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

OFFICIAL WORKSHOP TRANSCRIPT

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

PANEL FOUR
HOW CAN TECHNOLOGY HELP PROTECT INDIVIDUAL PRIVACY WHILE ENABLING GOVERNMENT AGENCIES TO ANALYZE DATA?

Moderator:
Mr. Kenneth P. Mortensen

Panelists:
Mr. John Bliss
Mr. Michael Daconta
Mr. William Gravell
Ms. Rebecca Wright

MR. MORTENSEN: I want to welcome you all to our second day of the workshop, as Toby said. Yesterday was a very enlightening day. Today is the fourth panel, but it's the second panel dealing with technology. And with the distinguished panelists that I have on either side of me today, I'm hoping that we'll have a conversation that will lead us into understanding how technology can help us enhance privacy.

Yesterday, we talked about the different ways that technology can be used, in terms of knowledge discovery, and there was a lot of talk of data mining, from the standpoint of its use in computer matching and finding information. Today, we want to see how we can use technology as the tool that is used to create an environment in which privacy is sacred. We want to show that privacy is not the enemy.

As was mentioned yesterday, privacy is merely what can be done with it. I always have a saying that the computer is a very dumb device; it will only do what you tell it to do. And stupid in equals stupid out. So, we have to make sure that, as the implementers

of technology, that we assure that privacy is part of the inputs, because then privacy will be part of the outputs.

What I'd like to do, logistically, is, I've asked each of the panelists to briefly introduce themselves, give a little bit of background as to them, and then we're going to make a round again, where I have asked each panelist to present their specific ideas on this topic, for a few moments each. And then what I want to do is lead into a conversation where we had, actually, a previous conversation we started before this workshop, talking about some specific questions on how technology can be used to enhance privacy, and begin that conversation. And then, as Toby had mentioned, we hope to reserve 30 minutes at the end for your questions so that we can further the discussion that way. So, if we could, Rebecca, would you please start?

MS. WRIGHT: Hi, I'm Rebecca Wright. I'm a professor at Stevens Institute of Technology, which is in Hoboken, New Jersey. I've been there for just three years. Prior to that, I was at AT&T Labs -- Bell Labs, before that split -- in their Secure Systems Research Group. So, I come to this from the research side. I work on research in computer security, cryptography, distributed computing, and especially privacy, lately. And I guess I'll tell you more specifics when we come back around for the second round.

MR. GRAVELL: I'm Bill Gravell. I'm currently the Director of Identity Management at Northrop Grumman. And, although I sit before you today as a buttoned-down, gray-suited businessman, for the first 30 years of my professional life I served as an officer in the Armed Forces. And I would suggest to you, and can demonstrate, as required, that that informs my understanding of, and feelings toward, this issue. I can tell you, with confidence, that there is no one in this room who has a more specific, more deeply rooted understanding of the rights and freedoms that Americans enjoy, vis-à-vis almost all peoples on Earth, nor has a greater passion for the preservation of those, because I have seen, through my own eyes, the consequences of failing to maintain such attention with diligence. I bring that passion to my understanding and to my work in this field and look forward to the discussions today.

MR. DACONTA: Excuse me. My name is Mike Daconta. I'm the Metadata Program Manager for the Department. I'm the Director of the Metadata Center of Excellence, which is inside the Office of the CIO. My responsibilities there regard to creating the data- management program for the Department, and, obviously, privacy is an important part of that. In addition to that, I'm working two other larger federal initiatives, one called the Federal Enterprise Architecture Data Reference Model, which is, how do all the agencies in the Federal Government structure their data -- and, of course, privacy is an important part of that, there is a security and privacy profile -- as well as involved with several information-sharing initiatives; specifically, the National Information Exchange Model with the Department of Justice, to more easily facilitate the

creation of exchange messages. So, I look forward to the panel and how do we integrate privacy into those activities.

MR. BLISS: My name's John Bliss. I am the Privacy Strategist at a product unit within IBM headquartered in Las Vegas called Entity Analytic Solutions, or EAS. We were a small software company until January of 2005, and then we became part of a very large software company when we were acquired by IBM. As to why I'm here, I guess I bring a bit of a personal touch to this, too. Almost to the day, my then-colleague Jeff Jonas, the chief scientist of our unit, and I were in New York, in Manhattan, and we were scheduled to be in the World Trade Center Towers for a meeting on September 11th. Two weeks prior to that meeting, the meeting was moved two weeks forward, and, instead, we were in midtown Manhattan when the attacks occurred. And it definitely shaped my outlook on this issue. And I think it brought into very fine relief really what is a great challenge. Some may call it a DARPA hard problem. And that is, how do we protect our national security interests without sacrificing our rights to privacy and our civil liberties? And our unit at EAS -- one of the reasons I was drawn to SRD at the time -- is -- has been embarked on a project to try and permit knowledge discovery without disclosure, without disclosure of personally identifiable information, through a form of an anonymization technique that we're going to describe in later detail today. I should also say that I'm not a technologist. I am a lawyer, by training, but, happily, am in remission. [Laughter.]

MR. MORTENSEN: Thank you very much, John.

MR. BLISS: Let the Court Reporter note that there was some laughter in the audience. [Laughter.]

MR. GRAVELL: I should point out that I flew out of Dulles on the 10th of September, 2001. I was to have flown out on the 11th, but, due to my wife's work schedule, we moved it ahead a day.

MR. MORTENSEN: One of the things I would like to mention, too, is, one of the things we've realized. 9/11 was one example of an emergency, and certainly the reason why our Department currently exists, but it also exists for things such as the tragedy that's happening down in the Gulf States at this time. And one of the things that we've been talking about -- and yesterday kind of brought this to light for myself and some other of the panelists -- was talking about how the issue of privacy is one that we can't ignore, even during times of great emergency.

And we have to make sure that we are prepared to deal with it. Now, I'm a believer that technology can be the tool that will help us keep privacy in focus even at times when we need to react quickly in order to save lives and to put people back to where they belong and get their lives in order. So, what I'm hoping today is that we're going to learn some new things and be able to combine some things in order to have a

conversation as to how technology can be that tool for that particular thing. Right now, what I'd like to do is have each of the panelists, kind of, talk about what they're bringing to the table, in terms of privacy-enhancing technologies, and perhaps some of the work that they've done in the past that you will find enlightening in this regard. And let's start with Rebecca again, please.

MS. WRIGHT: Thanks, Ken. So, as I mentioned, I've been working a lot on privacy. I come from a background of cryptography and computer security, and, in particular, for the last two years -- we're now two years into a five-year project -- I've been a member of the PORTIA Project, which is a five-year National Science Foundation-funded project. We're funded for about 12-and-a-half-million dollars -- five universities -- Stanford, Yale, Stevens, NYU, and University of New Mexico -- maybe about ten professors involved, all of our students, and then we also have a number of industrial and government partners.

And what we're looking at is, generally, issues of handling sensitive information in a networked world, and with a strong focus on privacy, but also on issues of correctness and accountability and other issues of how you handle the information.

We have five major technical themes, one of which, and the one I'm most involved in, is dealing with privacy-preserving methods of computing on data, whether it's privacy-preserving data mining -- so, how do you do data-mining tasks while having some mechanisms to protect some kind of privacy, and how do you balance the obvious tension between extracting information and protecting information -- or other kinds of privacy-preserving computations. So, it doesn't necessarily have to be pattern-based data mining, but just ways of computing on data that, again, balance the desire to allow certain things to be observable and computable and certain things to be protected or hidden, at least from some people.

