as a service provider, we offer computer services. And the services help our clients understand and make better use of the information that they already have. And so we see information from two sides,

one as a provider, and one as a service provider to help our clients use that information. It brings some clarity to what information is appropriate for what purpose.

For 35 years we have been delivering these services, but more recently to the government sector. Only in the last five years have we been engaged in this way. But we are engaged and exposed to doing the same thing that we have historically done for the private sector, and that is, again, helping them to identify errors in information that they have, and helping them to accurately integrate information from multiple sources, or to do comparisons of information from multiple sources, and then also offering any information we have with either intelligence about that data or in some cases additional information. As the questions go on, we'll dig into that in a little bit more detail.

One of the things that's often misunderstood about our business and other information providers is that we have this one big database that has everything on everybody in the U.S. about it, and that is not the case. We build specific information products for specific applications and specific uses, tailored to use the database product, or that information, and that tailoring includes -- clients would like us to offer the data sources that are being used, as well as both legal, and our own judgment call about what information is appropriate to be used in certain instances for certain decision-making processes. Also, we take our responsibility seriously as the information provider and the custodian of information about consumers, some of which might be sensitive, and so we have a robust audit process that we use both internally in our company as well as the clients that we sell it to, to use it.

I'm just about through with my introductory remarks. But, suffice it to say that I think what we've seen in the last five years is a set of similar needs from the standpoint of information management in our dealings with government agencies that we have seen in the commercial sector, and I think that the government has an opportunity to play some catch up and actually take advantage of the products that have been in use in the private sector for many years, but have not been as widely deployed in the government sector. And with that I'll end my introductory remarks. Thank you.

MS. DIBATTISTE: Good morning. ChoicePoint is both a data provider and also, I would say, a technology company, whose main purpose is to provide information to whoever the customer is -- whether it's the government, or insurance services, or Fortune 1000 companies, or marketing services – in order for the customer to make decisions to make the world a safer, more secure place, and really a better place.

The information that ChoicePoint provides is really in four different sectors, and then I'll talk a little bit about the technology piece. And that is, insurance -- insurance services provides insurance carriers with information to provide consumers insurance -- whether it's personal property insurance, or casualty insurance, or commercial insurance. In the business services vertical, ChoicePoint provides information to companies to do

background screening for their employees. It's called our Workplace Solutions, it's part of ChoicePoint, and that information is provided to screen hundreds of thousands of employees of major companies, in order to ensure that their employees are who they say they are, and that they can put the trust in them, the trust that is required for them to do the jobs that the companies have hired them to do. And that Workplace Solutions is both from a commercial side, and also supports the government in doing background screening for government. Also in the business services, what I would call, vertical, ChoicePoint provides information, public record data information to help businesses with identity verification, with authentication of customers and their possible consumers who, let's say, want to have a financial transaction -- want to get a mortgage, want to ensure that they're credit-worthy -- and in that business services vertical, ChoicePoint will provide information, let's say, to a financial institution, so that financial institution can provide a service to the consumer, whether it be opening a bank account, getting a mortgage, or any other kind of financial transaction.

In the government sector, which is not really new to ChoicePoint, but with the acquisition of several companies over the last, I would say, few years, it's grown in the government sector. ChoicePoint provides data services to the government in order for the government to do criminal investigations, in order for the government to do victim identification, and I'll talk about that in just a moment; for homeland security, for border protection, for immigration services, for entitlement programs for disaster assistance, for critical infrastructure protection. I would put the buckets in the government sector in three -- federal law enforcement would be one, homeland security efforts would be two, and entitlement programs would be three. And then in the marketing segment, ChoicePoint provides information for marketing purposes. That's very limited, and that's probably the smallest portion of the ChoicePoint business.

But let me go back to the government, because that's why we're here today, and in addition to providing data for law enforcement purposes and for homeland security efforts and entitlement programs, in many phases -- and we'll get into that in a little bit more detail. ChoicePoint also provides technology that assists the government in many respects, such as IMAP Data, which is a mapping technology which visualizes data and helps, right now, in the disaster recovery efforts regarding Katrina. But ChoicePoint also has another technology company called i2, and I think you'll be hearing a little bit later in the program from Jack Reis, who is the president of i2, and i2 is an analytics tool that federal law enforcement uses, and intelligence community uses to work on their cases to provide analysis, and critical analysis, in linking suspects or linking key pieces of information together, so they can identify and solve cases.

So, it's both a data provider, a data user, and also a provider of critical technology that assists the government in their efforts every day. And I wanted to tag on a little bit to

what Nuala said before I close, and that is that this data also helps in the hurricane Katrina disaster relief. Presently, we're providing visualization and mapping technology support through IMAP Data to assist in the establish of relief shelters, as well as cataloging infrastructure damage. We're helping to validate the identity of victims that can receive assistance through hurricane Katrina, and that means payment and benefits through online, as well as telephone call center applications, so the person that needs assistance, we can verify they are who they say they are, they really did live in New Orleans, and that way they are entitled to the benefit that they are going to receive. We're also providing vital records for the people that have lost their homes, and some have lost their lives, and also providing screening for the volunteer organizations that are now hiring up, of course, and getting several volunteers to help them have the volunteers who are really going to assist in this effort, and not try to take advantage of this effort to fraudulent activities.

MS. MASTALLI: I'm going to provide caveat in addition to what Anna said. Not only are there constraints placed by national security, but they're also law enforcement sensitive information that we can not and should not share. Therefore, I'm going to speak in generalities and of past well-known specifics throughout this panel.

Carol actually did a superb job of summarizing how homeland security, law enforcement and intelligence agencies use some data, with particular emphasis, right now in that were it not for the commercial data sources, we would have wiped out government records in three states. There are people without prescriptions, without driver's licenses, and it is the commercial data sources, in many instances right now, that are facilitating, not just placing people, but verifying their identities to the claims we get to make sure that entitlements go to the individuals who deserve them. In general terms, the Information Sharing and Collaboration Office is cross cutting DHS and works with all of the components of DHS, including the Privacy, Civil Rights, and Civil Liberties Offices, trying to coordinate not just information sharing, but data management, and data acquisition activities to assure that the multitude of different ways that data is acquired or used, are done so effectively, efficiently and in full compliance with the law. Unfortunately, there have been occasions when we explain to well-intentioned operational staff that just because they could buy data from a commercial aggregator did not mean that they did not need the same probable cause that they needed to get a piece of paper with the same data. It's an educational process.

Carol actually did a wonderful job of characterizing how data is used across the government. She didn't mention that we have sometimes used commercial data, not just to support identity authentication, but to assure the integrity of government data, and the accuracy of government data.

Unfortunately, in many respects, the commercial enterprises have done better jobs of organizing and, what I call "cleaning" data to eliminate errors in the data. It's not -- I won't say it's true of all data brokers -- but it's true of many. And, it's something that, we sometimes use commercial data to verify the accuracy of government-collected data. We also use it for research and analysis. On law enforcement and intelligence, it may be used for link analysis, for trend-pattern analysis. It's used extensively for disaster relief, for public health services and entitlement. It's used to help with visualization, it's used to help conduct risk-based analysis of threats and natural disasters, and to prioritize the allocation of extremely limited federal resources. It's used for prevention, protection, rescue, as well as for information, just as you use it. The services for research, for legal, financial and other data that are used in the private sector are also used in the government for exactly the same reasons, although sometimes there are more hoops for the government to go through in order to access it. Rather than belaboring the generalities and anticipating what Jeff is going to say, I'll summarize there, then, because Jeff and I have worked together in the past on a number of the ways that commercial data can be used.

MR. ROSS: Thank you. Can everybody hear me in the back that wants to? Okay. First of all, Nuala, thank you, and thank you, Toby, thank you for having me here on this panel. I'm actually quite looking forward to this panel as well, to get some better perspective on a broader range of how commercial data is used by the federal government, and give you some background on how we, or what I consider as part of the Treasury Department, use it in an office of -- fairly new offices owned by the bureau called Terrorist Financing and Financial Intelligence. It is comprised two parts, the one part is the policy side and the operational side, which I'm on, which we come and look at things largely across the country and internationally. Unlike many other agencies in the government, and unlike the one I was at, at the Department of Justice, we're kind of one-stop shopping for many crimes. So, whether it's a narcotics crime that has a financial component -- and they all do -- or whether it's a public corruption crime with a financial component -- and they all do -- or whether it's a terrorist financing crime, or almost any other crime, organized crime, my office is the one that's responsible for coordinating the policy position of the Department of Treasury, and also for using various techniques and powers and authorities we have under the Bank Secrecy Act and the Privacy Act to attack those.

