



Homeland Security

Department of Homeland Security
Data Privacy and Integrity Advisory Committee

OFFICIAL MEETING TRANSCRIPT

Wednesday, June 15, 2005
Harvard Law School
1563 Massachusetts Avenue
Cambridge, MA 02138

AFTERNOON SESSION

MR. ROSENZWEIG: We're going to start. We have a quorum I think, of members and they'll joining us in just a few minutes. I'm going to take over the job that I asked Ms. Sotto to do of introducing people, and just so that I can make kind of clear to you, we proven to have a lot more questions than we actually have time for, so I would ask you all to kind of keep your remarks about five to seven minutes before I start coughing loudly, and then that will give us a chance to ask us some questions.

This panel is a panel of experts who work technologically in the field. We're very pleased to have, from my left, Norm Willox, who is with LexisNexis and serves as their chief officer for privacy. He oversees their domestic and a privacy regulatory activities. He's on the Board of International Fraud Symposium, and is a founding member of the Internet Fraud Council.

Next to him is Daniel Weitzner, I hoped I pronounced that right, thumbs up, with the Worldwide Web Consortium or W3C. He's a leader of the technology and society domain there. He's also a principal research scientist here at MIT, where he teaches Internet policy use. He's the co-founder and deputy director of CDT. We'll actually be hearing from the CDT person on the next panel from a policy perspective as well. So a double hit for CDT today.

Next to Daniel is Dr. LaTanya Sweeney. She's an associate professor of computer science and technology also at Carnegie Mellon and director of the data privacy lab there. I should add that one of my first and formative experiences was listening to Dr. Sweeney talk at a CSIS symposium about three years ago where I learned with some of the uses of data technology issues.

The fourth person on our panel is Dr. Simon Garfinkel. He's an MIT professor here. He received his Ph.D. in computer science in AI. He is a researcher on computer security, and has received awards for his commentary on information technology. He was awarded the 2004 and 2005, Jessie H. Neil National Business Journalism Award for Best Regularly Featured Department or column. That sounds interesting. Sometimes you see these great things in the bios.

So to the four of you, thank you very much for coming. We look forward to hearing from you. And as I said, if you keep your remarks about five to seven minutes so if there's an opportunity for questions that will be great. We'll go in the same order, it's random. Mr. Willox.

MR. WILLOX: Can you hear me okay? Mr. Chairman, thank you. Good afternoon. Good afternoon members of committee as well. Thank you for ordering up this cooler weather since we just flew up from DC this morning and it's been somewhat unbearable. I don't know how you did that but we appreciate that.

My name is Norm Willox and it's my distinct honor and privilege to appear before you today. At LexisNexis I do serve as the chief privacy officer for LexisNexis. I also oversee industry and regulatory affairs and my office is located in Washington DC. I'd have to apologize, I was advised earlier that I thought or comments would be fifteen minutes, so I'm going to jump through my comments fairly quickly to comply with the Chairman's request of five to seven minutes.

MR. ROSENZWEIG: If you have written stuff, we do want to have it.

MR. WILLOX: Okay. I'll submit that afterwards. What I thought I'd do is the history of Lexis. I think some of you are fairly familiar with it, so I'm not going to get into statistics and history, so I will jump through that. But I do want to talk a little about the LexisNexis Risk Management Division, where a lot of our public record data is stored. I think that's an area that's probably of interest to you. I do want to talk a little bit about the important of privacy to LexisNexis, and certainly screening and identify authentication, which is an area of focus for us, and then lastly talk about data mining and what that means to LexisNexis.

The LexisNexis Risk Management Solutions group is the one area of decision-making that is common to all of those -- all of the sectors within LexisNexis as it relates to risk. LexisNexis has responded through the formation of this risk management group, which is devoted solely to working with our customers and appropriately assessing and managing the diverse types of risks that they encounter.

The LexisNexis Risks Management Group confronts risks in various forms, from government entitlement program, frauds and abuse to financial institutions safety and soundness, to critical infrastructure protection and to law enforcement. The fundamental

toll with which we use to aid in the management of risk is information, whether shared directly, processed through analytical tools or integrated and managed through exceptionally powerful systems. LexisNexis has been providing information solutions to the government for over 25 years. Although originally done to respond to specific federal, state, and local law enforcement needs, such as to help identify and locate and criminals and witnesses, this service has expanded to include information solutions that aid in assessing terrorism risk, such as in the Department of Homeland Security Screening programs, and in assisting government entitlement programs to help identify proper recipient and addresses for benefit payments.

Similarly, the solutions have become more intricate to include not just data but also technology and policy assistance. As global information sharing policies evolve and are more advanced and analytical applications are needed for productive decision. Policy-based trusted enclaves are necessary to meet proportional information sharing, requirements necessary to enhance privacy and security.

In order to meet our emerging customer and societal needs around information policy, LexisNexis has developed an array of technical information supplied chain management, and I'm just going to list them here. I'm not going to talk about them right now.

We have data fusion tools, we have textual analysis tools, international data tools and custom screening tools. The importance of privacy to LexisNexis. Some of the data that LexisNexis collects stores and distributes is personally identifiable. Of this data the most sensitive, from a privacy perspective, are Social Security numbers and drivers license numbers. Regulated under Gramm-Leach-Bliley Act, and the Drivers Privacy Protection Act, this data is further protected through application of the LexisNexis data privacy policy which is implemented through a policy review board and enforced through a regulatory compliance unit and overseen by the chief privacy officer.

LexisNexis data privacy policy, which can be found through a link available on each page of our web site, provides the following privacy protection, and I'm just going to list them. I'm not going to go into details with them. LexisNexis requires public records and non-public information from established reputable sources, and the government and private sectors. LexisNexis endeavors to accurately reproduce all of its information in its products. LexisNexis restricts the distribution of Social Security numbers and drivers license numbers, and we allow individuals to opt out of it's non-public data bases under certain circumstances.

We also provide public access to the information, and LexisNexis takes reasonable steps to maintain the security of its data facilities and systems. In providing data to the Federal government for Homeland Security practices, we are particularly mindful of the Department of Homeland Security Secretary, Michael Chertoff's, admonition concerning

the need to protect privacy while providing processes and systems to aid and protect against terrorism.

At a recent presentation at the Center for Statistic International Studies, Secretary Chertoff said that remote security does not require that we sacrifice liberty. Our goal has to be to maximize all the values that foster our way of life. Although no one disputes that information, including personally identifiable information is needed, if not required, to aid law enforcement combat terrorism and foster critical infrastructure protection.

Some of the ways that we have done this are in the following: We've imbedded privacy enhancements into our technologies. Some examples of this would be developing of scoring models that will limit the amount of personally identifiable information that is necessary to accomplish the specific Homeland Security purpose and the creation of trust enclave environment, to the greatest extent possible, the data that is needed to accomplish the Homeland Security objective remains in the possession of the data provider, while simultaneously ensuring that the government queries and results of those queries are seen only by the appropriate government personnel.

Number two, assisting government privacy officials ensures that privacy policies are strictly followed during the development and implementation of Homeland Security programs. In the government programs in which LexisNexis has participated that have involved the use of commercial data or industry data, we have looked, we have worked closely with the privacy officials at the Department of Homeland Security and Transportation Security Administration in providing assessments of all commercial data uses to ensure that these uses are completely in accordance with the departmental policies. We have provided skill privacy technicians that work closely with the government and contract of personnel to incorporate the policies into the commercial data use and monitor compliance with those policies.

And, lastly, providing government officials with approaches to potential government uses of commercial data that maximize the benefit from the use of the data while simultaneously maximizing privacy protection.

Since the last December when the Intelligence Reform and Terrorism Prevention Act of 2004 was enacted, LexisNexis has been asked to provide various agencies with suggested approaches from on the use of data for the information sharing environments, while meeting all the required attributes, including the protection of individual privacy and Civil Liberties. We have been actively engaged through policy outreach through meetings with various government personnel, seminars that we've conducted for government agencies, and most recently the convening of various information policy expert to provide their assessment of critical challenges to invite them in government personnel.

LexisNexis' involvement with identity authenticating and screening programs. Again, I'm going to try to move to quickly through this.

MR. ROSENZWEIG: I hate to do this to you. I'm sorry you got a bad message. If you could get to the gut.

MR. WILLOX: I really can't wrap it up in thirty seconds. But LexisNexis' involvement in identity authentication. Identity authentication has become an important component in the use screening, screening individuals or entities, and also as it relates to data mining and its absence in LexisNexis projects as it relates to the U.S.G.

To the best of my knowledge, there are no government programs involving LexisNexis data, in which any data mining occurs, recognizing the fact that data the definition of data mining is an important issue. So with that, I'll conclude. Thank you for inviting me and I apologize for the --

MR. ROSENZWEIG: I think the apology is as much ours for giving you bad range in the direction. Dr. Weitzner.

MR. WEITZNER: Thank you, Mr. Chairman. I appreciate being here.

MR. ROSENZWEIG: I've been asked to ask the speakers to speak a little more slowly and into the mic.

MR. WEITZNER: The stenographer has my power source in her hands, so I've promised to speak slowly. How is that? I apologize for hiding behind my laptop. I just can't read my handwriting anymore, so I hope you'll understand.

In the short time that I have, I want to touch on just a couple of quick points which really center around my sense of the increasingly transparent information environment that we are all living in, that's, obviously, particularly relevant for the DHS arena, but I think it's a feature of the larger information processing world that's important for your consideration.

Just by way of background, I am with the Worldwide Web Consortium. W3C is an international corporation that sets the technical standards for the web, so that includes technologies such as HTML, XML, a whole bunch of other alphabet soup that are inside the guts of the various web, pieces of web software that we all rely on.

And my comments reflect not particularly the views of the members of the consortium, DHS is actually a member of consortium, just for the record, but reflect my sense of the evolving technical landscape as we see it from W3C. I want to talk a little bit about some of the larger trends that are pushing us towards an increasingly transparent environment, and by transparent I mean an environment in which personal information is increasingly exposed.

I want to highlight one key public policy challenge that I think this raises and then concluded by suggesting that we really have to learn, I think, in many ways through the efforts of this body, how to expect a higher level of accountability from the information systems that we all depend on, particularly in this case. So let me highlight four technical trends very quickly.

The first is what people in the IT industry refer to as the end of stovepipes. Stovepipes are what IT managers worry about when they have one system over here that has payroll and another system over there that has accounts receivable and a third one that has the catalogue, and a fourth one that has something else, and they want to learn about information that's spread across those different systems, they can't because they come from different vendors or the data is in different format, et cetera. IT vendors, to their credit, are gradually figuring out how to collapse these stovepipes, to join these stovepipes. It's great news for enterprises, large and small. It has the affect also, as I think you're well aware, of removing a key impediment to privacy intrusion. This is not an impediment, of course, there was put there on purpose. It was there because of the limitations of systems. But as we see the collapse of stovepipes, we will clearly see a major privacy impact.

Secondly, query in general, the ability to search over a larger and larger volumes of information is becoming nearly free. We all have extraordinary career capabilities on our internet connected desktop or laptops are increasingly mobile. Mobile devices through Google. We'll have it from other places. I don't think any one of us could really imagine the web without Google, and, of course, the privacy impact there is clear. When that sort of query power is available, there are significant privacy implications.

Many information environments are becoming increasingly location aware. This is important not only because people are obviously sensitive of that issues of location privacy but also because location is a key data element that facilitates the linkage of many, many other types of information. I think we're only at the very beginning of understanding how to manage the privacy issues associated with location awareness. Yet, as I think we all see just see from the devices that hang off of our belts or sit in our pocketbooks, we're seeing a proliferation of services.