Our second major technical theme is identity- theft protection and identity privacy. And, you know, I should point out we're not only looking at, sort of, government uses of data; we're looking broadly at the use of data that we -- you know, that all of us need in our everyday lives. And identity theft can certainly become a big issue, especially with Internet use and Internet commerce and e-mail and people's susceptibility to scammers.

Another technical thing we have is database policy enforcement tools. So, if you want to have all kinds of sophisticated policies about how data should be handled and use those in systems that do advanced cryptographic things, you need a way for the database to actually help you enforce those policies so that you don't lose all the efficiency. Databases -- you know, the reason there's, like, a huge database industry is because there's a need to have very efficient computations that work, not just on a general-purpose computer, but really are optimized and tailored towards databases. So, we want to get as much as possible of the security and privacy enforcement in at that level so we can still retain those kinds of efficiencies.

Fourth major theme is the use of trusted platforms to provide trusted privacy-preserving services. So, if you have -- you know, again, sort of, trying to push things to a level where they can be done more efficiently -- so, if you have secure hardware or secure operating systems that can build in some of your privacy guarantees for you, how can you leverage those to do privacy on, sort of, the higher levels – algorithms.

And then, finally, our last theme is not so much technical but actually our philosopher among us, who's a philosophy professor at NYU, has put forth the notion of contextual integrity as the way to talk about privacy. Because, in fact, if I were to ask each one of you in the room what you think "privacy" means, I'd get at least as many answers as the number of you in the room, and probably more. So, "privacy" means different things to different people. It means different things at different times, depending on what the task at hand is. And, to me, the -- really, it's about expectations.

When do people feel there's been a privacy breach? When something that they didn't expect to happen with their data happens. So, contextual integrity is a philosophical framework to discuss notions of privacy, and one that we're trying to bring into our technical work, as well, and ask the question is this satisfying privacy or not, you actually need a definition to measure your system against. And so, that's our framework for those definitions. So, as I mentioned, the PORTIA Project is five universities, a number of industry and government partners. I also do some other work, some new work on something I call "incentive compatibility," which is looking at ways -- if you have multiple engaged in a cooperation of some kind, and you're worried some of them may cheat, if you can align the incentives properly so that you know no one has an incentive to cheat, that means you may be able to develop more efficient solutions than if you're working against a general adversary. So, I think I'll turn it over, there.

MR. GRAVELL: Well, this is a conference about data mining, and we heard a great deal about that yesterday. I'm not going to talk about data mining, except peripherally. I'm going to talk about identity management. And you will see that the two subjects, while not concentric and certainly not identical, have a degree of overlap, and that's where I'm going to try and tease out an understanding of the implications for privacy in the performance and pursuit of security.

I put it to you that a thing that is created to perform a function, as discussed yesterday very articulately by Pete Sand's group in the third panel, is an object designed by engineers, but is necessarily seen in a specific application or context. To recognize that both of those factors are in play is to suggest that baseballs and dodgeballs can be used interchangeably. It might make for an interesting watch for a little bit, but I doubt that would be viewed as a sustainable solution.

Now, if we apply that to the status quo, and we think about 9/11, we say to ourselves, what is the expected effect of that, as regards identity? Now, I'm skipping over

the part where I already presume and judge that identity is profoundly changed by 9/11. As my bio suggests, I made that call in 2001. I'd suggest that that's clearly true today. But to say, "We need more identity," it doesn't begin to understand the approach to the problem in either technological or social terms. I would decompose the issue into three components. I suggest that we will apply identity management and demand, quote, "identification," unquote, about which more in a minute.

In a broader set of contexts, you may or may not know that in order to park a car in an underground parking garage in this city, you must show the gentleman, who -- almost always a gentleman -- who may or may not have a very good command of the English language, a photo ID. Now, I'd suggest that this person, who may be a very tenuous grasp on American culture, laws, citizenship -- he may be, frankly, an illegal immigrant -- isn't someone who can be trusted to recognize the fifth of seven formats that the Utah Department of Motor Vehicles has used to identify car drivers, but that satisfies the, quote, "identification requirement."

And that brings up the second issue. The mechanisms used for identity management, in almost all cases, are insufficient. They are inadequate. They crack under pressure. And I could talk for an hour and a half about all of that. But I put it forward as an axiom that almost all identity-management systems today are technologically incapable of concerted -- of resisting concerted attack. The third feature -- the first being -- just to review -- the first being a broader set of applications, with implications for costs, scalability, manageability -- and here we bring up the question of data storage; the second issue being the strength of the systems; and the third being the frequency within which one is challenged.

You have to show ID four times in an airport. This has effect, if you have architected the system for serial processing, to create those interminably long lines, with implications for latency and the general drawdown in efficiency, profitability, versatility of the system, as well as -- and here's the soldier in me coming out -- the channelization of process, which is the beginning of true vulnerability in any system. It's the point at which you're predictable to your enemy.

Thus, the argument, I suggest, to do or change nothing in order to ensure that our rights are not altered, is to risk a far more dangerous outcome, and that would be a Draconian future that I shudder to contemplate in which, at some future date, a horrific terrorist event has been achieved on the basis of the defeat of an identity management system that allows some perpetrator to gain access to a nuclear power plant or a critical point in the pharmaceutical or food production chain or a chemical or -- refinery. You'll note that I haven't named the Pentagon or nuclear weapons storehouses. I'm talking about things that are in society, typically guarded by for-profit corporations. In those cases, I predict to you that we would find Draconian measures that would do great

violence to our rights, privileges -- would be enormously expensive, not at all effective, and it would take us a good deal of time to swing back and find the right equilibrium.

I suggest that it's far wiser, as a technological and a policy choice, to act soberly and thoughtfully before the fact. Toward what end? I put to you -- and here's where I'll wrap up -- that, as the point of video surveillance was raised yesterday, you may or may not know that today on this Earth the nation with the highest density per capita for video surveillance is the United Kingdom, Great Britain, a nation that is the fountainhead of all the liberties we enjoy, the place where Magna Carta was signed, and so on. Is this to say that the British, collectively or individually, are less fervent about their freedoms and rights than we are?

I tell you flatly that that is not the case, in my longstanding dealings with the British Government and individuals in many areas. What has happened in Great Britain is that, over a period of many years, they have become so inured by the continuous social pressure of what they call the "troubles," their indigenous terrorist problem, that they have permitted their concept of governance to evolve and to gradually accept this surveillance regime and all associated costs for its management as simply the price of doing business, as simply the way in which they retain some measure of physical security. It's an outcome that works for them. I put it to you that the difference between their case and ours is that we don't have 50 years; and thus, the only other gross alternative available to us as a nation is to embrace disruptive technology, a challenging term about which there's much more that we need to get into, but there are a couple of things that we could think about as how to do that.

One or two -- and I'll cite two and stop -- would be to separate the notion of identity -- that is root identification -- from its application -- typically a privilege. A second would be to examine the point in the system where data is held, whether on the tokens as -- or a front end -- my sleeve has been identified -- or in the IT back point. Now, these are technical choices that will be informed by the social compact between those that hold data and those that apply it. Thank you.