The other side of my office, is the Office of Intelligence and Analysis -- it's an interesting office in that it is kind of a counterpart to the policy side, it is the financial intelligence side of the Department of the Treasury. But prior to the creation of this office, there was no central financial intelligence component -- there were various parts of financial intelligence in various places, but they were all put into this one place now, and put under the authority of one particular secretary. Why is this important? It means that

where there's classified intelligence, and we have this commercial database type intelligence, the source intelligence, law enforcement intelligence, intelligence we generate internally through our ability to access our own internal databases -- all of that information can be put into one place, filtered, analyzed and then passed over to the policy side, which is my side, to work, and go forward and then work with the law enforcement agencies, and vice versa. As policy identifies the area, we can go back to the financial intelligence people and ask, please tell us what you think. So, it's a very interesting office.

So very quickly I will tell you who reports to use, and that will give you some flavor of the types of commercial data that we use, and how we use it. First, is from an operational perspective, IRSCI, the criminal component of the Internal Revenue Service, doesn't report to my office, but it's kind of a dot-dot line, working directly with me on more than a daily basis, I have a liaison there, and we both work directly with IRSCI. IRSCI -- just as many other government agencies on the law enforcement side -- works crime in from two different perspectives, and it's important that you remember this, because the use of the data -- although it may be the same data -- it could differ based on how you're approaching what your assignment is. One way to use commercial data, or to use any kind of data, whether it comes from law enforcement, from intel, whether it comes from commercially available or other public databases, is if you have a particular target in mind, and we're working the target, but as a component of that target, you're looking at the financial side. So, you have a target; what do we know about the financing? What do we know about the associations? What do we know about the movement of assets? What do we know about the wealth of this individual? Particularly in an IRS context it could be important, because they will explain wealth is a key indicator of possible tax issues. We work it that way.

Now, the second way in which you can work a case is a proactive approach, and there, you're actually working the data that you have, to try to extrapolate from that who the targets might be. So one way is to say you've got like a tree in a forest, and you're working the forest, you've got a way to work the forest to try to get a bad tree. We do that from both perspectives from the Treasury Department, depending on what particular assignment it is, and what types of crimes you're going after. For example, unexplained wealth is a perfect barometer to view proactively, to try to identify a target. Whereas, associational relationships, asset transfers, ownerships of assets, that sort of thing, works better when you've already got a target in mind, and you're trying to get lead information for the investigation. So IRSCI reports to us -- well, doesn't report to us, works closely with us.

Second, reporting to us is FINCEN. FINCEN is the financial intelligence unit for the United States, it is the one responsible for enforcing the Bank Secrecy Act across the board

to all financial institutions that are subject to intervene under the Bank Secrecy Act or the Privacy Act. It is the institution that supports all law enforcement and financial crimes, it makes no difference if it's the EPA, DHS, CDP, FBI, IRS, CIA -- it supports them all, equally, and provides platforms of information for them, so it then has databases available to them that they can use to try to work their cases, again, either proactively, or when you have a targeted money. Secondly, also you begin as proactive targets, it takes the financial databases that it has, the Secrecy Act Databases that you're all familiar with, suspicious reports, large base currency transaction reports, in-bound, out-bound currency movement report -- analyses those and tries to extrapolate out from those what might be criminal activity so you can pass that to the field at the headquarters. There again, it's going to tap into commercial databases to add value to what's already there in the field.

A third agency reporting to us is OFAC, Office of Foreign Assets Control. When you hear about new global terrorists "designated" by the United States, by the department of the Treasury -- OFAC is the institution within the Treasury and USG that implements this. Not only does it do terrorist financing and terrorism designations, but it also does narcotics designations. Particularly with the drug kingpins worldwide, in what's called the Kingpin Act, it also works on the specific program dealing with Colombian narco-traffickers. OFAC, again, is trying to add value to what it may know. DEA may know that this person is a trafficker, but if we wish to designate a particular trafficker, the way to get the bang for the buck is to find out if that person actually has invested in the financial infrastructure. If the person is just wheeling money back and forth across the border, it might be a nice press release, but it's not going to have any effect. If you can find someone who's actually invested in City Hall, if it's in Colombia, tens of millions of dollars in the financial infrastructure by working with the Colombians on their data bases down there, you can get at that infrastructure, you can designate it and cut it off in the U.S. in one fell swoop, which is what we've done.

So, commercial data bases are very important to us in the law enforcement area to be used proactively, we've tried to identify it, secondly, we have the targets and need information, where you are trying, also, to find a specific individual or entity that should be involved, what associations are, who could also be potential witnesses in a case, so in a financial crimes perspective, these data bases are extremely important to us, and we use them in a variety of fashions in the treasury. Thank you.

MS. SLOMOVIC: Thank you. We can continue discussing the benefits. This was a very nice segue way, Jeff, into the first question which is -- what are the benefits? Why is the government buying commercial data? What benefits is it gaining? And how do the benefits of using commercial data compare with the benefits of using government data that has already been collected, or that may be collected? Given that privacy protections -

- legal privacy protections -- that apply to government data are different than those that apply to data collected by the commercial sector, for commercial applications. Grace?

MS. MASTALLI: One issue is that until the government does a better job of sharing and authenticating data, we often get more accurate data from the commercial sector. In addition, the processes by which government agencies manage data, often makes it difficult to acquire, and needs great deal of labor intensity into making it usable and accessible to other entities. So, in one respect, the commercial data providers provide accurate data -- often more accurate than some that we have, because they spend the time cleaning it and verifying it, and have matching capacities that we in government often have not yet invested in to eliminate the 17 instances of an individual who has a phonetically spelled name being recorded as 17 people instead of one.

There's also, frankly, data available from commercial sources that we find we need that we never thought we needed until the moment we needed it. One extremely well-known law enforcement intelligence example from immediately post-9/11 was when there was a now well-publicized threat - - so I'm not violating any national security constraints - - that there might be cells of terrorists training for scuba diving underwater bombing, similar to those that trained for 9/11 to learn to fly -- but not land -- planes. How does the government best acquire that? The FBI applied the standard shoe leather approach -- spent millions of dollars sending out every agent in every office in the country to identify certified scuba training schools. The alternative could and should have been for the Federal government to be able to buy that data for a couple of hundred dollars from a commercial provider, and to use that baseline and law enforcement resources, starting with the commercial data baseline. One of the issues here is that, other than the name of the owner or manager of scuba diving schools, there was no personally identifiable data in that instance. And I give that as an example. There are thousands -- chlorine, paper trucks, et cetera -- data that is maintained for commercial purposes that we have tracked, not even thought of being needed, that could be needed for terrorism protection, prevention or response to a national disaster. In a well-known -- and Carol alluded to it -- you know where we got the best information about the location of levees across the country in the Gulf several weeks ago? We got it from a commercial data provider, because it was better than the data that we had in the government.

MR. ROSS: If I could just add to that, I mean, I look at the commercial data as adding flesh to the bones of whatever it is you're working -- the law enforcement perspective, you have certain government databases available to you, particularly the Bank Secrecy Act, you have mandated reports that have to be filed, either by banks or other financial institutions that will identify information about particular transactions, and you'll know a certain event happened on a certain day in a certain place in a certain amount, that's good, that's kind of the bones. But if you want to add a little flesh to that

particular transaction, you want to learn a little something more about a particular individual, you're probably going to go to a commercial database, if you want to find out information about a particular individual's activities, again, you may have to go to -- so what you're doing is you're adding flesh to a skeleton.

And I think Grace implied -- the Federal government has a problem, a fancy term for this, mapping, they don't map very well. And where you need your map is, is where you can take a data field here -- and I learned all of the mapping working with the FBI after 9/11, and it wasn't the FBI's fault, they were trying to map different data, but you can take data fields from here to here to here and they all translate to one place, that if you put it in one central place where then, you can examine all three pieces. Well, the federal data doesn't map too well, DEA's intelligence database, and FDS database and they don't work all that well. So they will be, these individual agencies, will be investigating crime and utilizing their own databases regardless -- that's just going to happen. But what you can do is, for instance, in a mission defense center, the center has a platform of data that is available to all U.S. law enforcement agencies, and a lot of state and local that are doing criminal investigations. And this allows them to uplift certain data that defense maintains, and then apply it to whatever investigation that they do. So regardless of how the sea elephant is configured, the FBI or DEA or IRS, the same skin can be added to it, based on what their particular needs are, so I think that's a particularly important aspect to remember.

And you've got to remember, also, these different agencies are investigating different types of crimes. Again, and I know I sound like a broken record, but I'm in law enforcement, so I can't help it. DEA is a one-crime agency, it's drugs, that's what it does, that's what it's assigned to do. FBI is the lead in terrorist financing -- for terrorism generically -- with terrorist financing being kind of a subset under the broader terrorist umbrella. But once you get beyond that, you can get into money laundering and into cross-border crime, DHS and IRS both have extremely broad financial kinds of authority, and IRS, of course, is the only one that has tax crime authority. So each one is going to be using the particular types of commercial data to support particular types of investigation, so that even if the same data is available, defense and IRS engaged in this part of it, DEA maybe interested in that part, and FBI may be interested in none of it, depending on what the investigation is. So, you've got to remember that, as the panelist said, it depends on what the particular purpose is.