Finally, and in some ways most importantly, the cost of data storage is approaching zero, if it's not zero already. Google offers us a gigabit, and maybe it's now up to ten gigabits, of data storage for free through Gmail service. Other online services offer the same sort of capability. And the importance here is that the decision about data retention has fundamentally flipped. It used to be the data retention policies dealt with the question of which data to keep, and which data throw away. The default now is to keeping everything, because it's simply less expensive, for any institution, large or small, to just keep everything, rather than go through what is an expensive and difficult and

sometimes liability-incurring process of deciding which to keep, especially if you change your policies around the time indictments come along. So data storage will cost, essentially, nothing.

And I can report that new data base architectures are, in fact, engineered on the assumption that enterprises will just keep all their information, whether in one place or spread across a number of data storage. So, again, the reliance that practical privacy protection is placed on the gradual destruction or deletion of data just won't exist anymore.

I think there's one key public policy challenge that this brings up. And the challenge has to do with the fact that, of course, what is revealing, what is threatening, I think from a privacy standpoint, is not so much the individual bits of data that are stored, and there are more and more of them, what is challenging is that our power to construct inferences across this increasingly large set of data is growing at a quite extraordinary rate. And I think that it's fair to say that, at least in the United States, privacy protection framework, the ability to regulate inferencing power is really quite underdeveloped. We don't have legal tools for defining the scope of inferencing that's permissible, versus the scope of inferencing that's not permissible. And that's something that I think we'll have to really come to take very, very seriously.

Seeing that my time is short, I want to jump very quickly to what I think is an important, both, technology agenda and public policy agenda for your consideration.

I was very heartened to hear Ms. Waterman's discussion of the interest in encoding regulatory and legal frameworks into machine computable rules. I think this is a tremendously important effort. It syncs up, as you probably know, with a lot of the leading edge technical efforts in the industry. And I think it offers tremendous promise for helping to balance what is an increasingly transparent information environment, with some transparency that individuals and regulators can actually take advantage of in understanding how information is actually being processed.

With all respect to everyone who puts a lot of time into writing privacy notices, I think if you did a little accounting for yourselves of how often you or your dear friends or spouses or children or anyone else around you reads privacy notices and takes the time to understand them, that you would see that our path toward greater control over our privacy information may begin with -- certainly begins with clear privacy policies, but does not end by just delivering the text of those policies to the customers, whether they be the American citizens who are the customers of the service. So I think developing technologies in the area of more transparent reasoning, enforcement through rules and secure audit to be able to check compliance is really quite critical.

As a public policy matter, I think that the critical goals that we have to look at is to come to terms with the shift from current limits that we have in data protection rules on the collection information, to doing the better and better and more and more quantitative concrete job of articulating what the rules are for the use of information, not just what's collected. I think that we have to give individuals, who are the subjects in these databases, the right to accountability and the right to audits that establish that their personal information was handled correctly. Not simply that established that the system overall looks like it's working okay, but in the case of the individual who is the data subject, that that person can have some clear indication that their information was handled correctly according to the rules.

If we've learned anything from the Internet, it's that all of these processes get pushed down to the level of the individual. We have to do the same thing for accountability with data protection practices. Seeing my colleague and in his bow tie, leaning toward the microphone, I'll thank you all very much.

MR. ROSENZWEIG: That's great. Thank you, and I will join in exulting our sartorial excellence. Dr. Sweeney.

MS. SWEENEY: Thank you. I've just started my watch to make sure I stay on time.

Chairman and the members of the board, thank you for this opportunity to address you and to talk today about emerging technologies.

MR. ROSENZWEIG: Would you turn on your mic?

MS. SWEENEY: That are impacting privacy. Let's see how long it stays this time.

MR. ROSENZWEIG: Start your watch again.

MS. SWEENEY: I better keep going while it stays red. I'm the founder of the data privacy lab which we started about five years ago at Carnegie Mellon, and we do a lot to work on real world problems. Often in academia, the problems are so abstracted away from the real world practicalities as to lead the details -- sort of the real world details get left behind and therefore the things that come out the other end of academies aren't always as applicable. So we sort of cut out the middle man, if you will, and began working on real world problems.

I'd like to share with you some of the results we found on four of our projects. We have a lot of projects. You should feel free to visit our web site at privacy.cs.cmu.edu. I have put together a testimony and at the back of my testimony, which each of you should have, are one pagers on each of the projects, and there's more information on the web site with papers and so forth. These projects are basically going to be about facing the identification, bioterrorism surveillance, identify angel and web cam surveillance.

So one of the things that we've noticed, and I certainly am sure that you're aware as well, is that there's a lot of common misbeliefs, and what we've been trying to do is really tackle those. There is a false belief in the public that in order to be safe, they have to give up privacy and what we show is that now in many of these cases you can have both. There's also a common misbelief among those who develop ubiquitous technology, in order to have data that's going to be useful, you have to have explicitly identified or controlled releases of the data. And if the data is really efficient of anonymous, that is therefore useless, and we're going to show that that's also not true.

Now, one of the problems faced that people don't understand, is exactly what does it take to make data unique or reidentifiable. Back in 1997 I had done some work to show how seemingly anonymous information, because it didn't have name, address or Social Security number on it, could be reidentified to identify the subject of the information.

At that time, the subject was medical information from employees from the State of Massachusetts have been released, and the record that was reidentified that got the most noteworthy conversation was that of William Weld when he was the Governor of Massachusetts at the time, and it simply required linking on zip code, birth date and gender to voting list to reidentify the population.

It turns out that 87 percent of the population is uniquely identified by date of birth, gender and zip code. And I use this as an example to say that things that look anonymous often aren't, and we have to do much better than ad hoc techniques even in they involve encryption and so forth because we've been able to demonstrate in a lot of these ad hoc approaches that they're not sufficient, and that instead, you have to replace them with the real scientific basis. So let me give you some examples of a scientific basis of these things - - how data can be give out freely with some approvable privacy protection.

In the first example I'll talk about is basically identification. After 9/11 there was a lot of people in law enforcement who wanted more access to video recordings without the use of a search warrant. But because of the increased improvements in recognition software, there was growing concern with using the data base of driver license photos, they would be able to track all of the people all of the time. So we became interested in solving the problem, how could we share video data in such a way that all of the facial details were made, where the person is looking, various aspects and characteristics about their face, leaving as many details as possible, but guaranteeing you that no matter even if face recognition software was perfect, people could not be reidentified. And one of the things that's interesting is that it sounds like a trivial problem, and we looked at a lot of ad hoc things like putting blinders over the eyes or over the eyes and nose, and showing only the mouth. We looked at pixelation like CNN and other places used. And it turns out that under ideal circumstances, none of these techniques work. Even though as us as humans they look like they're sufficiently anonymous. In reality, all of them could be

reidentified. And, in fact, pixelation which is often used on the news, actually improved face recognition. So clearly this was in the wrong direction. How we did find a solution was by averaging image components so that we would take K individuals out of the video clip, the K of the closest images, we would average them and then replace each one with the same average face. It left as many facial details as possible so it morphed the anonymous back on -- in the image, so the anonymized face is talking or looking wherever the other face was. This is quite effective.

And now if you need to know the identity because there really is suspicious activity, you can get a search warrant to remove the mask, if you will, and actually see the person. So that's an example in the face recognition phase, the identification.

Another example I like to give you is in bioterrorism surveillance. We became very interested in early detection. This is the case where a biological agent such as anthrax might be released. We want to know as soon as possible, in order to save lives, whether or not that has happened. And one of the problems or complications there that many people will act like they have the flu. So the real question is how do I know whether there's an unusual number of people who think they have the flu. And so one of the common ways this is done in public health is to try to monitor medical visits daily, to see how many people are complaining of respiratory distress.

The problem is that without changes in public health law, public health laws are reporting laws, that is, given a list of diagnosis, if someone is diagnosed with anthrax, for example, then, in fact, they are required to report. But it in this particular case, most of the people have the flu, that's not a reportable disease, and, at the same time, very few people would have anthrax, which is the reportable disease, and so trying to find it meant that they couldn't get access to the reportable data.

We developed a thing called a privacy-risk assessment server, which is not commercially available and used to show compliance issues to anonymity, and we were able to show that the ears algorithm that CDC uses was able to work with this anonymized data, so we were able to, one, prove that this data is anonymous, and two, show that CDC could work with it. So then what happens, if you see a spike, an anomaly in the anonymous data, it might be enough to say something might be wrong, but I'd only know if I could link the data together, but you can't link it because it's anonymous. So we provided a sliding scale, which is called selective revelation that says now that you have something unusual happening at the anonymous level, we can give you slightly more identifiable data only on those cases. And then if you see something -- if they are geographically related, for example, they might say, hey, something really is going on and lower -- to the explicitly identifiable data, which was going to be covered under public health law. So this kind of sliding scale of anonymity and showing that in general use provably anonymous data was sufficient worked.

The next example I wanted to give is that of Identity Angel. This is particularly related to identity theft and that is -- Identity Angel basically scans the web to determine whether or not there's sufficient information about anyone, any individual, that one could impersonate you fraudulently and transactions. In particular, we became focused on the acquisition of new credit cards. So is there information already out there over the Internet that I could impersonate you, in credit cards. We found thousands of such cases during those, and, of course, this relates very much to the General Accounting Office. It says that this is a national priority in terms of national security threats, economic prosperity, or economic prosperity.

So one of the threats that I'd like to share with you is basically online resumes, so these particular resumes will often have Social Security numbers or part of the information necessary to fill out a credit card application, but when linked with other information that's also available on the web, one has sufficient information in these cases to actually reidentify the individual -- sorry, to actually impersonate the individual in acquiring a new credit card, so we have several cases of that, and we're also able to contact a lot of these people, Identify Angels tries to find an e-mail address for them and alert them. And sometimes it's not the person who put the information out there and often they don't have control over getting the information removed, which is kind of a frustrating problem. In the cases where they were a party to putting, say, a resume that had their SSN in it, many people did, in fact, respond, and it took it off and they would appreciate the notification.

The last example I'd like to share with you is a mining images and publicly available web cams. In our previous project we had surfed the web and found about 6,000 publicly available web cams that are viewing people in public spaces in the United States. And one of the things we became interested were uses of these sort of low quality camera images. So we began storing images in ten-minute increments from many of the cameras that were there, and then we began counting and tracking people over time.

And one of the things that we were able to do is to figure out to what extent could you detect whether an unusual number of people were appearing or not appearing at a particular location, because obviously having additional kind of use to an emergence response situation as well as to bioterrorism surveillance.

The reason I bring that example too, is because it kind of echoes several issues. One of them is that it's a low cost national security system that's already available, but it's available to anyone in the world. That's not only a personal privacy concern but perhaps one of national security. On the other hand, if you do -- if you had gone out and actually put these web cams around the United States to do this surveillance, one of the problems is you get an overload of images, so one of key things that happened here is our ability to write algorithms that could give us useful information, how many people, where peoples,

what are their trails, from one camera to another, where a car is going, coming on and off of exits. So in some sense this has promoted the notion for us of smart cameras in which high level information gets reported not actual images.

In concluding -- so in concluding, the first two examples demonstrate how data can be rendered anonymous and shared freely. The third example, Identity Angels, shows that not helping the public protect privacy can lead to national security vulnerabilities.

The final example shows how surveillance can already be conducted using publicly available cameras but that smart cameras can be much more effective and less privacy invasive. This is a sample of the work that we've done in the data privacy lab and an idea of the kind of work that can be done.