MR. DACONTA: We've quite an interesting panel here. I come at this from a slightly different perspective. As I mentioned, I'm involved with creating a data management strategy for the Department, so I look at these types of issues in terms of the larger picture of managing all our data and what's the intersection between managing our data and privacy implications.

And the key focus that I bring to that -- and I want to bring to today -- is the idea that we are making strides in technology improvements in how we model information to get a greater degree of the semantics, or what the information means, to bear. And so, that's the first area that I bring to this panel, is the idea of -- that privacy data and privacy implications are one aspect of data management. And one of the key things that we're

trying to do in data management is get to a finer degree of fidelity in how we model that information. And that fine a degree has to expose the semantics, or the meaning, of it, in that we first need to know -- as I believe Rebecca mentioned -- there are parts of information that we need to protect and parts that we do not need to protect.

Also, when we get into protection of information, you -- there are certainly valid reasons for accessing information, and invalid ones. Well, all those things are going to boil down to how well have we modeled them, and how formally can we do that? So then, as an example, if I want to check if somebody had a valid use for information and a valid request for types of information across the Federal Government, I need a consistent way of modeling that so that I can check and have some auditing that works. Because all this type of auditing and stuff, it has to be done in an automated manner. It cannot be done by, you know, send people down there to check every query that was made on every piece of data. There are too many queries being made. It has to be able to be done in an automated fashion.

So, that's the first part, data management. The second thing, of course, I think we're going to get into is that we are in need of new techniques. I mean, we all know that our traditional mechanisms for protecting things like passwords -- you know, how many people here have either written down their passwords or can't remember all the passwords that they have? We have hundreds and hundreds of passwords. So, some of our traditional mechanisms for protecting our information are inadequate.

So, we've got to talk about some new techniques. I'm going to be talking about stuff like biometrics and different ways to encrypt things, anonymization. And we do need those new techniques, because we have to break the mold on the way we've done it in the past, because it doesn't -- it's not going to get us there.

The third aspect, I think, of course, is the policy issues. And it really gets into some of the cost benefit of the equation, in terms of, How much can we afford to monitor, and to what level of fidelity? Because anytime we try to ratchet up the fidelity, ratchet up the auditing, there is an associated cost with that. It's like the issue of ratcheting up screening. You know, we can have a perfectly safe world, but, unfortunately, nobody can leave the room, you can't go outside, don't get on buses, don't get on planes. And then we're perfectly safe. So, we can have perfectly safe data if we don't ever share it, we don't ever allow anybody to access it, et cetera, et cetera. So, we have to balance these things of legitimate access and legitimate use from concerns about privacy.

MR. BLISS: In my introductory remarks, I said that part of the challenge is how to balance our interests in prosecuting national security against privacy and civil liberties interests. To do that requires figuring out how to information-share in a privacy-enhancing manner. And it requires, sort of, an acknowledgment that we are in a digital

world and that the notion of privacy now is different than it was when our Founders drafted the Fourth Amendment.

But, nevertheless, the key, I believe, is to figure out a way in the digital world to adhere to the Fourth Amendment and adhere to its spirit. And that means that we don't want to be casting a wide net using pattern-based, probabilistic behavioral data-mining approaches to figure out who might be a bad guy, but we want to start with the known bad guy, with the predicate that that adheres closer to the reasonable and particular -- reasonable and particular tests of the Fourth Amendment than does the pattern-based view. I think the other thing we have to acknowledge is that -- and I think this was probably alluded to by Larry Ponemon yesterday, although I wasn't able to make his talk, is that the American public is willing to trade some measure of privacy for a benefit, that they engage in a cost-benefit analysis.

They're willing to sign that warranty card to get something, and it generally means some erosion of their privacy. They're willing to produce a financial statement to get a lower-interest loan. And it works, possibly, because they can see the reward; it's quantifiable to them. But when you think about what we have to sacrifice to protect our national security interests -- we're purchasing a bit of national security; how do we know that we got what we bargained for?

It's very difficult, because we're talking about the absence of something occurring. We can't tell United States citizenry how many people were caught today. So, you're not going to get the knowledge to let you know that that bargain was one that was worth it, in your view. So, what do we -- how do we get people to overcome that trust issue? When you want to share data, and you want to do it consistent with the Fourth Amendment, and you want people to contribute their data, one way to conceive of doing that is to anonymize the data, because by anonymizing the data, you essentially are engaging in knowledge discovery without disclosure, you're sharing without sharing, you're determining key identity attributes which are necessary for knowledge discovery -- i.e., who is who? Is John Bliss the terrorist, or is it the other John Bliss? And who is he related to? Is he related to a terrorist? -- and you're doing that in an anonymized space.

That is the area in which the argument in Las Vegas -- Entity Analytic Solutions has been working for the last couple of years to produce a one-way hash technology that allows end parties to share anonymized data and then correlate that data to figure out who is who, and who is related to who in the anonymized space. The development is that we've figured out how to essentially do fuzzy matching in the anonymous space and to do so in a way that doesn't require -- and we're going to get --

MR. MORTENSEN: Thank you, John. I think you can see why these panelists were picked to speak today.

We have a wide diversity of knowledge that we're bringing to the table, which says to me that we really have some solutions that are going to be out there. There are multiple ways to be able to find how to solve this problem.

I'm going to back up here a little bit and -- I was very happy to hear Rebecca say that there is a philosophy professor that's part of the PORTIA team, because, coming from a engineering and law background, actually philosophy is very important to me. And I want to, kind of, take something from President Madison that I have carried with me into this job here at the Department -- President Madison once remarked that our country is the hope of liberty throughout the world. And what that means is, is that we are the shining beacon. And, unfortunately, there are folks out there, as we saw on 9/11, and as we've seen in other attacks, that wish to target us because of that. Yet, one of the other things that goes along with that, from a philosophical sense -- not to take away from what John just said about a balance -- I'd like to say that it's not a balance. You can have privacy and homeland security at the same time.

And the reason why is because we have methods and ways of doing things. Each one of these panelists has spoken about that. There are techniques that we can employ that will allow us to be able to do our mission here at the Department, and certainly allow the government to do its mission in whichever part it needs to do, in a way that is not necessarily invasive to privacy. I think that Bill's comment about the difference between the United Kingdom and our country was enlightening because the United Kingdom is the cornerstone to the beginning of our legal understanding, and our core understanding of privacy does come from there.

In fact, the concept of privacy that -- and pardon me, Rebecca, but it's a little chauvinistic -- but it was every man's home is his keep, or, as we know here in America, his castle; and, within such, as long as he remains quiet, he shall preserve his privacy. That goes back to the 14th century, in terms of an understanding that the English had developed as to what we have an expectation to, to our own personal privacy.

But I think it's also important to understand that privacy is dependent upon the individual. One of the things I always teach my students when I'm teaching the privacy-law segment is that privacy is just the term that we throw out there; and really what we have to do to understand privacy is, we have to understand what the "right to privacy" is.

And the right to privacy, basically, is three elements. And that is that it requires a reasonable expectation of privacy. And to understand how a technology will affect privacy, you have to apply those three elements, and to look at them. First, we have to understand, what is privacy in the context of the particular technology system or program that's being employed? What about it would an individual feel invaded? You can define "privacy" very simply by saying that it is the intrusion upon seclusion. The invasion of one's privacy is the intrusion of others into a realm that you have designated that upon

yourself. The second part is to look at what your expectation is. As the individual having to interact with that system or that program, what is your expectation with privacy?