And, just to conclude quickly, what is very interesting -- and Grace alluded to this – we post-9/11, when we were over -- I was with Justice when 9/11 happened, I was training with the FBI to help set up a inter-agency working group that was focusing exclusively on financing, that's how the whole terrorist financing thing got started -- and we realized what we didn't know, we didn't know what choke-points were for some of

this stuff, in the commercial area. We didn't know who the best providers might be-- we need to learn from scratch who might be the best commercial providers to tell us exactly as Grace said. In the case of 9/11, it would have been nice to have a universal baseline on all flight schools in the country. Well, we didn't go get that, it wasn't there, it wasn't available. I can assure you the FBI has it -- it's available, they know. As you learn, then you say, you know, this is interesting -- you take this piece of information and apply it to this, and we can extrapolate something. The only way you're going to find that out is when you're actually working the investigation of the cases. It's kind of an ongoing process in anything that's ongoing today.

MS. MASTALLI: If I may, I want to re-emphasize something Jeff said, which is, we've learned a lot. One of the things we've learned is, we've learned that not all commercial data providers are created equally. And, we've gotten smart about what providers have listed. For four years a man living with his wife and ex-wife because they didn't clean the records, then who had accurate data. So, one of the things that may have happened immediately post- 9/11 in terms of commercial, use of commercial data by government, was agencies went out and they looked at everything and they treated all providers somewhat equally. Some agencies have learned the hard way that not all commercially-provided data is of equal quality, just as we know that all government data is of equal quality.

MS. DIBATTISTE: I would put the benefit of commercial data -- I would say there are three major benefits, and then possibly that -- not possibly, definitely -- that would lead to a fourth benefit. The first benefit, I think, of commercial data is what, well, actually, I think they commented on all of them, but the first benefit I would comment on is time saved. It's faster to use commercial data. And Grace gave the example of the scuba divers, I would give an example of -- again, related to Katrina -- we could send out people, or the relief agency, FEMA, could send out people to try to locate every doctor, every nurse, every pharmacy. And that would take weeks, if not months, to do that. And commercial data provided that to the relevant people in a matter of minutes. And that is a very valid example based on the times, and what we're dealing with today. I've also gotten some other examples for law enforcement investigative purposes -- what would take a law enforcement agency -- and this is through no fault of their own -- to go out with boots on the ground to get something, it might take four to six weeks, where commercial data, the capability there of getting it in four to six minutes. That's how extreme it is. I gave you the Katrina example, I can give you the investigative example, but it is much faster.

The second, I think, major benefit is that it is better quality, or more accurate, and again I think both Jeff and Grace touched upon that. And that is because it can be more precise, because the more information available, you can hone in and do matching to

actually determine who is the person, where is the place, and any other specific information that you're trying to glean to ensure that, let's say, that you're in an investigation, that you've identified the right person -- the more information you get -- and say you want to know where that person is living, you can go to commercial data and find out exactly where that person is, instead of going out and doing a boots on the ground-type of investigation, or doing phone calls, or looking at various records -- you can use commercial data to hone right in and get accurate and better quality data to help assist in the investigation.

MS. MASTALLI: And to make sure the warrant, when issued by a judge, is valid, because the information is correct.

MS. DIBATTISTE: And the third, I would say, is it's more current-- I think Grace, and Jeff again, touched on that, that the currency of the commercial data is faster and better. And for those three benefits, I believe there's a fourth benefit, which may be the greatest one in the forum that we're talking about today -- I believe because it's faster, better quality and more current that the benefits, a large benefit is, that it protects and minimizes the impingement on civil rights, civil liberties and privacy, because you're not doing these broad-based searches that are involving hundreds of people, you can hone right in very quickly on one, so the benefits to privacy and civil rights and civil liberties become greater when you use commercial data responsibly, and carefully, and those three benefits then turn into a fourth benefit, because obviously you're reducing the amount of false positives also, with commercial data.

MS. BARRETT: I'm not going to -- some very, very excellent points made by the other panelists -- I'd like to hone in on two at the moment, and dig a little deeper. One is this area that I think Grace first brought up in opening remarks and Jeff commented on, and that's the accuracy of information. Any more, today, any use of information requires combining information from at least two sources, usually dozens of sources, or comparing information in one source to another source to try and make some determination. And the accuracy of the information that you hold, and are getting ready to engage in either of those processes is critical to the outcome. And sometimes it's very simple, basic stuff, like knowing in a high-rise apartment that there are 12 individual family units with the same first initial and last name. So, if you don't have a unit designator, you've got a big problem. Now, that may not be -- that problem may not exist in rural America where we don't have the density, of the same names, but it may, because families tend to live near each other in the country, as well as in the city. So, very, very simple information like knowing in this particular location, I need to have an apartment number before I can make some combination of data, some mapping of data, to use Jeff's term. Or some comparisons of data against some, you know, an OPEC list or a watch list or something else becomes extremely critical.

The private sector took probably 10 years to learn that you spend your money first on getting the data you have right before you start engaging in all kinds of sophisticated mapping and integration and analysis exercises, because if the data's brought together wrong, everything that comes after that is wrong. I think the government is beginning to learn that, and they should take lessons from the private sector, they certainly have not been there as long we did trying to figure it out, in terms of focusing on the data you've got before you start down the process of using it -- whether it's using it for your own purposes, or whether it's using it for some comparison to commercial data. And commercial data plays a role in that accuracy.

The other piece of it that I'd like to touch on is the part that Carol alluded to. There are those that claim that the private sector, you know, doesn't operate under many rules and regulations when it comes to data. I think I can certainly speak for my company, I think I'm probably also speaking for Carol to some degree -- that we feel very, very regulated in many, many aspects of our business. There are laws that regulate us, also regulatory guidelines that we have publicly ascribed to, and therefore are subject to Section 5 of the FTC Act if we don't follow those guidelines, and then we have our own stated policy. And the law gets very, very complicated in terms of what data can be used for what purposes. We pass those restrictions on to our clients, and it is appropriate that our clients understand those. But I think that the private sector, and the information providers that have lived under these rules, actually provide a benefit to each other in helping them understand what the rules are around these private sector data. I think there's also things that can be learned in terms of how we manage those restrictions, because as stated, integrating across agencies, whether it's a combination of private and public data, or just public sector data, the same kinds of restrictions, policies, processes that need to be sure of following, need to come into play. And I think there's an area where we really need to work together and achieve a much safer environment for the consumer, and assure a much more appropriate use of information.

MS. SLOMOVIC: Let me ask a question related to this. Commercial data is collected for particular purposes -- people do transactions and they provide certain data to complete the transaction; people have interactions with their public officials, and they provide data for those purposes. When that data starts getting used for other purposes, the word "accuracy" changes its meaning. It may be accurate for one transaction, but not for another. In some cases, the two transactions may be sufficiently close to each other that it doesn't actually matter. But when we start talking about especially pattern matching for patterns where there really isn't a whole lot to go on -- unlike billions of credit card transactions, there haven't really been billions of terrorist incidents, thank goodness -- how do we know that the data quality of the data that was collected for one set of purposes is actually appropriate data quality for a completely different set of purposes?

MS. MASTALLI: I'm not sure I fully understand your point, Anna, because, if you took -- for example -- airline reservation data. It was either a person using what may be a false name, or not; may have made hundreds of reservations using the same credit card number, or not; I mean, the accuracy may not be that it really was just Smith who made the reservations using that credit card number, the identity may be wrong -- but the name, the credit card, the reservations made should be accurate. And, if you're looking, taking the example, the past history, since it's easiest to talk about past history -- if we had looked at anonymized airline reservation data pre-9/11, there would have been information available from the patterns of reservations made using one or two single credit cards that may have been then provided to intelligence and law enforcement -- without ever having looked at anyone's individual data -- the extracted patterns could have then been provided and led to investigations that would have uncovered individuals, but you would have, by then, had the legal basis for getting named individual information.

MS. SLOMOVIC: You're assuming that you would have known what pattern is the pattern that should be investigated.

MS. MASTALLI: I'm assuming that we can make informed and intelligent projections, based on past history, and that it's an evolutionary process. We learn more, we do red teaming to identify new possible patterns, we test them in theory, and we apply what's been learned about terrorism, for example, not just within the United States, but around the world, historically, and then you apply that knowledge in a fashion that impinges on no one's individual privacy, and test it in a living laboratory, if you will. And then -- Jeff wants to jump in, so please do --

MR. ROSS: No, no, I think it's a pertinent question, I give an excellent example of that, again, in 9/11 -- it was so horrendous and such a shock to everything that for a period of time we thought the whole world was just gone, which was good in a lot of ways, because people were cooperating much better than normal on this, but the point is, for instance, in the case of the hijackers. The hijackers -- I talked about two different ways to investigate cases, prior to the case, and the reactive case where you have a target in mind -- and the 9/11 hijackers are the best examples I can think of reactive, because they're dead. So then what you want to do is, okay, we know these guys are dead, we know who they are -- what can we find out about how they operated financially? So then you start to work on this, and as Grace said, then you're going to get data. So, let's say, the credit card data that you got on Mohammed Atta shows that what you had was an individual living a completely third or fourth class life, with first class air travel. Well, so what you've started with is a target, you've now identified a set of factors, now you can go proactively in the future, and this is what FBI did. As we would identify patterns of behavior that hijackers used, then we could try to extrapolate out to what the future -- assuming that

they stayed the same, and we're not so stupid to say that they will -- but if the same sorts of individuals are operating currently, and that's what we were looking at in post-9/11 -- we were looking to see if there were more terrorists of the Atta ilk in the country we didn't know about. And we were using the financials to try to identify patterns, and then go to the industry to identify if these patterns are being seen anywhere else.