But one major problem is a lack of funding. It's important to note that all of the work that I just described was not funded by any federal agency. It was funded primarily by those people in the lab, like myself, who donated our efforts and times and resources to do this work. We would welcome an opportunity for government funding. None has ever been forthcoming. Thank you.

MR. ROSENZWEIG: Thank you.

MR. ROSENZWEIG: Dr. Garfinkel.

MR. GARFINKEL: Thank you. So Danny Weitzner said a lot of the things I want to say, and I don't have to say them. Chairman Richards, members of the board, thank you for inviting me to speak. I have a quotation I wanted to read, which is, "In the past, personal and political liberty depended, to a considerable extent, on governmental and efficiency. The spirit of tyranny was always willing but its organization and material equipment were generally weak. Progressive science and technology has changed all of this completely."

So that sounds like what a lot of people here are saying, except Aldous Huxley said it in 1984. He was actually paraphrasing something that Tolstoi had said fifty years earlier. This is an old problem.

In 1997, CDM, I'm going to talk about the future because we've covered the present. In '97 a CDMA cell phone took about three chips to make. Today, a CDMA cell phone, like this one here, takes just thirty percent of a single chip, and you can do a lot with that that extra silicone. This is Morse law at work. So one of the things you could, is you could have telephones monitor their users for words like bomb, anthrax, kill the president. You can use a GPS chip and have it notice if it's making frequent trips near a nuclear power plant, and if so, report that, because somebody might be doing some scouting. Or you could even have batch list of cell phones, and if too many cells phone get close to the bad cell phones, you could have that be reported. Microphones can be turned on and off remotely or automatically.

So the second thing going on is that there's, besides being much more connectivity and much more computational power, there's also much more data storage, as Danny said.

In 1995, it took about a million dollars to store a terra byte of data. Today you can do it for about a thousand dollars and by 2010 you'll be able to store fifty terra bytes for about a thousand dollars. So the two things going on is a lot more computability, a lot more data storing. What this means is that much information that was previously inaccessible, because it wasn't being recorded or it wasn't searchable, is becoming accessible.

You've been focusing on computer records. With LaTanya Sweeney's work, it's clear that casual web cams can be made accessible, recorded, plate recognition, automatic speaker identification. The real question is are we going to move into a world in which everything is archived. Archived everything is certainly the model of Gmail, but it could also be applied to home users. For instance, the fastest typist types just a tenth of a gigabyte a year, and a typical Internet user is just four to eight gigabytes of web traffic and Internet traffic a month. The price difference between a PC with a 40 gigabyte hard drive and 160 gigabyte hard drive is about \$60, so we could just have PC's record everything ever typed, and everything sent to receive over the Internet connection, and you could think of that as a long-term persistent archive.

Should users be told of that archive? Should user be allowed to turn that archive off and on? Some businesses are keeping this archive on their employees. Who should have access to that archive? And if it's created, should it be usable for law enforcement? Should archives like that be usable for only active investigations or should they -- and replay or should they be usable for doing searches or for doing marketing.

And then, you know, we have this huge archives of things like Gmail or Google and so forth, and we've been only focusing on the data in the archive. What about the searches of the archive? It's very interesting if people are searching for bomb making instructions or anthrax instructions. A few weeks ago I was searching for information on how to disable the Northeast power grid. I found it, by the way, about ten minutes. Maybe those searches should have been recorded and that should have alerted somebody.

But we can do better. I've got five minutes.

MR. ROSENZWEIG: Couple.

MR. GARFINKEL: We have a lot of computation that's available. We could have bots that go out and make friends with people and chat them up and establish long-term relations. And those bots could target against people we're curious about, to sort of learn their behaviors and gain their trust. So this is five years from now, and those bots

could be backed ended by lots of people in China, they're operating under contract or pure artificial intelligence.

There's increased technologies for sort of peering into the brain. There's a company that is called Brain Fingerprinting Laboratories that has a technology that when they show you a photography, by looking at the brain you can tell if a person has previously seen that photograph or not, and they advertise on their web site that this can aid in terming who's participating in terrorists acts, directly or indirectly, aid in identifying trained terrorists and sleeper cells, and identify people who have knowledge or training in banking that are associated with terrorists teams or acts.

So I think it as the technology becomes much more powerful. We'll also have the ability to probe deeper into individuals. Again, we have to make a decision, what do we want to do and what do we not want to do with this technology? Because if we're only limited by what is technically possible, that's not going to be a limitation moving into the future. So that's briefly what I want to say.

MR. ROSENZWEIG: Thank you very much. You've certainly scared me, which perhaps was your intention. Lance, congratulations, you're the first non-prior speaker.

MR. L. HOFFMAN: Thank you, Mr. Chairman. This is a question for Dr. Sweeney. I noticed that you, at the end of your talk, you said, well, you haven't had any of this research, you and your students have done, funded by federal agencies. Is that correct?

MS. SWEENEY: Right.

MR. L. HOFFMAN: I'd like to know if you -- since we are a committee advising the Secretary of DHS, I'm wondering if you had any interaction with DHS or components, and whether or not you have, what your impression has been of the DHS research program with respect to privacy, and what it might do better or differently in the future to pursue relevant research and might get us ahead of the curve?

MS. SWEENEY: In general, yes, I have had a lot of interaction -- quite a bit of interaction with DHS on some of the projects that I talked about. And what's interesting, not just to DHS but throughout the federal government, especially whenever it comes to national security or law enforcement, is the privacy or the P word, as they call it, is a bad word, and no one wants to end you up in the way of TIA. And one of the lessons that General Poindexter are quick to point our with TIA is that, gee, if we had never talked about privacy, you know, maybe -- or privacy technology, they feel, even at DARPA, and Tony Tethers made it clear, that a funding privacy technology can generate problems, because it seems to draw attention to privacy. But this seems so counter-productive. It just seems so backwards. So what happens is, anytime the word privacy comes out, it gets farmed out to a commercial entity, which isn't really able to deal -- the things I'm talking about are new technologies, technologies that don't exist. They're trying to use old

technologies that have nothing to do with privacy in this context but has to do with yesteryear's privacy and trying to patch it together, and say now we've solved the privacy problem, and they haven't, because, in fact, you know, a lot of science behind it is brand new and not known in commercial settings.

MR. ROSENZWEIG: You wanted to follow up?

MR. WEITZNER: Just to address Mr. Hoffman's point. Let me just correct the record while I have the microphone. I'm not an employee of CDT. I was a co-founder and I'm currently on the board, but I don't want anyone to think CDT is double dipping.

I think that one of the critical points, I'm also, in some part of my life, in search of money to support this kind of research and it's -- I didn't come here to really talk about that, but I think one thing that's worth noting is that the research that I think is required in many ways falls outside of the traditional boxes, as you know, into which computer science research or just about any other academic research is really pursued. I think a lot of the traditional computer security research, frankly, is not exactly what is called for here, though they're certainly learning from those disciplines, but I think one of the challenges that both HSARPA and NSF have, that this is a new field. I think the NSF has made considerable effort to reach out to the existing computer science community and to sometimes recognize that interdisciplinary work is required. I think that other agencies that want to support work in this area ought to recognize that what they're doing fundamentally is building a new field, a new field of endeavor. It's not just funding existing work. Existing work is -- there's a lot of great security technology out there. I don't think what we need is more security technology. So I think that's part of the challenge. It's the challenge that DARPA faced, as you know, thirty, forty years ago in trying to create what is now modern computer science. So I think it's a non-trivial endeavor, but I think it requires more concerted effort.

MR. ROSENZWEIG: Richard.

MR. PURCELL: Thank you, Mr. Chairman. My question is for you, Mr. Willox. You stated that LexisNexis allows individuals to request or removal of their information from non-public database under your control, under certain circumstances. Mr. Weitzner had said that access to personally identifiable information by individuals for review and correction is an important component, and you, yourself, said that you have many government contracts that are information-based services that you provide.

My question is, how does a person know, or what efforts does LexisNexis make to allow a person to know or encourage any American citizen to know that you're in possession of their personally identifiable information. And secondly, what are those certain circumstances?

MR. WILLOX: You want me to start off easy, didn't we. See if I can -- that was a several part question. We do have educational programs in place that educate not only our customers, but the general industry, the government customers and society in general, as it relates to our privacy policies and practices and they take place through organizations, through forms, through brochures, through mailers and things like that. I'd be happy to provide you more detail on that if you'd like that, greater detail later.

As it relates to our privacy policy and allow people to opt out today, that's an important issue to us because we just changed it, I want to say, two months ago, thirty, sixty days ago, and the reason we changed it was because of the effectiveness of our data. Again, we provide our information of data to businesses and government. We don't provide our solutions or data to the general public, and we're very concerned about the effectiveness of our ability to provide to our customers with the ability to make enhanced decisions.

So our biggest concern, if you understand how the fraud mind works, the fraudsters will study those systems and try to find ways to circumvent those systems, and as a result of such, they'll try to opt out or try to create identities or create challenges for us as it relates to those processes and analytics, and as a result of such, we had to become much more specific about how we would allow people the ability to opt out of our databases, so if we do opt out for specific purposes, several of them are, obviously if somebody is a government official and they're in a very sensitive job, if they're -- if somebody's lives is being threatened or endangered, we allow them to opt out for that, so if there's domestic or civil issues, we allow for that. If somebody says that they can live with that, either before or after, we allow for, we allow for that, and there's a couple others as well, so we do allow for it, but again, it's a very specific process.

MR. ROSENZWEIG: Jim.

MR. HARPER: Dr. Weitzner, you hit on a favorite piece of conventional wisdom, and Dr. Garfinkel you took it a little bit too, so I wanted to just question a little bit, that's the concept in storage that is approaching freely, that Google gives you a big chunk of the storage for free, but then counter antidotes that I've collected, just because I'm paying attention to that kind of question, is a visit I made to a -- one of the credit bureaus to their data warehouse in Texas that had raised floors and special fire protection and extra generators and it was built below grade and all kinds of other things that they did to protect the data, and I think it was in tens of beta bytes or terra bytes. I know that's a big difference, but I don't know which one it was. The facility that they put that in would cost them 250 million dollars, and there were two of them, so that's a half a billion dollars to store data that, it's a lot of data but certainly not all of the data. So I question whether storage is close to free on the scale that matters to data mining, text of applications and some of the things like total information awareness. So I think the trajectory you're on is

right, but I think it's a little much in that the economic consequences, the ability to store is going to grow and grow and grow. It's always going to be a little costly to store a heck of a lot of stuff. And secondly, because I think we only get one question, the concept of stovepipes breaking down, while I agree is generally true and certainly can be done technically, isn't automatic, for policy reasons, for competitive reasons among companies. It's somewhere in the middle. There are plenty of companies that have data that they don't share, first for competitive reasons, digital, with each other, we need to preserve that, and I think it's very important we talk about privacy data sharing with the government, and I think that's the stovepipe that needs to be preserved between the two, that is, let's share between the two, that's the real nature of the Privacy Act, I think we can do a lot.

MR. GARFINKEL: Danny has given me his turn. It's very difficult to assure the data will be retained, and it's is very difficult to assure that data will be destroyed. That's my research. I bought all these used hard drives and found lots of confidential information that the owners thought they destroyed. The case that you gave just now is an organization that wants to be absolutely sure that a key piece of information is not lost, but whether you retain or lose information is a statistical question. There's a certain percentage chance, and so they're spending a lot of money to increase their chances that their information will be retained. But for an organization that doesn't need five nines or six nines of probability that their information will be retained, significant amounts of information can be retained very cheaply. I keep, at home on a hard drive, every e-mail message I've received for the past ten years, including all the spam. I have log files of everybody who touches my web site going back five years, every single click. My retention policy is I keep everything. And I have a backup at MIT, and yes, if we had a big fire here and at MIT at the same time, I would lose the data.