John mentioned one of the things that happen all the time is that our expectation of privacy sometimes is de minimus, because we hand in that warranty card. We are willing to give that up in return for -- another term that Rebecca talked about -- an "incentive." There was something that was returned back to us that we felt it was worth it.

So, we have to look at what the expectation is. And then, lastly -- this is the most difficult and probably the most debated portion of privacy -- is that, is that expectation reasonable? It's always kind of a fun thing to talk about with my students, because I always say, "Well, you know, as I'm talking to you right now, do I have a reasonable expectation of privacy?" And certainly if I pose that to you, most of you are probably saying, "Well, no, Ken, you have no reasonable expectation of privacy, certainly in the context of the conversation we're having now. You're speaking publicly, you're over a microphone, a PA system, there's a transcript being taken," and certainly that is probably the context.

But, in other contexts, such as teaching in a class, even though I'm speaking publicly to the class, the reasonableness of the privacy may expand, depending upon: What is the population we're talking about? Do I have privacy with regard to the entire society, with regard to this entire university, or with regard to the class, itself? So, I think as we think about these technologies, and talk about that, we can't forget that.

Judge Brandeis, over 115 years ago, looked at this concept in an article that he wrote for the Harvard Law Review, and what I find absolutely fascinating is that he was dealing with technology at the time. In 1890, when he wrote the article, "The Right to Privacy," his fear was this new technology called "photography" -- and especially something called a "snap camera," which was now a camera you could actually hold, and it was very portable -- was going to invade into the privacy of individuals just as they walked amongst the streets.

And I think what puts that into context with today is, much of what he talks about in that article, an article that, as I said, is 115 years old, the analysis applies directly. And what we're going to do is continue the conversation here right now in the same context.

And what I want to do now is, I want to open up to the panel and basically ask you -- start the conversation -- in the sense of talking about your different areas, but, how do you think that we can apply technology to enhance privacy? I'd open up to whoever would like to.

MR. GRAVELL: Well, I'll start what I'm sure will be a rich discussion.

I would -- I think that the adjective that Ken chose is exactly the correct one. Much of yesterday's discussion was an intellectual tension about the effort to preserve and retain and maintain. And those were the sorts of adjectives being used. Privacy, in the face of social pressures, perceived threat, and encroaching technology which was going and exploding in a -- in the ways of spiders and octopi.

I think that's the wrong understanding altogether. I think that one has the opportunity to look at technology as providing the potential not only to guarantee the preservation of privacy, but to enhance it relative to the status quo. I'll give you a simplistic example that has touched the lives of every one of you.

How many here have not ever presented a paper check to a merchant who did not know you personally by face, and, in order to have the check honored and accepted for the goods and services you were seeking to pay for, demand to see, perhaps to copy, almost certainly to write on the check, your driver's license number, your telephone number, very possibly your social security number? I've certainly had it happen to me many times. And the consequence of refusing to accept this completely inappropriate intrusion into privacy, the completely inappropriate application of data outside the regime for which it was collected, stored, managed, and intended to be applied, all of which richly discussed yesterday in the context of why the Social Security Administration keeps our SSNs and our career earning history, all of that is done violence by the simple act of trying to prove who you are for the merest of transactions in society.

Technology can overcome that problem with a stroke, and, thus, restore to us, in that small way -- but it's illustrative of a larger process -- the ways in which data, which is collected for a defined purpose, which has been regulated, announced, approved, and is subject to review and audit, can be maintained within those channels. The channels are different. The specificity and the application of data are entirely different for myriad governmental and commercial applications. And, as had been discussed, in some cases we consciously surrender the data for some perceived benefit, in other cases the law declares that we must provide the data.

We are required to pay taxes and required to provide certain information, as defined in the tax statutes. The issue is not that we have to pay taxes, at least not with most of us. The thing that would upset us deeply is the ability of people, other than those defined and authorized, to be able to access that information, much less apply it inappropriately.

MR. BLISS: You know, I have to comment a little on Brandeis. Imagine his grief now. I mean, he was worried about the right to privacy when we were talking about a snap photo. It's of one image captured in time of a person. What are we -- what are we looking at now? We're essentially looking at the capability of taking the digital snapshot of our whole life, not just one image, but who we were before, who we are now, our

banking records, our financial records, other personal history. I'm sure he'd be turning over in his grave trying to grapple with that problem.

And I think that what it also suggests and what might be emerging is another right to privacy, and that may be the right to not have data about ourselves aggregated. I think there is an emerging, sort of, belief among the American people. And, while they are willing to give up some of their privacy, somehow they have -- they're uneasy about the fact that, "Well, I'll -- yeah, sure, I'm going to file a financial statement with a bank, but I'm not so sure that I want that aggregated with every other transaction in my life and then made known to the government so that they can apply some predictive algorithm against me to determine whether I'm the bad guy." That might be an emerging disconnect.

We already have quite a number of disconnects. But, turning back to the question of, what can technology do in this area, particularly anonymization? Imagine a particular problem. Imagine that Disney Cruise Lines has a manifest of those that are about to travel. And imagine that the government would like access to that cruise line manifest to compare it against a watch list of terrorists, or suspected terrorists. Disney is not too inclined to give up that list and share it with the government. The government's certainly not inclined to give its terrorist watch list to Disney. So, we're really at an impasse. How does the government go about its job of protecting its national security interests and the corporation not give up a valuable asset and anger its customers?

One way that can be done, and one use of anonymization, is to take the data store of the customers, of the manifest of Disney, and, quote, "anonymize" that. And, by that, I mean to take the -- to take the names and other identifying identity information and apply a one-way hash to that information, which is essentially a digital signature, an irreversible signature -- as Jonas, in our office, likes to talk about, think of taking a pig and running it through a grinder. It's very unlikely that once that pig has been run through the grinder, you could return it to its pig state. That's what a one-way hash does. It's an irreversible set of indecipherable characters that are a product of names and dates and phone numbers.

You take that information, and you hash it all, so the Disney Cruise Line manifest is hashed. It's put into an anonymized data repository. The terrorist watch list is similarly hashed. That is put into the anonymized data repository. And, at that point, entity analytics are performed against that hashed data, indeterminate who is who and who is related to whom in that anonymized space. And you may determine that there is a match of entity 123 from terrorists watch list A to entity 456 from Disney Cruise Line's manifest. At that point, there would be pointer systems that would publish that alert back to someone with need-to-know. And here's where it becomes incredibly critical.

Technology is not the panacea here. It's one way of solving part of the problem, but, in the absence of business rules, processes, controls, oversight, audits, it may even

make things worse. But this is a technique to allow that information-sharing to occur in a manner that's much more privacy- enhancing. It's not privacy-protective, necessarily, but it's certainly privacy-enhancing. So, in problem sets where you have two entities that can lawfully share information, as the one I've just described -- the Federal Bureau of Investigation is lawfully within its rights to obtain that whole manifest, not just a particular person in whom they have an interest, under Section 215 of the Patriot Act without revealing the personally identifiable information of anybody else on that manifest, without revealing the terrorist watch list to the corporation, and, frankly, without even really revealing some of the other PII of the individual who's been a match, his health history and so forth.