Another example is when they used their debit cards, their ATM cards -- they basically lived off of those things. And, we just assumed that at least, initially, when an ATM card was in a city, Atta was in the city -- not true. The card was in the city, because these guys passed these cards around, and anybody could use them at any time. Another fact that you can then use to try to extrapolate future behavior. And this is exactly what Grace is saying -- you start from a point, as you do your investigation and you identify particular habits, and in that case, I'm looking at purely financial, but there are other things these guys would do, maybe the way they book their reservations, they would book one at the last possible second, for instance, they would use cash -- whatever additional factors could be added. Then you could project future conduct, and from that you could then tailor what types of commercial data you want.

The other point I want to make is that commercial data is going to be the same, all right? So, let's say commercial data from the scuba diving schools show who registered to take scuba diving classes across the country, let's say that such a database exists. But the value of the scuba diving database -- even though it's identical -- is going to differ based on what the user knows already about that particular individual. If, for instance, the FBI happens to know that so-and-so uses three aliases -- the DHS doesn't -- the FBI is going to get more value out of finding that alias, and that list, than DHS will. So, it's going to differ depending on what it is that the particular agency knows -- if you think about drugs, the DEA is probably going to have the greatest data base. Of course, in terrorism, the FBI should have the greatest database. If it involves the cross-border movement of smuggled goods, I would think DHS probably should handle the data. So, even though the commercial data can stay the same, the value to the particular agency or entity using it can differ, based on what that agency or entity already knows.

MS. BARRETT: If I could, maybe, add another dimension to this question of is the data appropriate and accurate enough for a particular use? I think it's very important, and responsible data providers do this, but I think it's also saying what our client or user, in this case, does to ask about the source of the data, to ask about the level of accuracy of the data. And I'll give a real-life commercial example, because it's easier to talk about this one that's easily transferable: We build products for marketing purposes, and we build products for identity verification. Some of the same sources of data go into both of those products, but many of them do not. None of our commercial clients would ever use our marketing data for identity verification application, because it is not designed to be

accurate enough. It is not current enough, it may not be complete enough, et cetera. It is fine for marketing purposes. Because the downside in that example is, you either are not, or you are, going to get a piece of mail, or a call or whatever, that you might not have wanted. So, the decisions that are being made on the data, and when the data is being used in an automated fashion to make a decision, or whether it's being used to augment something like an investigation, where you're putting intelligence and other known facts and bringing them to bear to help sort out what's the right answer, drive very specific requirements for how accurate is the data.

Now, I'm going to make a statement, I make it frequently, I make it on Capitol Hill, and no one likes to hear this statement, but it is the truth, so it's where we live in -- nobody's data is 100 percent accurate. So, what we have to do is figure out -- what is the level of accuracy in the data, and how do we mitigate against the weaknesses that the inaccuracies create? And the worse the data, the more mitigation you have to pay to play. I'll give another kind of example here -- you're not going to use data that you have significant questions about, or concerns about, for making automated decisions -- decisions that are black and white -- that someone is directing, one direction versus another -- you might create a suspect category. When information is incomplete, or when information is, when there's multiple responses to a particular query, and you're not sure which is the right one, it can cause further investigation, or further analysis to come to a good decision. So, we build systems and we build processes to take this data and to try to compensate for the inaccuracies that are inherent in the information.

I'd also like to say that the processes that agencies use for evaluating the inaccuracies are a big area where, again, they can learn from the private sector. The private sector has spent billions of dollars trying to find out, how do I determine what is accurate, and what is not? And so, it's a question that ought to be asked of any data provider when they're signed up, if they don't give you a good answer, it's an indication that they probably don't know, and to what Grace was saying earlier, not all data providers provide the same level of accuracy, and so it's one that really needs to be focused on.

MS. SLOMOVIC: Given the difference in data quality for different applications, and the fact that you build different applications differently -- is the government asking data providers to collect certain kinds of data? Or are data providers proactively looking out to see what kinds of data might be useful? It sounds like in many cases, the government doesn't know what data it might need until it needs it, and so I'm trying to understand a little bit about how the data gets collected, how do we know what to collect if we don't know we may need it for awhile?

MS. BARRETT: Well, we get asked for that all of the time. A lot of times we don't have it, a lot of times we can't get it, a lot of times we tell the client -- whether it's the

government or the private sector, it's illegal to provide that data for that particular use, because we are familiar with the laws and restrictions around it. If there is enough of a demand and a need for certain kinds of data, and it's economically feasible to collect it, I think data providers will respond to the demands in the marketplace. But we have to understand those demands. We don't sit around with a lot of data in our pockets, hoping somebody's going to come around and ask for it. It's expensive to collect, it's expensive to clean up and maintain, and even though the storage is getting cheaper, that's the least of the costs in doing data collection and integration and then building new products.

So, it's one where I think the private and the public sector need to work together, if there is information that is not being collected, and can be, and is available, I mean, availability is sometimes a problem. Often we have to tell guys to the agency to the effect that the data is just not available, it's just not practical to collect. So, to a large degree, data -- the nice uses that Grace described where you have a need, and oh my God, guess what? Somebody's got it -- unfortunately come more by accident than by any kind of planned strategy.

MS. MASTALLI: I was going to make the point -- I think that there are two different categories -- I think that data is collected for the collectors' purpose. I think that what we, the government, often do is we work with data aggregators or data brokers, and say, "Do you have x?" and so it's not a question that we ask anyone to collect it for our purposes, we ask anyone who has collected it for another purpose. Traditionally, if it is needed for a governmental purpose, the government will have been directed, authorized to collect it. And so sometimes what we want to do is match the commercially available data which the government hadn't thought it needed, to government-collected data. But I'm having a hard time thinking of ever being aware of an agency wanting a private entity to collect data, as opposed to finding out whether or not they can acquire it and provide it in an easily useable format.

MS. DIBATTISTE: And I would piggyback onto what Grace said, in dealing with the government agencies, the types of commercial information that most of the government agencies are looking for -- although I would agree -- no one is coming and saying, "Collect this specific type of public information," but it would generally be information such as address, and address history, phone history, real property records, drivers' license data, vehicle ownership, bankruptcy liens and judgments, because that's how -- and that's all publicly available, phone listings, business filings, and business directory-type of information, yellow page information, phone and business listings -- that's a general, broad categorization of public/commercial data, and then that data is used by the agents or the investigators or whomever, and I would actually tack onto that critical infrastructure data, where there is critical infrastructure, and that includes everything you can think of, and then the data behind that critical infrastructure - - if it's a

school, what's the data behind that school, where is it located? Who are the key players in that school? And that, obviously, helps in an emergency response, but the public record data is not something that the government agencies -- at least to the best of my knowledge, is coming to the data companies and saying, "Do you collect this? I need this specific piece of data." But they are not telling us to go and collect certain data.

The other key piece is the linking of the data. The tools -- and that piggybacks, really, onto what Jeff said earlier -- each of the government agencies, and I'm talking law enforcement now, might have information that they have particular to an investigation, they need commercial data to refine that information, to help them key in on, is this the right person? Is this the right suspect? Who is this person or suspect linked to? Who do they live next to? Who are they calling all the time? Once they've identified and honed in on a suspect, and agencies now have the ability to link that information. The agents working on different Federal government agencies can link it through certain analytical tools, and that helps them also increase the value of their investigation, and the speed, and the accuracy again, that we talked about earlier.

I wanted to make one other comment on Jeff's prior comment, maybe it was Jennifer's, actually, on the accuracy piece -- and that is, there is, to help on the private side, to help ensure that the data is accurate, there is what is called advanced pinging technology, that helps you refine, that you're sure that that piece of data that you have received is accurate, and there's also business rules that are put in place regarding matching, so you're matching that specific address to that specific individual, or that specific piece of data to that specific individual, business rules go into that matching, and also technology goes -- advanced technology goes into making sure that that data was accurate when you were providing it.

And the final piece I want to add is on use -- I viewed your question a little bit differently on use. When we deal with any government agency, the agency does have use limitation in any agreement on using that data that is given to them by a data company, and they agree to that use limitation, they can't use the data, say, for instance, for Fair Credit Reporting Act purposes -- to get someone a job or get someone insurance, they can't use the data, and they agree to specific uses of the data before they get access to it, so, I know we're going to get into that a little bit later, on what is in the use agreements, but that is a key piece, to ensure that the data is used for the permissible purpose, or for the purpose from which they are asking for it.