In a large distributed system, it's much easier to keep data and much cheaper than in a single system where a single entity needs to be able to do searches over it. So for your organization, where everything had to be online, immediately accessible, that's very expensive. But maybe a smart video system, such as Dr. Sweeney has proposed, where every video camera keeps it's last five years of everything it's seen, that can be done very cheaply because those hard drives are distributed.

MR. WEITZNER: Just quickly to the storage. I think you're absolutely right that good storage is not free, but there's a lot of storage that's not so good that is effectively free. There's no question that the management cost continues, and I think when we're concerned about managing according to policy, such as in privacy policies, obviously there's a cost there, and I think that's an important point.

To the question of stovepipes, I guess -- I think that you're right, that there's nothing essential about the decline of stovepipes, but I guess what I would say is that the

pressure to collapse stovepipes and to have general purpose data integration architectures, I think will come to overwhelm any specific desire to keep that data separate for privacy purposes or any other purposes, so in the generalized data architectures that enterprises have, whether they're public or private, we will have this sort of wide-scale data integration capability. Whether it's used or not, in any given case, I agree, is a matter of policy. It's not something that will be technically determined, but I think that what will be determined by market forces and user needs that extend beyond just the privacy question, is that those capabilities will be there to be turned on or off at will and to the point that several have made. We won't be able to rely on the fact that they're not there, as a sort of a tacit privacy protection, that's really the point.

MR. ROSENZWEIG: Michael.

MS. SWEENEY: May I just respond to one thing?

MR. ROSENZWEIG: Sure.

MS. SWEENEY: I would just say that in the case of the 6,000 web cams that are publicly available, we didn't put any of the web cams there, but under a less than a thousand dollars, we were capturing hundreds of their images in every ten-minute shot, and so that's an example of how much information can be stored in a lower bottom line.

MR. HARPER: I guess I take the point Dr. Garfinkel makes that, as a theoretical matter, it's all stored, but actually the point of concern is that it will be used. It has to be -- some expense has to go into preserving it in usable format, in usable ways, in usable qualities, in usable places. You can go digging through garbage dumps looking for hard drives, but there's no conceivable DHS digging through dumps for hard drives because conceivably they could have something, that's --

MR. WEITZNER: I think that the critical dynamic there is the meta data, the storage architectures, the database structures that are going to help -- that will distinguish these big piles of data from garbage heaps versus something that could be used more readily for whatever purpose, that those are going to be part of our general purpose information environment. They're not going -- you're not going to have to decide to put them there. Your video camera throws all kinds of, you know, temporal meta data and increasingly will probably put location meta data into the images you take. Whether you choose to have it there or not, it's just there and LaTanya can then do stuff unexpectedly with it. So I think the point again is these are sort of general purpose features that will be available that we're going to have to reckon with, not decide whether we want or not.

MR. GARFINKEL: For me, right now, this is real quick, right now we're saying, oh, the information is there but it's not searchable. Ten years ago, a person in your position said who could have imagine recording all that information, but it's a very simple matter to put a search API on top of all those hard drives. In fact, you could have a self-

indexing hard drive that automatically indexes all the information on it. So if that is done and you have a search API, then the fact that all this information is out there, the government could say, DHS could say, we want to have a priority access to be able to search anybody's hard drive, and that could easily be instituted. So don't rely on the difficulty of search as a way of protecting privacy.

MR. HARPER: Agreed. Luckily DHS can't for policy reasons.

MR. ROSENZWEIG: Not yet. We're moving on though. Just so everybody understands, we, at least check, had only one speaker in the public discussion minutes, thirty minutes at the end, so I'm going to exercise my prerogative and eat a little bit of that time for this panel and for the next panel, but that's a note for the people in the public, that if you want to get on that public speakers list, do it now because if you don't do it in the next five, ten minutes and you come later, you know, sorry. We've used the time for other purposes. Michael.

MR. TURNER: Thank you. My question actually is for the three academics and computer scientists on the panel. Just very quickly, we're wrestling with some issues in our attempts to develop an approach to assessing certain applications at DHS and in the context of a cost benefit analysis, and on the benefits side, we're attempting to engage the efficacy of potentially certain applications or programs, and we're putting quite a bit of intellectual thought into how we go about doing that.

In your role as academics, I'd like you to provide me a grade for the efficacy of any of the DHS programs, and feel free to be specific or more general, with which you're familiar, and furthermore, with respect to the use of technologies in the context of data privacy, data security and data integrity. So two separate grades but one question.

MS. SWEENEY: I'll go first. I think your notion about the double grades is right. I think there has to be -- there is a kind of optimization, a grade on the one hand as to how effective is a deployment. There might be a privacy invasive technology, how effective is it actually being? Then the question is to what extent are you being able to return that privacy invasive technology into a privacy enhanced one. I think DHS has had problems on both fronts. This is my view. I don't study DHS so I don't intend to be an expert about DHS. The problems I've seen in my encounters, and I have been -- this reliance on private industry, to provide technologies for problems that are really, we don't know how to do is a real problem, because it gives you bad grades on both. And the world, it's very simple, IT and IS, and many of the people in the commercial arena, are sort of stuck with the technology we have. They can put glue technology together to make the major pieces work, and to the extent that that's a solution to the problem, then that's great. But if it's something we don't know how to do yet, computer scientists don't talk about today's technology, they talk about what the technology of tomorrow will be, and so they're dealing with problems in a different, an erratically different way, because they get to

create new technology, and these are totally blurred concepts on both sides of the coin in the DHS discussions to which I've been a party.

The other problem has been confusion, as was pointed out earlier, about computer security versus this kind of privacy technology. These are not the same as encryption authorization/authentication issues. It's not solved by those problems. There's been confusion around that, so there's been funding going to security types of views thinking -- with the label on it, that it's solving the bigger privacy problems.

And the last part is that there's been a confusion about what exactly are the problems. You know, what are the top ten solid academic problems? Well, the best insight I had was in a visit of DHS panel in San Diego, where I actually talked to the people who were in the field who were trying to work with the data. I mean you've got this unbelievable richness to the problems that they actually face. And so there's this disconnect normally, and I was the only academic in the room, and to this, normally, this incredible disconnect between the academics and those who are really in the trenches. I would really love to see us move towards sort of like at the end of World War II, where Allen Turing, a very famous computer scientist, many of us consider the father of today's computers, was actually trying to solve a very practical problem in wartime. How do I break the code? And this led to a whole generation of computer science. I think it's that level of bringing the practicality to the academic that's really needed.

MR. ROSENZWEIG: Either of the other two. We'll take one of you so you flip a coin.

MR. WEITZNER: My TA is not in the room, so these grades are going to be arbitrary and unsubstantiated. I would say, as to the passenger screening systems, I think -- what did I write down? I wrote down A for effort, B, B plus for public engagement. I think there has been really a notable amount of dialogue with the public, which I think is to DHS's credit, and frankly, as you know, extraordinarily unusual when people are designing computer systems, there's not a lot of conversation with users or those involved in the system, so I think it's clearly an enormous challenge. I guess I think, on the kind of transparency that I would hope to see, the transparency of the reasoning and inferencing process, I would, frankly say, it's sort of somewhere between F and incomplete, because it's not there. And I think -- again, I put that in the category, in many ways, of difficult design problems that are novel in some ways to design systems of the scale that you're talking about the offer, users in data such as transparent access, I think that's hard, and I think on accountability overall auditability, I think there's kind of an incomplete, and there's clearly a lot of work to do that there.

MR. GARFINKEL: There's another access which is very important to consider and that has been the deterrent affect that these efforts have had. My contacts tell me that even if these screening systems aren't working, then even if they're inconveniencing a ton

of people, that bad guys have changed behaviors because of fear of these systems, bad guys have changed behaviors for fear that there's pervasive monitoring on the Internet. Whether or not it's actually happening or not, that's another question. But I am told that there has been substantive changes in behavior as a result of the perceived effectiveness of these systems.

MR. ROSENZWEIG: We have time for one more in this slightly extended group, and Joanne, since you haven't had a shot today, you're on. Microphone, please.

MS. MCNABB: Have you considered a concern that I think many people have, of the government, and Homeland Security in particular, using private sector databases, involves the accuracy or lack thereof in a lot of public record information captures and collated in databases, have you considered allowing individuals to object to, correct public record information that you have where perhaps the wrong records get associated, which I have experienced myself in public record databases?

MR. WILLOX: Yes, we have. In fact, we do have an access policy that allows individuals to get access of our records and be able to review them. We do have a redress where it allows them to come in and correct those records. Again, the quality of our data is critical or we wouldn't be in business, to be quite candid with you. And the ability for us to correct the data where the corrections are within our control, as opposed to at the same time working with a redress situation, whereby we may point that individual to the actual data supplier to correct it is something that we also do. We have a very large customer service group of over 300 people that work the folks in these issues.

MS. MCNABB: But if it's a question of, if the person is asserting that the wrong public record data was associated with his or her name, so going back to the source isn't going to do it, it's you who did it?

MR. WILLOX: If it's an instance whereby it's something that we did in our data fabrication or our data loading, we would correct that.

MS. MCNABB: And what are the standards of proof in that situation?

MR. WILLOX: Standards of proof.

MS. MCNABB: And to say, that's not me.

MR. WILLOX: And we would provide that data back to them for the review. Is that what you mean?

MS. MCNABB: And then what? How do I convince you that isn't me, that isn't my data?

MR. WILLOX: How do you work through that review process, that's what you're saying.

MS. MCNABB: Yeah. I mean is that laid out somewhere?

MR. WILLOX: Yeah, we have policies and our customer service folks are supposed to do that. Now, keep in mind that the fraud people will come in and try to change the data.

MS. MCNABB: Yeah, I know. That's what I want to know about, the standards.

MR. WILLOX: So we have authentication policies and things like that, that when you come in, you have to authenticate yourselves, or communicate via mail to the address and all that kind of stuff, so we have pretty strict guidelines around that.

MS. MCNABB: If I had another question. I would ask Dr. Sweeney --

MR. ROSENZWEIG: Everybody has plays.

MS. MCNABB: -- if Identity Angle has detected the kind of information it's looking for in any public record information online?

MS. SWEENEY: Definitely. I mean I can give you a large list of them, actually, of where we release information that's enough to be -- fraudulently represent people.

MR. ROSENZWEIG: Before we let you go, I want to thank you. I also want to say that we have begun a policy, in the morning, that we're going to probably avail ourselves with you, which is that, as you can see from the fact that there are one, two, three, four, five, six tent cards up, that reflects that at least six of my colleagues didn't get to ask question, so if I may impose upon you, we will be writing to you with additional questions that will give you a chance to expand on what you said and answer some of the unanswered questions, and that will become part of our web site, dhs.gov/privacy. I do want to thank you. This was absolutely fascinating information and panel.

If I could ask the next -- we're going to keep forging through because we are definitely leaving at 4:30. Sorry. We'll keep going.

MS. SOTTO: Can I go ahead and introduce the next panel?

MR. ROSENZWEIG: Yes.

MS. SOTTO: Welcome and thank you for joining us. Jonathan Zittrain is our first speaker. Mr. Zittrain is the faculty co-director at Harvard Law School's Berkman Center for Internet and Society. His research includes digital property, privacy and speech, the role of intermediaries within Internet architecture and the unobtrusive deployment of technology in education.

Ari Schwartz is our next speaker from the Center for Democracy and Technology. Mr. Schwartz is the associate director of the Center. He promotes privacy protections in the digital age and expanding access to government information via the Internet. Mr.