So, it has a particular use in this homeland security context, but I don't want to over- represent it. It is a technology which will promote information-sharing in a privacy-enhanced manner where there is already a lawful ability to share. It will, by its nature, help with the mission- creep problem, which is so central to this, and the re- purposing of information. But, it is probably years before we get to the point where we can cause federal governments -- foreign governments to share with the United States.

There are other issues that Rebecca can certainly go into involving cryptologic attacks against that hashed information that we need to be hardened against that kind of attack before it becomes ready for prime time at that level. But, in the first level, information- sharing between departments or between companies where they have a lawful right to share, I think it has significant value.

MR. DACONTA: I'd just like to -- well, I think there is a lot of exciting new technologies. There's a lot of exploration going on, on new techniques.

But I actually think we need to take a step back and focus on getting proper discipline into the process. There is a lot -- so, when we're talking about identity and the reasonable expectation of a citizen -- let me, for a minute, talk a little bit about Metadata and identity metadata.

Right now, so you know, there are really no strong, consistent rules across the Federal Government on how to model these things. So, as an example, you can go to a database, and I may have -- and both on the commercial and government side I may have a COM called "identifier." And if I -- if I'm a citizen, I know that identifier is something made up, it's some -- say, I'm going into a Blockbuster, and they make up an identifier number. I really don't care if you share that. It's a made-up number. You can do whatever you want with it. But if they have it labeled "identifier," and they're using my social security number for the identifier, then I do care if that's shared, because that has a higher degree of risk associated with that, and more potential for misuse.

So, the first question is -- is a measure of discipline and standardization on things that affect privacy. And what that will let us do -- so what we're -- and that's -- some of that is what we're doing -- what we're trying to achieve with the Federal Enterprise Architecture's data-reference model, consistency and standardization on modeling some universal or common entities and attributes across the Federal Government. And the intersection with that and the privacy profile is extremely important. And this is something that we can do today.

So, technology-wise on that, if we take that and we get a disciplined approach to modeling and we identify the key attributes of different entities that have privacy ramifications, we can then set up rules in expert systems that do some of that monitoring for us, as well as, if we're talking to commercial data, we can force that commercial data to be mapped to these models so that we develop one set of business rules that can do checks and audits against this.

So, my biggest area that I'd like to see a lot more work done is things -- and I -- and I've started work on this with the privacy office -- a set of privacy decision trees and rules on formally modeled data across the Federal Government. And if you think about -- this is -- this dovetails great, because trends like service-oriented architectures are all about: How do we separate the data models from the presentation layer of our applications so that they can be worked on separately? And part of that work-on would be stuff like policy engines validating that the data is okay to be released, I've checked the query, I understand who the query is from, what they're asking for, and whether it's acceptable to allow that access to go through.

MR. MORTENSEN: Actually, Mike, what you say there is very interesting. You took me back to one of my first cases as an attorney right out of law school. The case was called Higg-a-Rella versus the County of Essex, and in which case Higg-a-Rella was a company that assisted property owners in tax assessment battles. Basically, what they were trying to do -- this is in the early '90s -- they were trying to collect the tax-roll tapes for the properties from the counties, and they were going around, and the huge battle that the county said is, they didn't want to provide it to them in electronic form. And what it came down to was an argument of form over substance.

I think a lot of what we deal with, when we talk about knowledge discovery, when we talk about data mining, when we talk about the use of commercial data by the Federal Government, it is all a form-over- substance argument, because many times what happens is that people say, "I have less privacy because information is electronic." And, certainly, that has a lot of truth to it, but I think what Mike just said is that we have to view that in the context of its use, and to be able to define the business rules. Because, in the Higg-a-Rella case, we ended up winning because -- when I wrote my amicus brief to the Supreme Court, our argument was, "Well, wait a minute, I can walk into the county courthouse in

Essex, and I can get those tax rolls, I can photocopy them, and that's all the same information that I could get if it's in electronic form."

The only argument you can make against that is the fact that, "Well, you can find the information quicker." And what I'd like to do is, I'm going to turn to Rebecca and -- actually, and ask about that. How are we resolving that issue, in terms of a perception of privacy?

MS. WRIGHT: Right. So, there are a lot of things here that I want to speak to a little bit.

I think this is -- this is the notion of expectations, right here, when you talk about the difference -- the expectation most people have is that not a lot of people are going to go to that county clerk's office and stand in that line and get that information. And, you know, certainly, I think, in the last year, the public has become much more aware of the huge data aggregators and the fact that, actually, those aggregators are paying people to go stand in those lines, get that information, and then take it, in a digital form, combine it with other information, and, you know, have dossiers of far more information about each individual than most people realize.

And so, I think, you know, while legally it's true, the information is there, one way or the other, as far as how it affects people's perceptions and how it affects the ability to use that information to extract things people felt was private by virtue of being less accessible, it really changes the nature of that. And that -- I don't think we have all the answers there, but we're starting -- So, what I -- what I want to come back to, a little bit, is this notion of a balance, and -- Is there a balance? Is there not a balance? What is the balance between?

So, encryption has been around for a long time in the form -- and we know how very well to do -- if I want to send data from me to John, that John and I can both read, but no one else in between can, we know how to do that. We need a little bit of help from some identity management and policies and all kinds of other things, but we have a very good idea of how to do that. What we want to do now is use information in such a way that -- you know, perhaps John has some information, I have some information -- we want to learn something about what -- combined in our information, and we don't want to take the route -- we don't want to balance -- you know, tip the balance completely and say, "Security is the utmost concern, and we're going to put all that data in one place that we can look at it," you know, that kind of thing is proposed sometimes and usually rallied against.

People in our country have a very strong reaction that that's not the kind of place they want to live, that's not the kind of world they want to live. They may, for short periods of time and for certain purposes, feel that right now we have to, you know, do

something a little bit differently, but we don't want to create a society, long term, in which that's happening. And so, the kind of technology that John was talking about precisely does that. He has a list of people, I have a list of people, we want to learn who's on both lists without having to share our lists.

And so, that kind of tool that John talked about, and using anonymity, is a way that you can -- you can get some of the results of sharing information without exactly sharing it. And this is what let's you say it doesn't have to be a balance. And where I come from is, looking at -- so, we all know about this very simple use of encryption. The -- what John was talking about maybe some of you didn't know was possible, that you can use, in this case, one-way hashing to effectively create data that's encrypted in a way that you can never go back, but you can still perform certain computations on it. And, in fact, in the scientific research community, as early as the 1980s, there is a body of work, just beautiful and elegant solutions that say, essentially, you can compute anything you want on encrypted data.

We can come up with ways of encrypting our data so that it -- none of us can decrypt it, sending messages around with our encrypted data, and -- you know, it's complicated, but it can be done -- and, at the end, if what we're supposed to do is that, you know, one of us is supposed to learn some common elements on a list, maybe another is a -- supposed to learn some kind of -- you know, how many people fit a certain profile, there are solutions that can do that, and they've been around since the 1980s. What keeps them from having come into widespread use is that they are -- they're not really practical, and they are also very confusing. You know, even those of us with Ph.D.s in computer science have to work for many years to, kind of, get it in our heads and work with it, and then we can't talk to anyone else about them anymore. So, there's -- you know, they're complicated; they're perhaps not efficient enough.