MS. MASTALLI: If I may give one, I'm going to oversimplify greatly, but I think it's important to understand that the concept of using commercial data that is collected for other purposes to facilitate government implementation, and to force another warrant, I probably… Everyone in this room has, at one point or another, if they've ever traveled out of the country -- filled out a form that said I-94 or a similar form saying, "Okay, I am

going to be going to the Hotel Moonbean and staying for two weeks, and the purpose of my trip is tourism." In portions of immigration laws, which is an area that we learned about post-9/11 -- one of the things that we learned is -- not everyone tells the truth when they fill out an I-94 when they're coming into the United States. In fact, none of the hijackers did. Had we understood, we would have had far fewer absconders early on. If you can look at commercial data that says, "In Manhattan or Washington, or Dubuque, there is no hotel at the address, or by the name," it's not only at the address provided, but anywhere in the city. You know the person who filled out that form either was gravely mistaken, or perhaps, wasn't intending to comply with the law. The information about the location, or even to get into a lower level, whether or not there was a reservation made, or reservation available on that date is an oversimplified example of how commercial data collected for commercial purposes -- if there's, you can verify -- I can look at a map and say, "I have 10,000 I-94s claiming 126 Westbarn Avenue in Bethesda, and there is no such address." Then there's probable cause to look at these people as potential absconders. It wasn't collected for the purpose -- I never thought I ever needed to know where empty lots were, versus where hotels were, but it can make it as simple as that available from -- and timely, current and accurate make the law enforcement activity of ICE easier to conduct.

MS. SLOMOVIC: Welcome, Mr. Swecker, thank you for coming. We're going to take five minutes, and ask you to please tell us a little bit about how your agency uses commercial data -- what it buys, how it uses it. Please proceed.

MR. SWECKER: Good morning everyone, can you hear me? My name is Chris Swecker, I'm the Assistant Director of the FBI's Criminal Investigative Division, which gives me the responsibility for all things criminal, both in our domestic investigations and any investigations that are criminal in nature overseas. I have about 24 years in the Bureau, and I was a state prosecutor before I came in the FBI.

I testified on the Hill about commercial databases, and I have looked at it from the standpoint of the consumer of commercial database information, and we as an agency subscribe to several commercial databases, including ChoicePoint, Lexis-Nexis, Dunn & Bradstreet, and a variety of others. Our mission, of course, is not just in the criminal area, but in the counter-terrorism area, is to prevent terrorist acts, first and foremost. So, we try to support the counter-terrorism division as much as possible, leveraging our criminal resources against the terrorist threat. By that I mean, looking for fraud activities that may be -- the proceeds of which may be going to the prohibited charitable organizations. You probably heard about the armed robbers in California that were involved in some type of -- these are home-grown Jihadists who are engaged in some type of plans to attack facilities around the United States. So we look very closely within the criminal division for any leads or any nexus to terrorist activities.

But then, moving down the line, we're involved in corporate fraud, gangs, drugs, violent crimes of all varieties. As an investigator, you know, if you look back 23 years ago, if I wanted to gather information about a subject, if I had been drawn into any type of investigation, one of your first steps is to learn as much as you can about your target, or your subject. We would have to physically go to the courthouse to get real estate records, we would have to be sending these to another state to go get a driver's license record or a picture, we would have to go to a lot of different places, and manually gather this information -- if you were lucky, maybe the courthouse would have their records on microfiche, maybe not. I remember in the early days, actually going through the Register of Deeds, those big, giant books that you pull off the shelves. So, I looked at commercial databases as a way to efficiently gather information that's available to any other public citizen. If we access information that is not available to the general public, in my experience it has been that the general public -- especially businesses, for example, often can get more information than we can, legally. But if we access information that is not available to the general public, we will either have to get a grand jury subpoena, administrative subpoena, if it's a drug case or a kiddie porn case, or one of those cases where we're permitted to do administrative subpoenas, or whatever type of process is necessary to get the information.

We're mindful of the Right to Financial Privacy Act, and other limitations, so as we access this information, we have to get -- we may or may not have to get legal process to access it. For example, we access credit information in a routine inquiry, say through ChoicePoint or Lexis-Nexis simply to get name and address. If we go beyond that, we have to get some type of process with the information. So, I look at it as a way to save investigative time and resources. We have about 1200 fewer agents in the criminal division than we had prior to 9/11 because of the need to move agents over into the counter-terrorism area. That means we have to be efficient in what we do with fewer agents. We've had to prioritize what we do, so accessing commercial databases saves countless, countless hours of investigative time. And there are, I could go through a long litany of examples of how it has helped in various investigations, but I'll just stop there, and maybe just take questions.

MS. SLOMOVIC: This is actually a good segue way into contractual arrangements and use limitations. I would ask Jennifer to start, please, to talk a little bit about that.

MS. BARRETT: Obviously, whenever data is provided to the government, it has to be under contract. I'll make a general comment about the government procurement process, which is quite different -- at least in our experience -- from the private sector, and that's that, to a large degree, the terms and conditions, which is where the use limitations would typically be placed, are part of the GSA contract, and become, I think, in the minds

of some agencies, actually the people using the data, part of the boilerplate that we all know exists, but that we rarely look at. We kind of all know the sorts of things that are in the boilerplate -- you know, all sorts of liability and other kinds of issues. But, I think it's extremely important that -- and we've actually gone to the step of providing the terms and conditions that go with the data contract to the agency, making sure that the people who are going to be using that agency tool understand those terms and conditions. And quite frankly, we're sometimes a little surprised at the fact that they've never seen it before, which doesn't necessarily make it comfortable for us to be data provider for that particular agency, particularly if it's data that they've not used before. And I think with any agencies we're looking at expanding their data uses over a lot of situations where there's resources that they have never looked at and they may think it looks just like this other kind of data that I've got, but the rules around each use are going to be very different.

As I mentioned, there are three kinds of rules that we have to build into our terms and conditions. One are legal restrictions around the data, and we're going to talk about that, I think, at the next panel; the second are industry guidelines or practices that industries have adopted around the use of certain kinds of information; and then the third is our own policies and procedures around that information, and to some degree you can say there's the fourth, although we consider it as part of our policies and procedures, and that is that much of the data that any information company brings to market -- they're not the original source on it, and so, there may be restrictions placed on that data by the source. In the case of a public record, it might be a legal restriction. In the case of a public record, it might be a legal restriction, in the case of data from a private sector, it may not be a legal restriction, but it's a contractual restriction that applies to the recipient of that data, that has to be passed on downstream.

So, we're very sensitive to all these types of regulations, and all these types of uses. And it gets very, very detailed. For instance, the use of a date of birth as part of information from a credit file, or not having that date of birth in that credit file is a very distinct differentiation, because now I have fallen under the guise of the Fair Credit Reporting Act, when I include a date of birth with the name. If I just give the name, I don't. So, it's very, very specific, and it does require a fairly significant amount of communication during both the sell process, as well as the start-up process.

MS. DIBATTISTE: I would only add to what Jennifer has said, and say first that the vehicles that we're familiar with that the government uses, or the traditional ones that I'm sure all of you are aware of, and that is, you know, the full and open competition, GSA schedule, the sole source purchases, or other schedules such as FedLink, which is through Library of Congress, and elsewhere, and I think also some government-wide acquisition vehicles sometimes are used by the government for procurement, and that would be the

same in working with a commercial data provider. And then once -- let's say that it's an RFP and, there would be a service agreement, and as Jennifer said, agreed to between, in this instance, the government and the data provider, and I have read it carefully, it's a five page or six page agreement, and I would say that there's a lot in there, a lot of legal liability terminology, the purposes for which that is going to be used, but I would put the terms of the agreement into some of the following key categories that I think would be of most interest at this forum, And that is credentialing -- the government has to go through a credentialing process, and that's in the service agreement, just like every other customer has to go through one, and that's one that, at least from a ChoicePoint perspective, that we've strengthened and improved upon over the last several months.

There's also many terms and conditions regarding use limitations that Jennifer talked about, I spoke about earlier, saying that whoever it is we're entering into this agreement with is the end user, that they're not going to re-sell the information, that they have to comply with Graham-Leach-Bliley Privacy Protection Act, any motor vehicle state requirements if they're going to use motor vehicle information, and cannot use any of the information for Fair Credit Reporting Act purposes -- so that's all laid out pretty specifically in these agreements. The third category, I would say, there are terms that say, protecting against misuse of the information, and that applies both to the giver of the information and the receiver of the information, and also what to do in the case of a breach, that's also contained in the subscriber or user agreements, and then there's portions on account maintenance -- how are you going to ensure as the holder, or the user of this data -- that you are maintaining it properly, and also there are terms and provisions that talk about privacy principles, ChoicePoint's privacy principles, and that we consider important, and there are terms and conditions in the agreements that talk about agreement to follow those privacy principles, again, two ways.