Schwartz regularly testifies before Congress and executive branch agencies on privacy issues.

And our last speaker of the day, my apologies, Mr. Steinhardt, Barry Steinhardt, is the director of the Technology and Liberty Project at the ACLU. We heard from your colleague earlier today in the public session. Mr. Steinhardt shared the 2003 Computer Freedom and Privacy Conference and was co-founder of the global Internet Liberty Campaign. Thank you very much for joining us.

MR. ROSENZWEIG: Before we begin, I want to just apologize for the reporter for the seemingly hurried nature, but we've got a hard deadline, so I want to get as much substance as possible. And with that in mind, could you keep your remarks about five to seven minutes. If I start coughing miserably, that's -- because we have many more questions than we can get to. Jon.

MR. ZITTRAIN: Thank you very much, and thank you to the committee for coming up to Boston and convening here and for convening generally to keep an eye on such important issues in these times. So I'm delighted to be here. Feel honored to be on a panel with these colleagues and a tough act to follow from the last panel indeed.

So as I thought about what to use my five to seven minutes before coughing for, oddly enough, I found the law professor in me winning out over the technology geek in me for what I wanted to talk about and that surprised me and I thought I should explain why.

Among the big questions that a group like this needs to ask and, of course, the Department itself and the government itself needs to ask, are what kind of data do we want to collect, or are we going to collect? Is there any limit we want to impose on ourselves in the name of privacy on the data that we collect, even knowing there might be some other utility in collecting it.

And secondly, what are limitations on use? Should we collect it? Are there rules we should come up with on the use of that data in the name of privacy? I think these are both important questions, and I will not try to answer them today. In part, because I have a sneaking suspicion that a water will always find its level, and that if there is useful data out there that can be gathered in a relatively inexpensive way and used to search an important end, privacy will, most of the time, yield. And knowing that reality, I think we want to look to additional mechanisms, complimentary mechanisms to try to vindicate privacy interests in a very real way, even as we know that data is collected and used. And for that, I want to focus on a very distinct problem and one that is probably raised in a somewhat timely fashion, given the revelation of who was Deep Throat in the preceding week, and that is the government abuse of sensitive personal information.

When we turn in a tax return and essentially tell the government every important financial transaction in which we have engaged in the past year, we do so with some expectation and hope that we can trust that data will not be abused by somebody out to do something bad. And if there's anything we learn here in a law school environment, we're etched over our library cornice, in Latin is not under man but under God in law, it is that we can't and should not, as a free society, simply trust that our officials will always be as pure as the driven snow, even though any given official that you would present to me, of course, I would assume and presume is.

What we have then is a situation where we'll be gathering information. We have a compatible, as the last panel explained, to gather far more of it than possible in the past, because the cost of doing drops so precipitously. And a situation in which more people will have access to it. In large part, thanks to the fact that it's all networked, so it's awfully easy as an analyst somewhere to get to the data, if you can make a case that you have reason to do it, and for which we have people accessing the data in unexpected ways, because we want to encourage people to think outside the box as they try to solve problems related to terrorism and national security.

So that makes me think back, and this is where the law professor part comes in, to the warrant requirement. The kind of, I won't dare call it quaint, but the notion that one would bother a magistrate in his or her chambers and say hi. You don't know me, but I'm the police officer from down the street and I'd like to tell you a little story of somebody we're investigating. At the end of this conversation, I want your signature allowing me to undertake something that without your signature, or that of a fellow magistrate, I can't undertake. And I think there's been a certain talismanic value to the warrant requirement, often colleagues of mine on this panel are fighting for it, versus such things as administrative subpoenas or no warrant at all, and I think it's worth very quickly looking at why.

One is the idea of before versus after. You get the warrant before you do the search. You have to get the clearance before you actually intrude on the privacy. I think that aspect of warrants have been receding under the exigencies of needing to do a fast paced investigation, without having to shake a magistrate awake in the middle of the night with a random story, that now the magistrate is caught between a rock and hard space, between saying no, no, no. You'll have to borrow me tomorrow, and it might be on my head if there should be a problem in the meantime. And with the problem of having then to sign and approve, without really feeling like the magistrate has given it the real shake.

The other of three points of a warrant is probable cause. That there's actually some standard that has to be met, and we do not vest in the official wanting to make the search the discretion to say, yes, in my own view, yes, I think that threshold of probable cause

has been met. And there, then, what I find most of interest is not whatever the threshold is, probable cause or reasonable suspicion based upon articulable facts, or whatever it is we're teaching at law school these days, but rather that the standard is vetted by somebody remarkably independent of the person wanting to do the search, literally in a different branch of government.

And finally, I think actually most easily overlooked, is the solemnization aspect of obtaining a warrant. You actually have to make an oath to a judicial officer that what you're doing is in the interest of this country, and that it will advance something, and that independent officer hears your oath and you feel somewhat bound by it. And if you were thinking of straying from your oath as an officer of the United States Government Executive Branch, you might think twice before actually, just as you get on a stand and swear on a Bible or not, but swear, that before you perjure yourself, it's just different from happening to tweak the facts a bit say when a journalist calls you.

So all of that then leads me to ask what this committee might be able to do by way of recommendations to the Department to, as much as possible, capture the most important aspects of the warrant requirement, even for those searches, which turn out to be most of them, of course, for which no warrant is the constitutionally or statutorily required. And for that, looking to see what role, for example, the Inspector General might play as a quasi independent voice within the Department, at the moment a post that was filled with a recess appointment and now is being held by an acting Inspector General, but somebody that can actually have the authority to ask to look over the shoulder of somebody doing those searches and say, wait a minute. What are doing here and why? That could actually uncover whether it's low grade, third rate burglary kind of misuse of personal data or far more nefarious uses that might be demanded by a senior official who, put flatly, may, in some at future date, some future official none of us has met yet, choose to abuse his or her power, I think that's where this committee can make a real contribution. So whether it's through an inspector general or privacy commissioner or even the officer of the chief privacy officer, having the authority to do it and perhaps even having to be consulted, if not before the fact then after the fact, for certain kinds of uses of data and having a technical capacity along with it, so that the report is not made to that officer of certain kinds of searches, it will be much harder to hide. These are the kinds of practical things that I think are quite important to maintain the integrity and trust with which we rightful now hold our government.

MR. ROSENZWEIG: Thank you. You're probably unaware, but we have in the room a representative from the DHS Inspector General's Office who sat up just as you mentioned their names.

MR. SCHWARTZ: Mr. Chairman, Ms. Vice Chairman, members of the committee, thank you very much for allowing me to testify again here today on behalf of CDT. CDT

has had two other people testify on our behalf, our president Jerry Berman spoke mainly about some of these same issues that we're talking about here today, and our Executive Director Jim Dempsey spoke on behalf of the Markle Foundation's task force at the last meeting in Washington. Therefore I'm going to try and go one level deeper into detail and talk about some of the laws and policies that cover private sector data by use of the government.

Simply put, use of private sector data by government was not contemplated in the creation of most of the legal framework for privacy protections for the use of government data. Therefore, there is currently substantial debate about whether the law or policy actually applies to today or how it should apply. And to go into specific detail about what I mean by that, I'm going to talk about the Privacy Act.

The Privacy Act, as most of you, I know know, among other protections, prevents creation of completely secret databases, for the most part, insures data quality, limits collection only to that that is necessary for the purpose at hand, limits the sharing of data collected for one purpose to be used for another and creates transparency by allowing individuals to access information held in the databases.

The controversial over the private sector data comes in, in a few different areas. First of all, under Section M of the act, which is the contractor clause of the act. Contractors are specifically covered, but then there's a question as to whether use of private sector data is considered a contractor under the Act or not.

Credit reporting agencies are specifically exempt, but non-credit recorded agencies are not specifically exempt. That's left the question open for debate. Many agencies do not consider private sector databases covered. For example, the IRS and the FBI use data, from what we know, seem to use data that is not specifically part of the credit agency under the Fair Credit Reporting Act, yet they do not consider that to be Privacy Act systems of records. Therefore, there is some controversy where that comes in. Also, they're having cases, including at DHS, where information has been mixed, where private sector data has been mixed with government data, and in DHS it's mostly been done for testing purposes. And in those cases, from what we know, again, we don't know all the details, we do know, perhaps some more has come out in the Inspector General's report, but from what we know, CDT believes that these probably aren't Privacy Act systems of record that have been created, yet they were not disclosed as Privacy Act systems of records. Again, they were only done for testing purposes, so there's a question of how that shakes out in contrast with the law.

Another area of where this comes up in concern is for privacy impact assessments. Under Section 208 of the Government Act of 2002, new databases, or databases have been substantially changed, that contain information on ten or more people, have to create privacy impact assessments. Privacy impact assessments allow agencies to gauge the

potential privacy impact on individuals of a particular new project or policy or a specific technology. Some of the agencies have taken us very seriously. DHS, under Nuala O'Connor Kelly, has done an excellent job in really taking the lead in creating those practices in the scope and qualities of these PIAs. However, the law did not address the issue of private sector data again, and OMB's guidelines specifically exempt requirements for PIAs for private sector databases. Again, that's not saying that you can't do them, that's saying they're exempt from requirement to do that.

So to follow up -- to address these two points and what our recommendations are, I guess our recommendations are that you recommend in this particular case. DHS should, number one, make clear that Privacy Act applies when data is merged. When you take private sector data and merge it with the government's existing data for testing purposes or otherwise, the Privacy Act should apply. You're creating a new system of records that has privacy consequences and that's the reason that the Privacy Act was put into place.

Number two, to build Privacy Act-like protections for quality, data quality for disclosures, for redress on private sector databases that are used by the government, or alternatively, just make clear that Section M, the contract clause, applies to these types of databases.

Number three, draft PIAs or require contractors to draft PIAs for any program that's using private sector data. So those are the simple three recommendations that we have and I look forward to any questions.

MR. ROSENZWEIG: Thank you. That's great. Exactly the types of recommendations. I don't know if we'll agree but that's exactly what I was looking for. Barry.

MR. STEINHARDT: Thank you very much, Mr. Chairman. I appreciate the opportunity to speak here this morning to the advisory committee. I too wanted to think about how to focus my testimony, given the relatively brief period of time that we have this morning, this afternoon. And thought that what I want ought talk to you about is something I want to put on your radar screen. It's something that occurred after the committee was brought into creation, and that's the passage by the Congress of the Real ID Act. Real ID is probably the watershed event. The passage of Real ID is probably the watershed event in the modern struggle over privacy rights. Now, why do I say that? I say that because real ID will, in fact, if it comes to fruition, if it's implemented in the way I think it was intended by it's sponsors, create what amounts to a national identification card.

As many of you know, in the form of a uniform state driver's license that is to be designed, mandated by the Department of the Homeland Security, and to which the states would have to adhere if they wish to have their citizens recognized for, essentially, any

Federal purpose, boarding airplane, opening a bank account, or any other purpose, which the Federal Government might impose.

But Real ID is not your grandfather's national ID card. This is a modern 21st Century ID card. Some of us think of it as a national ID on steroids. It's an extraordinarily powerful card. It will not only contain all of the identifying information on the front that now sits on our drivers licenses, mine, but it will contain new information, for most states, including one or more biometrics. For example, digital photograph for the purposes of facial recognition. It will also have a much more significant, what the statute calls machine readable component on the back of the card. It will be in form either, could be an RFID chip, it could be some other form of magnetic stripe, but it will be a much more powerful form of data storage that now sits on most driver's license. It will contain all of the information that is in the front of the card and much of the information that's in driver's databases that are not theoretically secured from authorized access by the Driver's Personal Privacy Act.