And so, what one of the major goals of this PORTIA Project is -- and, in fact, actually, there are other researchers in the world that care about these kinds of things, as well -- is to take these kinds of ideas and these kinds of techniques and start to bring them into practices. You know, I think John's -- and he wants to talk a little bit in a minute -- is, you know, a great first step. It lets us do so much more encrypting -- on computing with anonymized data, in his case -- I like to think of it is as encrypted data, because it's not necessarily just names you want to remove -- but, you know, allowing certain computations to happen without putting everything there.

Now, that said, coming back to the balance issue -- so, what this says is, it's not, kind of, a direct tradeoff. You can get more security and more privacy than most people think. I still think there is a balance, and it's really more for the public-policy end and the philosophers even, perhaps, to decide if -- you know, so I'm saying we can get -- we can break the balance between privacy and security concerns by computing unencrypted data,

but there's still something you're computing. You're still trying to answer a question. You know, so maybe you have -- you can compute on a bunch of data to get: Who are the bad people on this list? Or maybe you can get: Who are the common people? Or maybe you can get: What's a, you know, base net representation of some of the variables on the list? But someone can see the answer to that question.

And so, that's where we need a better understanding of policies of who should see that, what the privacy implications are, what we think, as a society, should be allowable and should not be.

MR. BLISS: Just a couple of comments.

One to Ken. On the tax-record issue and the notion of, "It's -- well, it's difficult to physically go down and get it," versus electronic, it goes to my aggregation point and the right to privacy in not having data that's actually aggregated.

And I had started thinking about this a bit more, and I was, like, you know, I'm curious just to see if the Framers thought about this. And, lo and behold, in the Federalist Papers, there is a passage involving the discussion about the proper dispersal of power, vis-à-vis the State against the interests of the individual. How do we protect ourselves against too much power in the State? And they were worried about this very notion. And what they designed, or intended to design, was a system in which bureaucracies would be bureaucracies. They would inept. They would be inefficient. They wouldn't well share information.

And here we are now at a point where we actually are at the cusp of being able to be efficient and to share information. So, you have to wonder what the Framers would view about the current state of affairs.

Now, there's a sort of a retort within that passage where they said, "Well, but we're willing to allow that power equation to twist a bit where national security interests are involved." And here we are again talking about that very notion, which is, "Well, can we aggregate data about individual citizens?" Maybe the Federalist -- the Framers would say, "Yes, but in the context of identifying terrorists, but not in the context of catching deadbeat dads."

Now, to the point of -- that was made about anonymization, you know, it strikes me as, we're almost, sort of, talking about a Zen cone. We're talking about sharing without sharing. And a perfect example, an easy-to-understand example is in the mergers and acquisitions standpoint. Let's say that I -- I'm the CEO of a corporation, and my colleague Steve Adler is the CEO of another corporation, and the principal capital asset is a customer list. Now, we have a pretty good idea of what -- he has a pretty good idea of what my customer list is. I have a pretty good idea of his. But it's not exact. Therefore, the valuation of this deal is inexact.

If we anonymized our customer lists and shared them, we would get great precision on the degree of that match. It would allow us to know what value his customer list has to me, and vice versa. At that point, we could agree to agree and merge, or we could agree to disagree. No harm, no foul. We've shared without sharing. It's just an example of how it could work.

MR. GRAVELL: Let me throw out a quick one before the point is lost. I would like to add my support to the thesis that has been general on the panel about the absolute inescapability of seeing the rule set, the policy structure, in the context of technology.

The phrase I tend to use in this environment is that I believe that this whole series of discussions, whether one is speaking of the collection application of data, the sharing of data between organizations and government, the application of governmental data, commercial data by government, or identity management, what these have in common is that they lie at the intersection -- indeed, at the conjunction -- of hard science and social science. I have -- I'm, by the way, a code breaker, by nature, and I have sat in rooms with amazingly talented people of a highly structured, precise, narrowly channelized technical understanding who have sought to gain insight of our nation's enemies via strictly technical means. And I have sat with some amazingly creative philosophers who have sought to do the same thing. And if one does these things separately, one gains some understanding.

But the true power, and, I will suggest, the true successes, only occur when the best of both are combined. That's the model here. I would add a second point or two. We've spoken about the relationship between security and privacy. We've generally agreed that it's not a zero- sum relationship, that one need not be sacrificed in the name of the other. I vote for that, but I add a third, and that is what I call "functionality." You may apply a different term, but what it is, is the recognition that in the scope -- demographic, topical, informational environment -- will overcome all previous models of scalability, affordability, manage- -- maintainability.

All the classic models of X dollars spent for Y heads under cover must necessarily be thrown out the window, because nothing can be done affordably, nothing could move forward effectively in that environment. Hence, we return to disruptive technology in order to restore efficiency to the total system, in order to permit American ingenuity to find expression and not be channelized and forced into a process that renders America essentially, in a small way, a controlled economy, such as the ones that we have encountered and defeated in the global market for the last century.

That functionality factor has to come into and inform both the creation of the technology and the formulation of the rule set.

MR. DACONTA: I'd like to segue off that, relating to the rule set.

We have to be -- and this is going to be in the context of information-sharing -- because, of course, what'll happen, you know, the pendulum may be slightly swinging away from -- you know, as 9/11 fades, swings -- you know, the pendulum will swing away from the need to share and more back to the need to know. So, we have that need-to-know/need- to-share continuum. And after 9/11, of course, we stressed the idea that we need to do a lot more improving and holding the standard that the standard is no longer whether somebody has the need to know, but whether you have the need to share.

And one of the things we've got to -- I want to make sure we get our priorities straight, because there's two ways to look at this. Some of the things that we were talking about, like encryption and de-identifying data are, sort of, all about producer controls. How do I, as the producer of information, hide certain aspects of it and only let you see what I think you need to see? And that quickly runs up against the problem of context. As John mentioned, in some contexts it's perfectly valid and reasonably to say, "Here, have my data." As an example, for me, going through a screening checkpoint, "What do you want to know about me? I'll let you know anything you need to know, because I want to get through this checkpoint quickly, and I have nothing to hide."

So, the question comes down to -- and this is very important -- if you look at some of the doctrine that the military is professing, they basically know for a fact that need to know and producer controls on information costs lives, because it's too slow. They need dynamic, real-time situational awareness. That means they need to move that pendulum much more towards need-to-share.

So, what I would argue is that, instead of thinking about producer controls -- and they're fine, and they're important -- but we need to think about consumer controls. And what does that mean? That means that if somebody's going to express a need, there are techniques available for you to formally express that need, formally model context. We can do that model. Now, it does -- it does cost, because it take cost to model things accurately. But we have technology today where we can formally model need and context. And we have one other good point about this. And this is in the area of things like the Semantic Web.

And, actually, the Semantic Web is all about knowledge representation, which brings us right back to philosophy. And, as an aside, being a graduate of NYU, I'm sure Rebecca and -- they're doing great things, and they should be fully supported. But I'm not biased.

[Laughter.]

Anyway -- back, for a minute, to the Semantic Web. So, in this idea of the ability to formally model context, that then allows us to, with certainty, say, "This request is valid, because I understand its context; and this request is not."

And the best thing about this is, this is not just the government pushing this. The same techniques and the same technologies that will enable that are the same technologies that are required for software agents to act on our behalf. If you look at the Semantic Web, it's all about: How do I take the World Wide Web and take it from a human-readable web to a machine-readable web? So that machines -- so, as an example, if I say to my software agent, "Go -- I'm having these kinds of symptoms. Go find me a doctor in my area that can treat these illnesses, and schedule an appointment."