And then, finally, there are audit provisions, where the government is agreeing that we can audit. Now, obviously we don't see their searches, they are blinded, we don't see what they're conducting the search on, obviously, they're sensitive investigations. But what we can go back and audit, when they're using one of our online products, every time that online product comes up on the screen, they have to certify -- the user has to certify that they're using it for a permissible purpose. And we can audit that piece of it, and they also agree to random audits and other types of, you know, we get audited by motor vehicle, states doing audits of us, are we using and maintaining the information properly, well, the same thing is in, are on these types of agreements, saying, "We're going to come in and audit randomly, and sometimes not randomly, your use of this data, and are you using it for permissible purposes."

MS. MASTALLI: I think that the answer to the question has been given very well by the commercial provider representatives, in that commercial data is accessed and its

use is limited by every contractual arrangement you could possibly imagine. Fixed priced, volume of access, short-term through RFPs, through direct source off of a schedule, but it occurs to me that perhaps because of the, some of the to-be vision issues, what is little understood about most government-accessed commercial data right now is that it continues to be, with a few exceptions -- one person accessing one record at a time, and having a specific reason for doing so. The audit trails are different, and one of the things that my office is trying to do is to bring some greater standardization to, you know, how DHS efforts, and it components, and sub-components acquire, what limitations they agree to and assure that the standards of the Privacy Office, and the Civil Rights and Civil office are embedded in all of these agreements on a standard basis, and the implementation is common across the entire enterprise. And that we have not just the external agency provider audits, but also our internal audit process in place. And the fact that there is a huge number of acquisition mechanisms out there complicates this to no end. And the fact that two people sitting in the HCOM, for example, may have accessed through their home agencies the same data under different use agreements is something we need to work on. Anyways, possible -- communicating training to one level of use agreements across the enterprise, across the federal government is part of future panelists and they will talk about that.

MR. ROSS: Grace, I just want to add one other thing to this -- Chris alluded to it initially. Particularly in the law enforcement area, when you're dealing with a cadre of people, investigators -- criminal investigators, in particular -- but even outside of the criminal area, who are used to data use restrictions there are no more stringent protections for information than, for instance, 6C -- information obtained for the grand jury process. All criminal investigators are very frugal with that, there are criminal prosecutions that can come out of misuse or abuse or discipline for that. In the IRS context, I can tell you that the 6103, which is tax return information and provisions in Title 26, are among the most stringent that exist -- the ability to obtain tax return-type data, and that's a pretty broad range to determine the term.

The restrictions on disclosure are so difficult that, in a lot of cases it's virtually impossible to make the disclosure under the law, no matter how important a particular agent might think disclosure could be in a particular context -- the Privacy Act, Bank Secrecy Act have their own protections, the Trade Secrets Act is not used much, but it exists, so the individuals who are accessing -- whether they're accessing for the FBI, or the DEA or the Treasury or the IRS, are extremely well-versed in protecting data.

MS. MASTALLI: If I may, one of the things, we've in fact in this very past week run into use limitations and concerns that have presented some difficulties. An example related to Katrina is the commercial providers are so conscious of their responsibilities, that when, for instance, prescription information requested to try to address restoration of

benefits -- in some cases, immediately save lives in an emergency situation -- have resulted in, shall we say, a small snow storm of paper and requests for waivers which -- under emergency provisions aren't really necessary -- and documentation and indemnification to ensure that even though this is, in the case of prescriptions, personally identifiable data desperately needed to make sure that individuals are receiving the immediate care, medication and treatment they need, and that is available readily by commercial providers,

I've got to tell you -- I'm not necessarily happy with the use limitations and requirements that -- and there is a small team at HHS, DEA and the Red Cross that's been working on use limitations issues related to melding commercial and private sector data right now in an emergency, because in fact, we have set the bar so high. Or it is understand to be set, perhaps even higher in fact than it is.

MS. SLOMOVIC: We've heard about training, people training, understanding how to use data, we've heard about legal limitations, we've heard contractual limitations, do we know anything about technical limitations? Are there technical limitations on access to data? Other than somebody having to certify on the screen that, yes, I'm using it for a permissible purposes. Are there technical implementations in place that link the query and the purpose of the query to the kind of data that is being requested?

MS. MASTALLI: At DHS and in my experience with a number of the inter-agency activities, we've been moving toward role-based access. I think what most of us are familiar with right now are technological barriers to the use and access of data, which are unintended problems, in terms of how the data is structured and how it can be made available and through what mechanisms. We have talked about the to-be vision to a far greater extent than can be realized on the government technology at this point in time in terms of the ability to access and know. It's one of the biggest problems post-9/11 with data which was publicly available, and did not require any particular protection, which was working out mechanisms with the data structure. We did not have common data records models, and we've just begun to move forward to resolve the technological barriers to be able to access and use information which, for totally legitimate purposes. The FBI might want to address those other issues.

MR. SWECKER: Well, first, I think it's real important to understand that in my portfolio, a criminal investigation, we cannot, legally, just data mine information. We have three levels of inquiry -- actually prior to 9/11 we only had two -- one is a preliminary inquiry, one is a full investigation. And these are governed by Attorney General Guidelines that we follow. A full investigation is what the name implies, you can use all investigative techniques, including database checks and that sort of thing. The standard there that you have to have is that there's a reasonable indication that criminal activity is being conducted by an individual or a group, or an organization. That standard

is below probable cause, but it's higher than just a mere suspicion. So there has to be a minimum threshold where we can open a full investigation, and those investigations are approved at the supervisory level -- usually an Assistant U.S. Attorney is consulted prior to opening an investigation -- we done want to open an investigation that's not involved in a criminal violation that would be prosecutable.

The second level is called a preliminary inquiry, they are 90 days in length, it's a much lower standard, it's not, obviously not probable cause, it's not reasonable suspicion, it's a -- someone has provided information, there's been some tip, some lead that has come in, but it hasn't been corroborated by a second source of information, or a reliable source of information, and so the preliminary inquiry is opened up for a 90 day time period to do certain limited investigative techniques to determine if, in fact, there is criminal activity afoot, and we can not use the techniques of undercover operations, we can't use wiretaps, we can't record, you can't do surveillance -- a preliminary inquiry is pretty much limited to checking our internal indices, checking other agencies' databases, checking public databases, commercial databases, and possibly an overt interview.

A new threshold that's been developed since 9/11, and that is called threat assessment. That threat assessment is somebody that is called in with a very general tip or lead. Post-9/11 we were getting thousands of phone calls, very, very general in nature, very generic, but woe onto the agent that didn't follow up on a terrorist lead -- no matter how general, no matter how unspecific it was, and believe me, as an SAC or an agent in charge of the North Carolina office for five years, right after 9/11, I got phone calls at home from soccer moms and friends and associates, "I saw three Middle Eastern people meeting at Starbucks," I got that phone call at home. But we had the dilemma of, "What do you ignore, what do you follow up on?" So, they developed a third level, which was a threat assessment, that was simply, check your internal databases, and check the commercial databases and see if this person has, in the past, has had any nexus to terrorism. So, that is only in terrorism investigations, it is a third tier, it's very limited, it's a very limited public database check.

So, with that backdrop, I think it's important to understand, we don't just simply dive and data mine our information. The closest thing to that would be a batch query, for example, you run across a drug pusher, you've executed a search warrant, you've got his address book -- is it a list of customers? Is it a list of clients? Is it a list of money-laundering contacts, facilitators -- you've got an address book with 200 names in it. We might do a batch query -- in other words, we might check all of the names in that address book to see if any of them have been linked to prior drug investigations, for example. But I think it's important to distinguish between that and data mining, because there -- this is part of a predicated investigation -- you've got a drug trafficker, you've got a gang-banger, you've got some type of criminal activity and an identified subject, and you have

names that you have to do something with. To ignore that address book would be tantamount to investigative negligence. Even more so if it were a terrorist investigation, and you found this address book in a cave in Afghanistan of a known Taliban or Jihadist. So, I think it's real important for everybody to understand where the thresholds are to trigger any type of inquiry. We just do not have the legal authority to simply dive in and do database, or do data mining. And I think a lot of entities and a lot of people don't make that distinction between data mining on a fishing expedition, and checking information in connection with a predicated investigation.

MS. BARRETT: Let me maybe give you a way to think about the question, because it has lots of dynamics, and I think when you talk to an individual user or an agency, they tend to frame the question in terms of their particular way of doing business. There are three dynamics to the issue of the technology of the access. And the first is a little bit of what Chris is talking about -- real-time query versus batched queries, which come back at some point later, it may be a few hours, it may be a day, but after a period of time I get an answer back. Or what we typically would think of, at least in our vernacular, is a data mining, which is doing more analysis on the data, we're not doing queries for individual data, but I'm doing analysis on what's going on today, that's one dynamic.

The second dynamic is -- are you getting a confirmation or a verification of the data that you have, or are you actually retrieving additional data? Because both instances occur, in both a real-time and the batch queries, and that creates a different dynamic. If all I'm getting is yes, it's the right data, or no, the social security number doesn't match, that's one kind of situation. If you're actually pulling down additional data that you're doing to integrate into some process, then there are other system issues that have to be deal.