The consequence of Real ID is going to be extraordinary. Not only will Americans be expected to present their ID card all the time to do ordinary things like open a bank account or to board an airplane, not only will this ID card be demanded by police officers and other law enforcement officials all the time, but it's going to be, no pun intended here, magnetically attractive to the private sector.

Every time you go into convenience store, you go, not just your bank, a convenience store, you go to rent an apartment, whatever you do, somebody is going to say, let me see your card. I want prove that you are who you say you are, and they're going to take that card and swipe it through what will become essentially universal uniform reader. Because, after all, every state will be required to have the same technology, same formats, the same data on these cards and they will store it. And there's absolutely nothing now, certainly Federal, although a few state laws may apply, nothing Federal law that prohibits that from occurring. There's nothing that prohibits that enterprise, 7 Eleven, whoever, from selling that information to a data aggregator like Choice Point, for example, let's just pick one. And for that data aggregator turnaround, and take all the data that they know have about you and these electronic dossiers and to augment it with what it becomes the gold standard of data in the United States, the driver's license.

And so then in effect all the protections that we now have in the Drivers Personal Privacy Act against the disclosure of this information are simply rendered irrelevant, by placing this information on the back of the card in a machine readable form, and by having that timely access, not only by the government but a wide variety of private actors, who will create a parallel set of databases. And ironically, the government itself, in squaring the diagram here, will probably, in the end, purchase that database. Because

after all, it will augment the database that it has, because the private sector will have the capacity to capture information that the government entity itself may not have the authority to collect, whether it's the Privacy Act or state stature, or whatever it is, but you suddenly have the ability to merge the governmental or not governmental data.

Now, I tell you all this because I think it's extraordinary important, given the fact that it is the Department of Homeland Security, which under a fairly tight time schedule, must create the standards for the Real ID, standards which are to be effectively imposed on the states in 200 million plus Americans.

I believe it's important that this committee, that the Chief Privacy Officer at DHS take a hard look at these regulations. I cannot think of anything that you could do that will be more valuable than that. This really is a watershed event. Much of the technology and much of the database fears that we had all over the years, about all the data being collected in one place, will be available in one place, be all tied together would be this card, and may well be this card. So I urge you to put Real ID on your agenda and to take a hard look at the proposal as it comes out of the Department. Thank you.

MR. ROSENZWEIG: Thank you very much. I see a few, I'm going to, actually, go first, exercise my prerogative, and this is for Professor Zittrain. I share your fondness for the quaint methodologies of the warrant requirement, and actually I think you left out one of the really good advantages. And it's not just that it's pre-approval for the sake of pre-approval, but because it freezes the record of the justification for the intrusion, and even if best minded people are inclined to have hindsight that is much clearer than their foresight. My question to you, then, with that particular application prompting question, is whether or not you think that it is only independent human review that can provide the same function or whether or not some aspects of the warrant requirements principles can be advanced technologically. For example, the one I just mentioned, about freezing seems easily accomplished technologically, rather than through personal appearance, so I would welcome your thoughts upon the extent to which it enhanced technologies that we're dealing with can address the principles that you are suggesting.

MR. ZITTRAIN: That's a great question, and you're absolutely right that, for example, the important function of freezing the record of justification for a search before the search actually takes place and whatever fruits fall out of the tree fall, you could see simply filling out the form, having it a time stamped technologically, and then it's just in that database to be looked at later should the occasion arise. So in that sense, yeah, I can see a technological piece of it.

I suppose the real issue behind the question, is how we want to think about the analysts and others who have such ready access to such previously unthinkable powerful databases. They're so easy to use and can be so used so casually. The idea that I can go into Lexis right now and find out how many times the word onion has been mentioned in

past year, I mean it's just a trivial search to run, and yet would, before Lexis, would have taken so much energy to have done it, makes it seem casual. But, of course, the very power of the database makes every search potentially privacy infringing in a way that regular searches that require warrants are not. With the only other difference being that maybe it's more humiliating to have a physical search made of you. You have to stand there while they rifle through your drawers. Where if they actually do sneak and peek, you don't experience the humiliation because you have no idea it's been done.

So all of that leads me to say that if this committee thought hard about what a quasi-warrant requirement would look like for the kinds of searches the data mining and screening searches we're imagining are becoming, if not already, will be a common place at the Department, that would be a great service. And coming up with operational recommendations as to how somebody sitting down to do a hundred searches an hour can do it without having to fill out a hundred forms indicating what they're hoping to find. In fact, it might even just be I'm on a fishing expedition, but asking them to assert, in some solemnizing way, that they realize the power of the device they're using, and it's privacy implications, that would be great. And I apologize if that answer is a little more vague than, by all rights, it should be.

MR. ROSENZWEIG: No, it's vague because the problem is vague, or indeterminate. Charles, you haven't gone yet today, so.

MR. PALMER: This is a question for Barry. By the absence of the tie, you notice I'm the token geek. I'm going to keep that up. When you think about identification, we've been talking about identification schemes for a long, long time. We're getting better, and we have some new idea and technology industry and research and academia. My question is, and I don't mean this to be unkind, but can you conceive of a technology or a day when it would be enough protection and audit and so on, that the opinions you represent would actually go along with an ID. I mean if I go to MIT and talk to Simpson and his crowd or somebody at Cambridge in England and tell them, here's what I want to be able to do. I want to be able to access every access of this data. I want it encrypted in such a way, I want this, the geeks will do it. And we will, perhaps in time, find a way to make it tractable to actually be used. But my question is really one more of, can we even try to solve this problem or will it really just not going to get away from the idea that it is galactic ID. Is it technology or is it society?

MR. STEINHARDT: Well, I think this is always a mix of technology and policy. I would say your technology and law, which I'm going to cite as different than policy. Law meaning rules to which people are bound, including government agents.

The problem is that this technology that may well be theoretically possibly, to allow us to segregate data and protect data in ways where rules are followed, laws are followed, you know, may happen, but it's not going to happen in the next three years. In

the next three years, every state in the Union is going to be required, three years from the May 11th of this year, is going to be required to issue the Real ID. And if things stand where they are now, there will be neither technology nor laws, rules, which cabin how the Real ID can be used. That's the reality with we face every day, which the clients that represent face every day.

And, you know, it's an interesting question that you raise. Is it possible to develop a series of technologies to cabin this, I hope so, but those technologies are going to have to follow the rules, and we haven't developed either the technology nor the rules.

MR. ROSENZWEIG: Ramon.

MR. BARQUIN: I'd like to follow up. It's very much along the lines of the technology and law, and I think both Dr. Weitzner and Garfinkel touch on it. And this is the whole issue that right now we are struggling because we are in a frontier where the technology can do all sorts of things. I think it was Dr. Zittrain that said that the privacy will yield because the demand, the marketplace will often take other directions, and what we're struggling with is a model for governance as we move towards a higher plain. And when, initially, you know you opened up, Jonathan, I thought you were moving in that direction, because all of those early words, I was going to put down my tent card, but then you went off into the war, which it's an interesting issue, but I'm really asking whether, because this is something that other technologies all throughout histories have presented us in the past, are there governance models that we should be looking to as ways to muddle along here, until 200 years from now maybe we've got it down pat, and there is law where there needs to be laws, and there are policies and protocols and accepted rules of behaviors, so.

MR. ZITTRAIN: I can take a quick comment on that, and I'm sure others have views too. It is a really tough problem. The only thing I can come up with, knowing that -- I think the only time I can think of that this country has explicitly said no to a new technology, really.

(Cell phone ringing.)

MR. ZITTRAIN: Aside from cell phones, which we don't say no to, was the supersonic transport. We just decided that it was okay to have to wait six hours to get between New York to LA, and we didn't want the sonic booms, but that's kind of the exception, in a way, that shows the rule. And for that then, that's why I actually, in my remarks, made that, seemingly, ninety degree turn into the warrant requirements because thinking that that simple phrase, calls maybe monitor for quality, when applied to keep a salesperson on the right track, when selling a 995 trinket to somebody, could have a lot of power for somebody running searches knowing that their search may be monitored for quality and for privacy compatibility, that might be not a bad operational stopgap.

Somebody outside the Department still trustworthy maybe watching what's going on, until we figure out what the right norms and rules are, writ large as policy, for the use of this very powerful set of databases that we now have. Joe. Barry, I'm sorry.

MR. STEINHARDT: Thank you. Two things. One just to answer your question. There are, of course, models for the rest of the world of privacy laws. Those models actually follow on principles that were developed through the United States, the bare information principles back thirty or more years ago, but, you know, you can look at European laws, Canadian law. There are laws out there that begin to grapple these problems. Now, to some extent those laws may be having some difficulty in to keeping with modern technologies, but I think eventually they will. So I do think there are models that we can look at.

But Jonathan just said something that I thought was important, and it needed some additional edification, which is that too often we ignore the question of whether or not a proposed security measure will actually work in the sense that it will make us materially safer. There is an example of another technology which has been prohibited, some contacts certainly not accepted as contacts and that's the polygraph. You probably know most courts will not allow polygraphs to enter into evidence as truthfulness or untruthfulness, is generally prohibited from use in most private workplaces. That was a decision that was based not only on the particular technology notion that one could judge truth and when civil liberties consequences were. There's also the question of whether it worked, whether it was a reliable way of gauging truthfulness, which most experts will tell you that it is not. There's a an awful lot of technology that the Department of Homeland Security has experimented or used. There may even be huge systems, like Secure Flight, where they haven't demonstrated, in fact, it's going to make us any safer, that, in fact, it will work. And then there will be important questions. We don't even need to begin to debate the question of whether or not a particular program or particular technology is going to rob of us our privacy rights, or our civil liberties if we don't get any benefit from it. Why bother have that debate, if it's doesn't make us substantially safer, materially safer. So that's one of the things you need to look at. One of the advantages of our technologist on your panel does the stuff that's being proposed here actually work, so we got to look at it.

MR. ROSENZWEIG: Joe.

MR. ALHADEFF: Thank you. I think, in some ways, there's a question that underlies a lot of the questions, and I don't think -- it's kind of like the elephant in the room that no one is discussing. I don't think it's because any of panelists who are presenting to us today are shy to discuss it, because lord knows none of them has a track record for being shy. But the question really that I think isn't raised because everyone just assumes it's obvious, but I think it kind of needs to be stated. And what we're really

getting down to is the question of trust. You're providing information under a certain set of circumstances that may be downstream uses, you may or may not be aware of them. You're looking to prevent unintended consequences arising from that information. How do you trust the conditions under which you give that? Do you have greater trust because you have a warrant? Do you have greater trust because you have technology? Do you have greater trust because you have a law? And the problem is I think we're asking this question sequentially, when it's actually a multi-dimensional exam. Because all those become factors in the answer, and there's no silver bullet and it's not one size fits all, and it's lots of different aspects of lots of different features come out to provide the solution. Because there are consumer benefits, there are law enforcement benefits, there are benefits to lots of different stakeholders, to different types of technological solutions, different types of uses of data. They're also questions of effectiveness. They're also questions related to how you deal with it. So I guess from the panelists, and I'm not sure if one or two of you want to answer, the question becomes, how do you get to a multi-dimensional approach to the solution because that's, in many ways, what we're tasked with and what would be useful advice?

MR. ROSENZWEIG: Ari, your light is on.