There are all sorts of data modeling and context implications in that, but there is a huge amount of business benefit in being able to make that happen. So, I just want to make sure we get our priorities right and first focus on consumer-centric controls and then implement producer-centric controls.

MR. MORTENSEN: Actually, I think that's very interesting, because the business rules -- the model that you develop is going to be the core to doing that. But there's only so much technology can do with regard to that. And, actually, I want to throw a question to John. We had a conversation, as part of this, and I liked what you said about the Federalist Papers and how the inefficiency of government actually creates privacy.

And perhaps what we're talking about here, especially what Mike was just talking about with Metadata and building these rules, is that the inefficiency is linked to the need to know, and, because it's inefficient, you didn't get to know everything; but now, because we have these efficiencies, you could have a full transfer of information. And what we need to do is to limit that transfer of information. But, even still, technology can be circumvented.

We were having a discussion about tracking the transfer and the next step to me is, how do we make sure that it's working?

MR. BLISS: So, along this, sort of, spectrum of data that I was describing -- so, let's -- we -- we're now at the point where we have a match.

We've anonymized everything, there's a match. Who gets to know that? Who gets to know that there is a match between a record of Disney Cruise Lines and a terrorist watch list? That ultimately becomes a very critical question for which business-process rules must be designed. And then they've got to be enforced.

So, let's assume we've figured that out. I mean, it's going to be a particular agent within the Federal Bureau of Investigation. How do we know he truly kept it to himself? How do we know if he used it for the use to which it was intended and not to help somebody else out with a deadbeat dad problem? That -- therein lies another problem. And there's been some work done in that area on this notion of an immutable audit. Some work done at R&D labs and IBM, some work done by Doug Patagar, out at Berkeley, and elsewhere. The notion is that you -- it's sort of like an audit on steroids. It

makes sure that even the DBA couldn't have been suborned by somebody, couldn't be an insider who's been extorted by the bad guy. It leaves footprints in the sand that are irreversible, that can't be covered up.

Now, that, in turn, creates a problem, if you think about it, because what have we done? We now have resident in one place some very, very important information. So, if someone could hack that audit, they've done as good a job as if they had hacked the intel first instance. So, a lot of work needs to be done in figuring out how to harden that immutable audit against that kind of theft.

MR. GRAVELL: And yet I'm sure you'd agree, John, that, picking up on that very point, you can -- you can mitigate against that risk by doing things like anonymizing the root data, picking up on your earlier point.

If you recall, a few years ago, a case in which employees of the -- I believe it was Social Security Administration, or it might have been the IRS, were cruising, browsing through files, looking up the names of celebrities, which were well known, and just seeing President of the United States? Probably a few, but not hundreds -- and just wondering what the President spent on -- paid on his income taxes in the recent past.

Well, that was a scandal, and they did a variety of things, and threw some people out. But if the limited number of persons who had necessary professional access to the file structure of the IRS did not have an ability to find any given individual, whether it's a jilted former spouse or whoever, for whatever reason, they just couldn't find a person based on a search for them, except through a defined, technological, and rule-based process that was highly audited and associated with rules and privileges assigned to them in their work, that is another component of your total security.

I would stitch together two or three different technological kernels that have been thrown out here, all of which I think we would agree -- John and we would be the first to agree -- collectively comprise the kind of technology and rule regime we're all talking about in a distributed way. Anonymizing of data, separation of root identity from its applications, compartmentalization of the data applications in controlled ways, auditing of access, digital signature, which provides, for free, features of data-content integrity and, done properly, confidentiality of data, as well. All of these things collectively comprise the kind of regime that is, frankly, readily available in technology today, it just has to be architected.

And I believe that, at least anecdotally, we're moving toward that sort of a regime.

MR. BLISS: And just two other parts of the puzzle: a robust press, an inquiring press, as we've seen in full evidence with Secure Flight and CAPPS II and Total Information Awareness; and real, not apolitical, congressional oversight -- not oversight

for political gain, but oversight to actually get to the bottom of the systems and see if they're working properly.

MR. GRAVELL: Completely agree.

MR. DACONTA: Let me throw one idea in this related to -- some of you -- I think these are all absolutely great points, and I -- and I really like the idea of immutable audits. I think your point on, you know, having access to the file system. And, of course, immutable audit goes beyond access to the file system, but the operating system at -- actually letting that audit trail be unalterable, even with deep knowledge of the system, deep knowledge of the operating system, et cetera.

Let's combine that, and let's add the notion of standard -- obviously, I'm a big standardization kind of guy -- let's add the notion of standardization and formal modeling off that audit log for this. What about -- think about the comfort that an American citizen would get if they could do an automated FOIA request on themselves.

I want the ability to do an automated FOIA request on me. And I'll do an experiment here to prove my point. And you have to be honest. How many people have done a vanity search of themselves on Google?

[Laughter.]

All right? There you go. Now, wouldn't it be nice to be able to do an automated FOIA request that says, "Oh, by the way, I want -- I want the ability, in an automated fashion -- and, in fact, I want it to come back as fast as Google results come back -- I want it to be able to mine all those immutable logs and know precisely what you're doing with my data." And if we could achieve an automated FOIA request, we'd really have something special.

MR. BLISS: I mean, now you're really getting to David Brins' notion of transparent society, that, you know, let's -- if we have transparency, if we know what the government is doing with our data, there might be greater trust engendered in this process. And it also, sort of, is a sop to the notion that, you know, "There is no more privacy. Let's get over it."

MR. MORTENSEN: I think this actually goes to a point that was made yesterday during the Privacy Act discussion where we were discussing the semantic of the difference between the word "retrieve" versus "retrievable."

When we first were looking at this, the Privacy Act, and looking at the issue of the government using vast amounts of data, the inefficiencies that existed then led us to say, "Only if you could retrieve by identity." But now we understand there's this idea of being able to cross-reference the information. And, really, what we're talking about is "retrievable." And, actually, Rebecca, I want to turn to you and talk a little -- have you talk

a little bit about that context and how can we apply the technology in this new age of being able to cross-reference everything? Because I -- if I want to do it, what Michael has just talked about, we're talking about precisely that. And I think that links to those sorts of things.

MS. WRIGHT: Right. Yeah. I mean, I think there's both privacy benefits and security benefits that can be gained here, and there is concerns regarding that.

So, I said before that I think sometimes anonymization is not actually enough to protect privacy and protect identity. And I think one of the, maybe, sort of, mathematically geeky ways I like to say this is, all of my data is an error- correcting code for my name or my identity. But you don't necessarily need to know my name to find it out. You don't need to know my address to find it out. You don't need to know -- if you're looking at a detailed history of every place that I've been in the last year, with my name not attached, that may be, alone, be enough for you to figure that it's me, especially if you know me well, especially if you have other sources of information about -- public information about the fact that I was here in Washington, D.C., today.

And, similarly, with Internet records. And then you get -- you know, or telephone records -- you get so much detail. Now, of course, that's part of the point. That's why that information is so useful, from a security standpoint, is that it can help -- even without knowing exactly who are the bad guys that we'd like to catch, sometimes other information about them can be useful. But you need to be a little bit careful there, as well. And I think one of the points that I think is critical is that technology can only go so far. And it can go farther than maybe we thought it could, or that it could a few years ago, even.