And the third situation is one where we see less of the latter, is a person going to be receiving this information and doing something with it, or is the system going to be receiving this information and doing something with it? And, you know, I think we have moved quicker when the people deal with data then when we have systems deal with data. That's only from our experience, but they, Grace can comment on it, but putting someone online to be able to get to data whether it's real-time or batch, is easier and faster than it is writing a system interface that's going to accept data, if we do, in some automated fashion, whatever process comes next. So, keep in mind that we're talking about the technology accessed data, all three of those factors -- how do you make the query, is it just to verify or is it to retrieve data.

MS. DIBATTISTE: The only thing I would add to that is on the online, the way the government accesses data, from ChoicePoint, at least presently, is one of three ways, and I just want to stress from a technology standpoint, the web-based interface, which is online, is secure, so it is a secure, web-based interface and as I mentioned, a person is coming in, and every time they log on, they are asked if they are a legitimate user, ad if they're going

to use it for permissible purposes, and that is audited, and also there's policies in place through the user agreement that I talked about earlier, that also require the changing of the passwords every certain amount of days, the deactivation of the account if the user hasn't used it for a certain amount of days, and basically the DOJ standards on user IDs and passwords are incorporated into each user agreement.

The second way they access is through the batch that Jennifer just mentioned, and I'll give you an example from a homeland security example on batch, and I know Chris also gave one from a Bureau prospective -- if a batch inquiry is sending a file to us, that would be encrypted, so again, it comes over secure, and that file they would -- let's give an example, they want to know where all these fugitives are, and that's a batch inquiry -- giving the name of all these people, we want to know what their current address is so they can maybe go, we can maybe go locate absconders or fugitives, and prioritize the order in which they want to do it -- that would be a batch. That technology is up and running and secure, and encrypted.

And then finally the system to system, that is done in cases where -- and that allows the end user -- let's say it's a law enforcement agency, direct access to the ChoicePoint system, but that is, again, that user is, the agency goes through a pretty strenuous credentialing process, service initiation process for subscriber agreement, and that system, the system again is encrypted and secure. So, that is only where the technology permits it on both ends, and that is one that I would say, where work needs to be done. Because that is obviously the most efficient, the most secure, and we need to do more work in that area to have more users using the system to system than going through the web-based and, well the batch is pretty secure in and of itself also.

MS. MASTALLI: It's notable that, and I think that some of the people in the panels later today will address, but in many instances, if we had more system interaction, we'd have higher standards of privacy protection. It is the human engaged in the middle that often proves the most difficult to monitor and audit in these processes. There are several examples of very intense negotiations that went on immediately post-9/11 to try to develop systems and some arrangements, and there are arrangements, you can look at GAO, of course, that outline, but from a technology perspective, and from a process prospective, it takes major negotiation to get information about data structure to do programming on the government's end to build in the protections and the utility necessary to interface, and we've made strides since 2001, but the process of the future vision of how it could work, and how it could enhance privacy if it worked better, it's still a vision.

MS. SLOMOVIC: Thank you everyone. I would now like to open the panel for questions, please. Please come to the microphone that is located in the aisle to my left, and before you ask your question, please identify yourself for the transcript.

MR. VON BREICHENRUCHARDT: Good morning, my name is Dane von Breichenruchardt, I'm with the United States Bill of Rights Foundation, and I'd really like to thank all of you for coming. This is very helpful, and this meeting is very useful.

My question, or questions, I suppose, go to, of course, privacy issues, and it occurs to me that, when the government gathers information on citizens and holds it in a database, it has to abide by the laws and the Privacy Act. And even when the information is being gathered, particularly if it's being gathered directly from the individual, they have certain questions they have to answer -- you know, what are you going to use with it, what's the authority for this information, who else are you going to share it with, that sort of thing. And I'm concerned, that it seems that, I guess just because of the growth of technology and I think things have kind of gotten out ahead of us, is that when the government goes to private brokers, information brokers, and even though I understand you to say -- and I believe I understood you to say that it's not that you're creating a new product for them, you're not going out and data mining something that they've asked you to go do, but you're compiling information that's already available. So, you're concentrating a lot of information on individuals for your client, the government, and of course this is a big -- you said yourselves -- that this is a big, growing area.

And my concern also, aside from the Privacy Act, is that there have been a lot of breaches -- ChoicePoint and Lexis-Nexis and other brokers have had severe breaches of security. And so you have, when you hire someone who works for a government agency, the kind of clearances they have to go through to work for the United States government and to deal with private databases, or databases on private individuals, versus those who are hired by ChoicePoint. What concerns me is, are we getting the same level of protection as to who you're going to hire versus who the government will hire.

So, my overall question is it seems to me that private databases, we do not enjoy the protections of the Privacy Act, and that even though you do have internal restrictions and agreements and contracts, and you audit yourselves and you do all of these things, the fact is that it sounds like very little of this is under the law, particularly the law of the Privacy Act itself. So, my bottom line question is, do you think this is now an appropriate time that Congress take a look at this industry, consolidating information on people, particularly, you talk about the date of birth requirement. But what's interested is that we have this new program coming up which is really CAPPS III, but they call it something else, where they're going to ask for your name and date of birth, and of course they're going to check that against a no-fly list, but then they're going to give it to a data broker, and they're going to data mine it and see if they can come up with matches, and you get sort of like an APGAR score as to whether or not this person needs to go to another level of clearance. And the problem is that, as I see this, is that even though you're not data

mining specifically for a new product, you're consolidating a lot of information on individuals.

MS. SLOMOVIC: Thank you, sir, let's let our panelists answer the question, as I understand it, the question is whether or not data collected and consolidated by private sector companies enjoys the same or appropriate level of protection as the data collected by the government, and whether there is something that the government -- and whether there is something that the Congress ought to be looking at given the changes in technology and the consolidation of data in the private sector.

MS. DIBATTISTE: Thank you, yes, sir. First, the Privacy Act itself doesn't directly apply, as you so accurately stated, to the commercial data providers. But, as I tried to portray here today in my comments and as you've also mentioned -- there are many policies, practices, procedures in place that the private sector -- not just the data providers, but the private sector and our customers, also, have to comply with. There's the Driver Privacy Protection Act, Graham-Leach-Bliley, the Fair Credit Reporting Act, all the internal privacy principles that we have, I know every company, every major data provider has, and obviously the government has privacy principles that they comply with.

We are developing new policies to ensure protection of the data, so although -- your second part of the question is, should Congress be taking a look at this? They are taking a look at it, and we at ChoicePoint welcome their taking a look at it -- we want to be part of the solution, and we think it needs to be carefully thought out and all of the principles and practices and regulations that are already in place that we want our customers to comply with to protect privacy, civil liberties and civil rights -- we want to only enhance that. So, although the Privacy Act specifically does not apply, I know our company has laid out privacy principles, we've initiated policies, we've strengthened our credentialing procedures, we've complied with all of the laws, we've put them in our user agreements, we've done everything to ensure that the principles of the Privacy Act are engrained in our institution, and in how we operate. And yes, we welcome the Congress into a thoughtful discussion of how this could best be done. And I don't think it would be helpful to any of us if it was done on a piecemeal approach.

AUDIENCE: The breach of information?

MS. DIBATTISTE: Yes, and there have been, I think we've counted about 104 what you would call breaches in 2005, and a lot in our company, of course, education industry, other data -- not just data providers. I guess I want to make that a clear statement that it hasn't just been breach of data providers. I think if you count the 104 incidents, you're going to find most of them with institutions of higher learning. We have been keeping track of that. But that's not to say that it wasn't important, that it didn't happen to us, and I would say that ChoicePoint has done everything humanly possible to be out in front on

this, to ensure -- and they were already doing it, I'm on board for four months to ensure putting extra procedures and practices in place, as I mentioned, strengthening our credentialing practices and procedures, and that was a piece of our breach, as you know, and ensuring -- there's never a 100 percent solution, in anything, in terrorism, in transportation security -- there's not a single silver bullet. Will we ever be able to ensure there will never be another breach? That's unrealistic, and if I told you yes, I would not be credible. But we are doing everything possible to ensure that our customers are credentialed properly, to tighten up the credentialing procedures, to do site visits on customers, to ensure that all of the laws are complied with, to enhance our privacy policies -- we have a consumer breach notification policy now, we have a new user access policy to enhance our user IDs, our password protection -- everything humanly possible to be done is being done and will continue to be done to ensure that there is not another breach. But, can I tell you 100 percent there will not be? No.