MR. SCHWARTZ: It's a large question. But I think that -- I do think that it goes beyond just trust. I think that there -- that when government -- if government, you know, intentionally abuses the use of data, they have data, they're using it for a purpose they weren't supposed to be using it for or making up reasons why they're using it afterwards, that has -- that goes beyond trust. That has to do with fundamental fabric of society and how government functions and functioning of government. You can say that that's a trust issue, but then in that case almost anything is a trust issue. So I do think when you're talking about government use of data, it goes far beyond that, and that that's why we have -- we've had laws in place, and, you know, in some ways we've seen some of that stripped back, and I think that that's where you start to -- and then as that starts to peel back, we start to see movement toward new uses of the data, things like Real ID Act. And without questioning whether it actually works and without questioning what the controls -- the controls on government for the use or the private sector for those uses are. So, I mean, speaking in terms of the big picture, I think that, you know, that getting at the idea of oversight of building in the transparency of the system is essential, but I also agree with Barry's point, that, I mean, the first thing to say is why aren't we talking about effectiveness? Why does privacy always come second when it comes to these questions of, you know, whether these programs work at all? I mean first we could check the effectiveness, then we can talk about the privacy concerns, and then we can move forward with the program, instead of moving forward with the program, checking the privacy concerns and then checking for effectiveness.

MR. ROSENZWEIG: I'm going to interpolate a quick question, because you both mentioned effectiveness. I want to know whether you think we have the skill set. I mean, we were all picked for our interest and ability to understand privacy concerns, less for our ability to measure the effectiveness of algorithm --

(Phone ringing)

MR. ROSENZWEIG: Whoever that is, please turn it off. Thank you. I appreciate that.

So, you know, do you think that that's something we can do, and if so how, since I'm kind of skeptical that I have that skill set.

MR. STEINHARDT: Well, I think you could do it. Look, let's take a real example here. Let's take Secure Flight, for example, or it's predecessor, the CAPPS program. The question that I think needs to be asked of the components, meaning the government officials who want to build this system is, not simply have you tested whether or not you can, in fact, protect the privacy of the data that your going to collect, but have you tested the question of whether or not you can really pick out the bad guys out of this? And how have you tested that? How have you convinced yourself that this will, in fact, work? And I don't think, and that's something that policy makers have to do all the time, ought to do all the time. It doesn't necessarily require the appropriate Ph.D to do that. And I would urge you to do that. There are, of course, people on this committee I know who are very skilled technologist by training and practice, but I don't think that even those of us who are lawyers have an inability to ask those questions, and to judge whether or not the answer makes sense.

MR. ROSENZWEIG: Jon.

MR. ZITTRAIN: My heart is with Barry and Ari, but I fear that for many of the most entrepreneurial in a good sense, ways in which we may put technology to work to solve hard problems and preempt terrible disasters, we won't know the effectiveness until we try it. It's hard to know whether in Federal Express is going to be a great idea until you have a bunch of trucks and you can tell people that you can take their package anywhere overnight. And at that point, well, we have the trucks, we might as well keep it running for a while.

And that then leads back to Joseph's questions of as this stuff gets built, if it should get built over the objections of people who are much more, perhaps rightfully cautious, of where that balance should tilt between entrepreneurial uses of technology and caution about privacy.

In addition to the regular multi-faceted formulas we would use, of having the people who run the stuff be democratically either elected or appointed by democratically

elected people, having transparency of their operation, so that you actually have a sense of what they're doing and why, and you believe that what you see is actually what's there.

And finally, accountability of officials, high and low, for wrongs that they might do, they actually have to pay the price for it. The added dimension here, when we're thinking of technology, might be this idea of sauce for the goose, which is to say, if you are going to be using a technology that has privacy intruding possibilities in its routine use, then you should, as the person using it, expect that your uses will be monitored and monitored by somebody you might not know and somebody that doesn't answer to the same boss that you do. And that while that may cramp the style of the person searching, that may be one of the most useful and technologically buildable, if it's done before you deploy the trucks. You know, when you build the technology, you can make it so that it really is detecting blips and usage that may be different from other usage of that system that could indicate searches that aren't from the kinds searches that out to be made.

MR. ROSENZWEIG: Thanks. John Sabo.

MR. SABO: This is going to be tough, moving from quantum physics to neptunium apple Physics. Let me try this. If it's true from the prior panel that, in effect, technology, the reduced cost of storage, let's use data because, you know, it's there and therefore we'll make use of it, or givens, then is a way to start, well, we deal with the monumental issues of efficacy and so on and so forth, to pick up at Jonathan's point about applying some breaks to the trend. In other words, let's take one example and I'll just try to get some your reaction.

Notice, if one of the issues with some systems is that inaccurate data about citizens might actually impact a significant or measurable number of them statistically because it's inaccurate or it was collected or keyed improperly, and it's being used by of government systems, would using notice provisions, post notice provisions, you certainly can't give notice to someone, you could by law, I guess, but given -- you've already collected the data for commercial purposes and now you're mingling it with Federal data, so notice provisions to the passenger or to the citizen, with the ability to request access to and amend records, or even go so as far as to opt out of use of that data, and where there's an overriding issue, for example, terrorists watch list, DHS could, through its own warrant process or core process or administrative process waive that particular record and opt it back in. And therefore what I'm getting at, is that a practical, an example of a practical way to store, that would probably use a common issue of law or regulation policy and technology. Otherwise, if that's not a starting point, what options are there, so that's my --

MR. SCHWARTZ: This is what I was getting at with my recommendations. I mean I was saying that the Privacy Act should apply. We have the rules out there and that's why the Privacy Act was written, and the Privacy Act was written for government use of the data. The assumption was, in 1974, that you were going to get this data and

store it internally and use it internally right, or you were going to have a contractor do that for you directly, right?

MR. SABO: My main point in that is that the Privacy Act, as a requirements of notice from the data collector, which is the government, and all I'm trying to say is that's not in the Privacy Act, even though you were saying it applies, so that's where I was going with the notice provision.

MR. SCHWARTZ: Right. I think, you know, that's -- but there's the notice provision, there's the notice provision that the system exists and what the system collects, and if you are making an exemption for certain purposes, you have to say what those purposes are and you have to live up to data quality standards, assuming that they don't waive them, if they can, as we've seen in some cases of DHS. I mean there is an assumption there that we're putting these protections in place, and that's why the Privacy Act was created. I don't want to -- I'm personally of the mind that the Privacy Act is getting outdated, that the term system of records doesn't work today because it's a very seventies idea about what a database should be. However, I do think that the concept and the general ideas are the right ideas, and that you can apply those same concepts. I think you can apply the law directly to these same databases, but if there's doubt about that, I think you can apply these exact same concepts using the same structures to private sector data, in the government use of private sector data.

MR. ZITTRAIN: Just a quick note on that. It might be sensible for the government to think of almost two categories of data that it's carrying or having others carry on its behalf. The range of speculative data, which is stuff that's been kind of scraped from the great universe out there, and that, you know, it's Delta says you flew on this plane on this date, American Express says you charged that, et cetera, et cetera, that might be used in aggregate to assess threats and for many useful purposes. But when that data is to form the basis of the actual judgment by the government about one its citizens, or just about anyone on its soil for that matter, such as a No Fly List or a watch list or something, at that moment it seems that to offer a citizen anything less than the rights that citizen enjoys with respect to Choice Point or Transunion or something on your credit rating, the idea that your credit rate is more important to give you the right to understand what they know and to correct it than whether or not you can fly on a Delta shuttle at all or without a total shakedown every time you want to go, it seems to me that that's exactly where, yes, you want the citizen who is often in the best position to say whether the data is accurate, to be able to step forward, to have some, as automated as possible process, for knowing what's there and to say, actually, I have three kids, not two. I don't know why you need to know one way or the other, but you're wrong on the number of kids I have. And that would be great to build into it, at least for the data that we push from the realm of the

speculative into the hard data on which decisions that implicate our fundamental rights that that government can grant to withhold where those come in.

MS. SOTTO: Ari, you actually answered the second part of my question, but ask the first part. What are the legal underpinnings of your argument that the Privacy Act would apply to private sector data that's obtained by the government for government use? And the second half of my question is assuming that, in fact, it does not apply, what legal principles would you apply to use of such database? And I think what you said was you would apply the Privacy Act principles.

MR. SCHWARTZ: Privacy Act and, as I was saying, privacy impact assessments as well from the E-Government Act, which is sort of a notice of requirement. And I mean it's to do two purposes. One purpose is to make sure that the privacy has been weighed and that it's been thought about, but early on in the structure. But also to make that information public where we can, and I think in the US-VISIT case is a perfect example of that. I think they answered a lot of questions that privacy advocates and people in the public had about how US-VISIT was going to work in the privacy impact assessment that were answered nowhere else, and it's a perfect example of how privacy impact assessment should work, and maybe raised a couple questions, but then we could ask those directly, about the redress process in that case.

But to go back to the other question, The legal underpinnings and why we think that the Privacy Act does apply. When we're talking about merging data, as we saw in the testing cases, where they took, in one case, there was an example of taking passenger record data and taking, I think it was Axciom data and taking it in and merging it, making sure that the information was correct, and then keeping the Axciom data in the data base. That's a new database. That has new information in it. It's being used for new purposes. Under the definition of the way that system of records notices are done, there should be a new system of records notice for that kind of database. For merging, that's the way that we look at it. The one theory that came out, sort of from the Inspector General's report, was that they weren't actually searching on this data. Well, then how were they matching this information. I mean if they -- at some point they had to tie this information together, otherwise there was no purpose in bringing in this other data unless they were going to use it at some point. So they were planning on using it at some point, and searching on it and searching on least identifiers, because they want to find individuals, that's the point. They weren't using this in the aggregate, from what we can tell. It seems as though the Privacy Act should apply. So that's the underpinnings in the very specific case there of merging data, and we think merging data is going to happen more and it probably should, but it should happen under the existing law.

Then the second piece is the contractor clause and how the contractor clause works. The contractor clause is granted as vague, and that's why I separated out, and I'm not, you

know, I think the argument could follow either way, whether Section M does cover these kinds of databases. As I said, the Fair Credit Reporting Act, credit agencies are exempted, are specifically exempted from Section M. Well, why didn't they exempt other things then? The truth is that they didn't contemplate any of this stuff beforehand, so you can have it fall either way. And I think that that's where the confusion comes in, so you can either say directly that they should apply in these cases, because if the government is using the data as if it had it in house, in other words, they shouldn't be able to get around the Privacy Act by -- the purpose of the contract clause is that they shouldn't be able to get around the Privacy Act by privatizing the function. Right? If they're going to use that data, they shouldn't be able to simply have a private company and hold the data there instead that they would be using internally, so that's the reason that that whole piece exists. If that's the basic function that's going on here, then the Privacy Act should apply. That's the basic, at least from -- you debate whether or not it does apply but that's the basic underpinning. If you don't think it applies, then we're saying you should come up with Privacy Act-like functions. If you do think that it applies, then let's say that directly and have the Department of Homeland Security say that too.

MR. ROSENZWEIG: Howard.

MR. BEALES: Like everybody else, I have a compound question.

MR. ROSENZWEIG: I should say, if we try not to ask too compound a set, we can do all three of the remaining tents and then do the public comments.

MR. BEALES: Great. I'll do my part. First I was interested in your comment on Real ID Act, and I'm not familiar with the statute, and I'm wondering how much discretion there is in the rule making process as to whether it would make it worth while for this committee to get involved in or not, and where, within that discretion, what kinds of limits do you think we ought to be looking at.

And second, I was interested in the comments about efficacy, because I certainly think it's certainly important to assess efficacy of these programs from the get-go because otherwise you can't think sensibly about what's the right balance. But in some cases efficacy seems like a very straightforward inference. For example, the core of Secure Flight, and it's not all of it, but the core of Secure Flight is find people's whose names are on the list of known terrorists, and keep them off of airplanes. I'm pretty willing to infer that that works and would enhance security. I mean I'd like to know if you share that inference or if you think there's something more that's necessary or even there.