MS. BARRETT: I'll echo a lot of what Carol said, I'm not going to repeat it, but I want to add a couple of points. I've testified three times -- we are also involved in the discussions on the Hill -- I've testified three times in the six months about this issue, and we are on record as both endorsing a national breach notification law, as well as a safeguards requirement to the private sector, including data brokers, but not necessarily limited to data brokers, that we think would provide a very high standard of a permanent requirement to safeguard the data, to credential your clients, to make sure that it's properly protected, whether it's an employee or whether it's being delivered to a client. And unfortunately, and I guess this comment is directed to your Congressional staff, it is probably unlikely that we're going to get something passed this year with recent events and other issues and the session coming to a close, and quite frankly we find that very disappointing.

MS. SLOMOVIC: Next question, please?

MS. GRANT: Good morning, I'm Susan Grant from the National Consumers League, thanks to the Department for holding this public workshop. It's pretty easy to understand the use of commercial data in investigations if you are looking at a particular suspect or if you're having to do something like identify all of those scuba shops. But I think it's the proactive use of data that gives rise to more concern. In the example that Grace gave, people flying first class who appear to have lower incomes than would justify that -- in order to find a pattern like that and identify those people, it requires looking at the information about all airline passengers and their incomes. And then while that may be able to be done in the aggregate initially, once you have identified a suspected pattern of something, it then requires identifying those particular people, to look into them further, and I think that probably none of them would be guilty of anything more than misstating their incomes over their careers. At any rate, what other kinds of information

about people is the Department of Homeland Security looking at to try to identify patterns and trends, and if you can't reveal that, is it fair to assume that there's all kinds of information about us that you're looking at that we'll never know about?

MS. MASTALLI: Actually, that example was for Treasury, not DHS's regarding financial enforcement, but I will -- (Laughter.)

MS. MASTALLI: I think that there is a misplaced fear regarding the focus of what is done proactively, and I'll be the devil's advocate to suggest that we would be more secure if, in many cases, more of the kind of work that is posed that the intelligence agencies -- the FBI, and DHS do, as some of my colleagues from the other agencies said -- most of the work that is done is done on a very, very highly aggregated level to try to identify -- to apply lessons learned from the past as I think everyone said earlier, and to look at patterns at a very high level. I will not speak for the other intelligence agencies outside of DHS, but there is, each national security agency has different specific responsibilities in how they do their work. From my devil's advocates position, too little of it looks at U.S. persons information overall to try to identify what might be occurring in the home-grown terrorist front. A vast majority of it is focused at non-immigrant alien information, and is grounded initially in government data collected about non-citizens.

There is important information that is used commercially for other purposes, such as preparedness, a number of issues regarding the infrastructure, but other than the extent to which the limited -- and they really are limited data mining data exploitation efforts -- I think that we need to focus on the actual work being done and what needs to be done within the existing and planned activity and address our audience and our legislative activities to the "to be" future with the recognition that in the cases that we -- and in the real risk-based analysis of what we decide to prevent from being done and what we will allow to be done -- the Privacy Act has far greater flexibility than I think the Federal government has availed itself of it in terms of examination of existing records, but I confess I'm less concerned about the scenario you just described, because it's really not something that is on the drawing boards, for the most part, in any of the agencies that I've worked with. Does that answer your question?

MS. GRANT: Thank you.

MS. BARRETT: Susan, maybe if I could add one comment to this from the private sector side, and these are some lessons learned from private sector use of vast amounts of information, where we don't have exactly the same -- maybe the same concerns, but we do have significant concerns, and we do have laws that protect it.

Quite often the analysis of what data is indicative of something, in fact in almost all cases, for practical reasons as well as privacy issues is known in an anonymized fashion. To determine what is the model, what are the rules, or what are the set of circumstances

that I need to be looking for. When the circumstances are validated as, "This is something that I need to watch for," regardless of what, in these examples, but how you actually implement that, you have vast choices to do. And that's why, when you talk about the technology, I brought up the concept of a verification -- as opposed to a retrieval -- of data. In the example you gave, and like I say, I hate to use it, but just for illustrative purposes -- if it was determined that it was highly significant that a person with low income is flying in first class has a high likelihood of being a terrorist, you don't have to give the income data to DHS to execute on that kind of decision. It can be a model that DHS makes a query against the private sector repository that has the financial data you apply the model to, and you get a yes or a no back. So, we have ways, I think, of taking advantage of the sophistication of the data and what it can do to help us and minimize the invasiveness or the distribution of personal information beyond the fact that it's already residing and already being used, and I think we need to keep that in mind when we start talking about the more sophisticated uses.

MS. MASTALLI: I do want to adhere to that, in fact, the commercial sector has done a better job of anonymization than we have yet been able to do in the federal sector. As one of the examples of how evolving technology will improve what we do, both within the government as the adoption of private sector practices.

MS. SLOMOVIC: Thank you, next question, please?

MS. STIRLAND: Hi, Sarah Stirland, from National General Technology Data. Grace, I was just wondering if you could walk us through the scuba diving example, I wasn't quite clear that I understood you properly, are you saying that those are so special that terrorists were taking scuba diving lessons, so that instead of sending FBI agents to go and participate in these classes and find out what was going on, you just buy data and find out through --

MS. MASTALLI: And Chris may well be familiar with this. This example was a post-9/11 example of, there was information suggesting that non-citizens might be seeking training, and the goal was simply to identify where all of the training schools that provide scuba training and certification around the country. And the FBI sent out to all field offices, I don't remember where you were then Chris, so that the special agents in charge in all of the field offices directed their teams to find out where in their districts of responsibility there were scuba training schools.

MS. STIRLAND: Thanks, that was my only question for you. And my second question was to Carol -- you said that you've tried to ingrain Privacy Act principles into rules (INAUDIBLE) allowed to ask the government what information they have on them, and what it's being used for, and if the government changes the purpose, you're allowed to ask again what they're using the information for and correct it. Is that true for your

products as well, I mean, can I or anyone else in this room write to DHS or something, and ask what information ChoicePoint has on me?

MS. DIBATTISTE: It depends, I don't know if you can write to DHS and ask what -- if they're using it for a law enforcement purpose, I don't think they're going to be able to tell you, but you can… When we, if it's a consumer- based transaction, when that information is collected from the consumer, they know what it's being collected for, let's say it's to get insurance, or to open a bank account, they know that they are providing information that is going to then be used by that insurance company or let's say, by that financial institution, they will then come to ChoicePoint and say, "Are they who they say they are, and are they credit worthy?" And they, yes, they are advised of the information that they are going to -- that the insurance company or that the financial institution is going to be collecting, is going to be asking about them. The customer would be advising them of that. And then there is opportunity for a consumer to go onto a website called Choicetrust.com where you can, a consumer can access public record information that is available about them, and where a consumer can also, in certain circumstances, access other information that is available about them.

MS. STIRLAND: Okay, my last question is for all of the panelists who want to answer it. Given that you use all of this data and sometimes you don't actually know that you want the data that's very useful, is it practical to have these privacy impact assessments that you're all supposed to be filling out? As I understand it, these privacy impact assessments, you're supposed to tell people, I mean, through the federal register, what you're doing with all of this data, and in these fast-moving circumstances, can everyone be issuing privacy impact assessments constantly, all of the time?

MS. SLOMOVIC: I'm sorry, to whom was the question directed?

MS. STIRLAND: Any of you, all of you who have to deal with it.

MS. SLOMOVIC: Well, the government folks have to do a privacy impact assessment, and the private sector folks don't, so --

MS. STIRLAND: Well, the government people who are using commercial data.

MS. SLOMOVIC: I guess, are you asking whether the government does privacy impact assessments when they use commercial data?

MS. STIRLAND: Well, I'm asking are they doing them, and is it practical to have to do these things, given that you've just told us of all of these circumstances in which you have to use the data and you don't know whether it's going to be useful or not.

MS. MASTALLI: I'll take it. The short answer is the government is doing everything it can to fully comply, including issuance of privacy impact statements. The answer to whether it's practical is, it's one of my great frustrations, and I'll give a short

anecdote, and again, it's now a publicly known person. Immediately post-9/11, the Department of Justice, the FBI among other agencies, was extremely interested in FAA data regarding non-citizen and citizen pilots. And the FAA sent over that data to the FBI for use, and I had the misfortune of being the first person to open up the data and look at it and read the System of Records Notice language, and have the FAA's privacy impact analysis and System of Records Notice in front of me, and I told the very unhappy senior FBI, in fact it was an Assistant Director, that he had to send it back to the FAA because we could not use it. The Systems Records Notice had allowed the information to be used for all drug enforcement purposes, the purposes that it had been requested for were counter-terrorism, and it was not legal at that point in time, without the FAA issuing a new Systems of Records Notice and a new privacy impact statement for us to use the data on alien pilot certification that included some U.S. persons data for counter-terrorism purposes, so I, from practical experience think that there are impracticalities associated with full compliance, and I do not have the policy solution for that.

MS. STIRLAND: Okay, thank you very much.

MS. SLOMOVIC: I'm sorry, I know people have been waiting, but this panel is already over time limit, and we need to seat our next panel. I apologize again for the people who have been waiting to ask their questions. Thank you very much to all of the panelists for coming today and providing so much useful and interesting information. Thank you. [Applause]