MR. STEINHARDT: Let me take your second question first. I agree with you. We do want to keep people who represent a threat to the aircraft passengers off the plane. The question of efficacy, though, isn't as simple as is that a legitimate desire, a legitimate goal? I mean we do look at the question, for example, of can we actually identify an

individual based on the structure of the program. Do we actually have lists of those individuals that we have confidence in, those questions, lots of reports have been written by government's own agencies questioning whether or not the list has been properly managed. So those are real questions that I think need to be looked at before we go forward and say that we're going to subject every airline passenger to this system.

To go to your second question -- I'm sorry.

MR. BEALES: How much discretion is there in the ID Act and where would you suggest limits?

MR. STEINHARDT: I apologize. Yes, there is some discretion in the Real ID act. There is discretion, for example, on what the machine readable component on the driver's license is to be. Is it an RFID chip? Is it a classical max stripe? What is it exactly? There is discretion about what data will be required. Some elements are required by statute, others are discretionary. There is going to be a very, very significant question about how long the states are given to implement Real ID. Is it reason -- even if you decided this was a good idea to do this, is it reasonable to believe that the states are going to begin to implement it in three years, given it's complexity. At each stage in which there will be a regulatory language, there are going to be decisions to be made, and I think that some those decisions may occasion the Congress to reconsider its decision.

It's probably useful for you to know the history here. The Congress has already reversed itself once on this. Last December, when the bill that created the National Intelligence Czar's position, the National Intelligence Bill was passed, the Congress set up a process of negotiated rule making between the various parties, including the states, the motor vehicle departments, the Governor's Association, conference that legislators, including the ACLU and CDT, we were sort of the public representatives.

MR. SCHWARTZ: Law enforcement too,

MR. STEINHARDT: And law enforcement. And as we were literally one meeting into that process, which surfaced all of these issues, which are very complex and difficult issues. When the Congress then reversed itself and said, no, no, no, that was mistake, I guess they said, we're now going to simply give DHS the authority to impose rule. But DHS will still have to create these rules. Somebody is going to have to come up with the money to implement this. It's not going to be cheap, and money is not appropriated certainly by the Congress, states are very concerned about it. I think there's a lot of room for play here, and I think in the end, it may be that the complexity and the cost of this, at least at this stage will cause the Congress to reconsider.

MR. SCHWARTZ: There is -- let me add on to that, real quick. There are a number of things that -- I mean there's less discretion than was in the 9/11 Commission building that Barry referred to earlier, but there's still some discretion in there. A lot of the pieces

were just simply undefined or left open. I have a pet peeve that really concerns me, which is the confidentiality laws in the states on permanent address on the front of the license. The new law says that you have to have a permanent address on the front of the license, but half of the states out there has laws that say victims of domestic violence, law enforcement judges, can put their -- can register their real address, but then put a P.O. address on the license itself in order to protect their privacy and security. So there's a question as to whether Congress debated that issue, whether they thought that out, and what permanent address really means, because there's no definition of permanent address and there's no legislative history of what permanent address means. So certainly I would hope they didn't mean to put law enforcement at risk. I don't think that that helps Homeland Security at all. So, you know, I'm giving the benefit of the doubt saying there is room for you guys to work in this and protect security and privacy through it.

MR. ROSENZWEIG: Lance. And I'm afraid Jim is probably the last one, I believe.

MR. L. HOFFMAN: I'm wondering if there's a definition of effectiveness, because I've seen some tension in the comments already on is that. I'm also concerned about effectiveness, and I like the analogy, I think, Ari, you gave it of, well, first effectiveness and then privacy and then okay, let's do it if we sort out those things. I'm not happy with the idea of using the American people as an involuntarily alpha testers for every new system that is introduced or suggested. I'm wondering what any of you would think of the any of a, we already have PIA requirements, something like an effectiveness or rapid prototyping or some kind of requirement where one could not stifle innovation but not go all or nothing. But say, okay, if you have a potentially good system, you can try doing it in a rapid prototyping mode, meaning certain constraints. But it can only go so far before you come back and present a report and get a result. Is that too much pie in the sky or is that a possibility, or is that clear?

MR. ZITTRAIN: I like the idea of pilot programs and of trying the thing out, trying it out on dummy data, trying it out on a circumscribed set of real data, even having volunteers step forward, although, of course, that skews your data set. But sure, I think before you throw the switch on a full thing, my Federal Express analogy, notwithstanding beforehand, hey, if you are running Federal Express, and you could find a way with to do it with a tenth of the trucks just to try it, you would, and the technology may offer ways to do it that doing it with trucks didn't. There's some irony to the fact that the Division of the Defense Department that had done some of the programs that had been most controversial, like Total Information Awareness, I think styled almost all of these, from that to the market and information about terrorists acts, all as pilot programs testing things out, and, you know, good idea, even if the rest was wasn't.

MR. SCHWARTZ: I tend to agree with that. Although I would say that if you're using real data, you still need to have privacy impact assessment. You still need to have --

the Privacy Act still needs to apply. That's where the questions came about right after 9/11, with some of the things that the Army were testing and some of the things that TSA helped out with, and some of those were a lot of questions, I think, arose with the collection of passenger information.

MR. ROSENZWEIG: Jim, we have a minute, so you can go.

MR. HARPER: It looks like two minutes. Just used up ten seconds then.

Mr. Steinhardt, I appreciate your focus on Real ID, and it's something I've been studying as well, something I'm concerned with. I don't know which of you panelists were here earlier when I talked about my testimony last week on Registered Traveler, where I highlighted the card, the privately issued card that will be used in the Orlando pilot, and one of things that I find notable about it is, in fact, it is a privately issued card, subject to a privacy policy, which says we will retain data only at the site, that is, data about how you use the system and then dispose of the data within twenty-four or forty-eight hours. I emphasized the fact that a contract is enforceable on the earlier panel.

But as we look for ways out of what I think is a real problem of Real ID, some states, apparently, are going to offer non-compliant IDs, which is one not complete way out, and I wonder if this privacy protected marketplace offering is another way out from your perspective, none complete but all important parts of resisting and national ID.

MR. STEINHARDT: Let me take a quick look at that. I don't know that it's a good alternative on issue of Real ID. The statute provides, as you know, I'm sure, is that the states can issue a form of license that is only good for the purposes of driving within that state to persons who cannot present the sufficient proof of legal residence in the United States, can't be used for identification, for any referral purpose, which is quickly become most purposes for which driver's license are used now, and will be used in the future as identification.

There is some talk in the states about issuing a parallel license for that very reason, but the cost here is going to be really daunting. It's expensive stuff, to run a motor vehicle licensing system, and to have a parallel system, I think we're going to find to be very difficult. It doesn't mean that there won't be some people who decide to opt out of this system, perhaps buy into a parallel system, but as a general matter, I don't think that that is in the end going to be the solution to that particular problem. Driver's licenses are something that is, you know, uniquely a governmental function, and I don't know a way around that problem.

MR. ROSENZWEIG: With that, I think we could probably go for quite some time. I must thank the panel for your time and energy. I'm told that we have only one request, two. Did they sign up after three o'clock? No, I'm serious. Two people, so if you gentlemen would clear, we'll take each of you for three minutes and then be gone. When

you come up, please identify yourself, any affiliation you might have, and this is, actually, we got time. We got fifteen minutes here. Sir, the microphone, standing right there. No, the microphone standing behind you. Thank you.

MR. SOBEL: I'm Richard Sobel, and I'm a fellow of Psychiatry and Law at Harvard Medical, though I speak for myself here. If I may, I checked and was told I was the only person, if I could have a little more than three minutes.

MR. ROSENZWEIG: I'm afraid the rules apply to everybody, and it's not a matter of numbers, it's a matter of equity amongst everybody else, so you may have three minutes.

MR. SOBEL: Some of what I have to say would summarize --

MR. ROSENZWEIG: We would be very pleased to receive anything you wanted to submit in writing to our a privacy committee at dhs.gov, e-mail address for posting on the web site.

MR. SOBEL: Okay. I'd like to thank the Department and Berkman Center and highlight three features: principles, rights, and pragmatism that make America great and also keep us safer. Simply put, I'd like to recommend adopting the principal of 39 that protects these three distinguishing features. Because surveillance systems, like Secure Flight, Real ID and US-VISIT, not only fail because they are violations of the right to privacy and of travel and of human rights to privacy and dignity, but also fail under pragmatic principles, the question of whether things work was just discussed, DHS policies needs to drop those plans and instead focus on pragmatic cost effective principle 39 solutions.

Cost effectiveness means targeting police resources and using technology appropriately to protect rights and resources. Miranda warnings, videotaping arrests protect rights and save money. Targeting resources on violators rather than drag netting innocent persons is more effective and protects rights. As Jonathan Zittrain said, adhering to probable cause standards not only protects rights but targets resources and is cost effective. If an agency or an agent can't get a judge to issue a warrant, then the case needs to be improved or the procedures changed.

The Boston Globe indicated in an analysis of the Justice Department's of list of terrorism prosecutions have found that 39 people have been convicted of crimes relative to violations of national security. Similarly, the Washington Post found that there have been only 39 arrests out of approximately 65 million visitors under the US-VISIT program, which costs up to 10 billion dollars over ten years, that's a billion and a half since then or about 39 million dollars per arrests. Those funds can be used better if they are targeted. A targeted approach would take the 15,000 suspected terrorists on the IDENT list, not

focus on 300 million Americans or fifty million visitors and apply that budget to targeting those by hiring more agents. This is good, focused public policy.

The Department of Justice Inspector General said that the FBI missed five hints that might have prevented September 11th. Getting a computer system that could have correlated flight school threats in Arizona and Minnesota would have been an effective use of resources. So I would urge the committee to urge DHS to drop programs such as Secure Flight, Real ID and modify US-VISITS to pragmatically be evaluated for cost effectiveness.

This committee can play a very strong role by looking at the Real ID issue, which was not reviewed by Congress. There were no hearings. The Senate refused to put it on its bill. It was simply pushed through. Calling for hearings and even considering calling for its repeal, if you want to protect privacy, that's a tremendous way to do it. People recommended privacy impact assessments, that's a great idea. Every government program should have a cost effectiveness program. Does it work? Is it worthwhile? Will it protect rights? By setting this as a fundamental principle your committee, you can help protect rights and save money. Jack Bennie exhorted the principle of 39 as ageless and money saving ideal. I'd recommend to the committee this is the approach to cost effectiveness and saving rights.

MR. ROSENZWEIG: Thank you very much. Lane, was there one more?

VOICE: Yes, there is.

MR. RAFFRAY: That's it.

MR. ROSENZWEIG: With that, we've reached the end of the agenda. Let me close with a couple of thoughts. First, I want to ask all the people who testified before us if you have your materials and in PDF or other suitable electric format that makes it easy for posting to the web, we welcome those. We also continue to welcome public comments to our e-mail address, privacycommittee@dhs.gov. I, for one, have learned a great deal from this, and welcome come the opportunity to learn. I want to take the moment to express, on behalf of the committee, our thanks. First, of course, to the witnesses who gave of their time to come and educate us, then of course to the members of the staff of the Privacy Office at DHS who made many any of the arrangements for us and arrange for us to come here. And then most particularly to our designated federal officer Rebecca Richards, who took the lead in making those arrangements. We thank you very much. We stand adjourned until the next meeting in September. Thank you all for coming. Oh, sorry, yes. Thank you Harvard again. We did that earlier, but thank you again to the people of the Berkman Center for the room. Thanks a lot for everybody coming.

Concluded at 4:22 p.m.