But it's -- you know, it's critical to remember that sometimes the data going in will be bad, sometimes there will be attacks on even the most well-designed systems. And so, it's very important how you use the answers to these various kinds of queries, not just who's getting them and that it's logged, but that -- you know, it would be a mistake to design a passenger screening system that just plugged some data into a computer program and got an answer, bad guy or good guy, and then -- you know, if it said it was a good guy, let it on the plane, no further search, "No metal detector for you, sir. You're a good guy, please step ahead." And, similarly, it would be very bad to go the other way and say, "You're a bad guy," you know, just, you know, arrest you, detain you, whatever, without any kind of due process.

And we have a lot of that process set up in a legal framework. It's also important to encode it somehow formally in the policies and to carry that through into the training and the culture around how these systems are used. I've forgotten, a little bit, where I was directed to go with the question. I just --

MR. MORTENSEN: Go on --

MR. DACONTA: I just want to jump in, one minute, on -- because I do think -- you know, we've touched on it a few times, but I think we need to slay this dragon. The idea of legality of aggregation.

You know, aggregation, of course, brings up a lot of this, right? Because people see, with automated methods, "Wow, it's really easy for them to aggregate." I had my dad actually tell me about some Websites a few weeks ago and saying, "You'd better go check this out, because I can't believe what they know about us."

So, the question we have to ask, you know, is aggregation bad? And, again, it depends on the context. I'd say, no, aggregation is not bad. In fact, we take advantage of it. Who would say that, in fact, "Oh, it's absolutely great and good for people to aggregate prices for when I want to buy something, because I want to do price comparisons and I want to get the cheapest price on whatever I'm buying. So, sure, go ahead, screen -- screen every single vendor on the planet and then tell me, at that precise second, what is the cheapest price on the book that I want to buy, or any other product?"

Now, of course, from those vendors' perspectives that's a bad thing. They may even claim, "That's an invasion of my privacy." But, from the consumer's perspective, guess what? We don't care. You know? So, think about that when you think about the notion of aggregation. All of a sudden, you don't care about the vendor's privacy, because that aggregation is benefiting you, but -- so, we want to make sure we're not hypocritical here.

MR. BLISS: Let me just play the devil's advocate.

When you're aggregating data for commercial purposes, the worst thing that can happen is: you get a piece of mail that you weren't supposed to get. When you're aggregating data for purposes of homeland security, your liberty is at issue. You're detained. You could be arrested. You could be shot. It's a wholly different kettle of fish requiring a whole different level of precision on the notion of what can data mining do to produce a result in that construct.

MR. MORTENSEN: Well, let me just play off of that, then, and pass this back to Rebecca to, kind of, answer. In the context of our constitutional framework for this, aggregation existed. It was one of the things that was fought against. If you look at the Fourth Amendment and the requirement of probable cause, that was to prevent the aggregation of charges, if you will. They were just basically saying, "You're lumped in with everybody else, so, therefore, we're going to throw you in." We need to have a specific purpose and know that we have some reasonableness in going and invading your privacy. How do we deal with that with the technology?

MS. WRIGHT: Right. Right. I think, in particular, my view on this is that there is this real difference when you're dealing with government use and commercial use. I mean, both can actually inconvenience you pretty substantially. But I think that, in particular, our laws in the country are designed to say that citizens have certain protections from governments and government use of data.

And this is where I think the use of commercial data for government purposes can really cross the line. I mean, if you have the government outsourcing its work to the commercial data-holders who do everything that the government wanted to do but wasn't allowed to, that, again, is when you get a justifiable complaint from citizens who say, "This doesn't meet my expectations, and it doesn't meet what I think our country should do."

And so, I think the immutable audits and transparency through both technical means and policy means is really critical here. When the -- when we have government use of data, it -- just like with wiretaps and all the law there, at some time, the fact that a particular telephone line is being wiretapped is supposed to be private. It needs to be -- you know, that needs to be secret, or the things that are -- that the government is trying to hear be said won't be said, necessarily. But, later, that fact needs to be in public.

And -- or, you know, for wiretap law, there would -- there's -- there is a lot of law around that, and policies for doing that, and we're really -- you know, I think we need -- we need new policies and oversight on how government use of data is going to happen, and we need both to create technologies that can support those policies and allow creation of more sensitive policies, and we need to make sure that, through meetings like this one, you know, that what technology can do is known so that the policies can be defined in a way that don't say it's an either/or kind of a balance, but that you can really -- you know, maybe you can have some kind -- some levels of transparency through things like Semantic Web that let you say very specifically what kinds of aggregation can happen, and who needs to know that it happened, and who needs to know when the results were used for what kinds of things.

And I think that's the kind of thing that computers can do better than people, in some sense. Once you automate certain policies, the right alerts or notifications to the right parties can happen at the right time.

MR. BLISS: I'm, sort of, returning full circle to where I began today, which is, I think the challenge that we really are talking about here is how to confer on the government, in a digital world, no more power than is currently conferred upon them in the physical space by the Fourth Amendment. Grant them no greater rights than they have in the digital world than the Fourth Amendment would confer upon them to conduct a search and seizure in the physical realm.

And what is critical is to get policymakers working with technologists to come to agreement on just where that line is. And that's, sort of, what this whole workshop has been about, in some senses.

MR. GRAVELL: I think that's such a good capstone for all that we've said that I would vote for that as the last comment of the panel, subject, of course, to our moderator.

[Laughter.]

But I will add one supplemental comment that I was going to make.

[Laughter.]

So, think of this as the penultimate comment just sneaking in under the wire before John's excellent summary.

I put it to you as a thesis that, in discussing the exchange of information and its application between grossly governmental and grossly nongovernmental environments, the tendency is, in every case, to try and cross that domain, for the simple reason that government, in the course of seeking, for legitimate reasons, to understand the backstory, the richness of action, conduct, motive, intent, and so on, to the extent that those things apply to legitimate governmental concerns, must seek to understand what any given individual, any of us in this room, does outside of a governmental context, because almost all of us have a much deeper, richer, wider story outside government than we do inside. On the other hand, in the commercial world, invariably, what are we told to do?

To show a governmental attestation of our identity, of our integrity, of our fidelity, because it's understood that the business case for Visa International doesn't support the kind of effort that the government routinely expends, even -- well, except for the most trivial cases, such as walk-up DMVs, which I believe are going to be going away shortly. Any -- many of -- but by no means all -- of the government's so-called identification processes are at least marginally more rigorous; thus, governmental identification is slightly better trusted, although I'd suggest that's a Potemkin village.

The real requirement here that will fundamentally change the social and technological order in this case will be the emergence -- and I believe that it is inevitable -- of rigorous commercially-based root identification to facilitate and enhance a set of applications in our private lives, including by those people who, for one reason or another, try assiduously to avoid their engagement with government, try to minimize government's visibility into, or influence over, their lives. These people will see the personal selfish benefits, in purely commercial context, of strong identification.

And, once that emerges as a technology policy opportunity, I think we will see the entire complex in this area redirected toward regimes that will be able to achieve many of the goals we describe here today.

MR. MORTENSEN: Well, I'll finish on a little bit different note, but it's, kind of, taking in context what John and Bill have just said. I like to think back to when I'm teaching my law students. And one of the concepts -- coming from an engineering background mixed with the law, there's a desire, many times, to say that technology will change the law. Because what we're talking about, ultimately, is our jurisprudence here in the United States, specifically in the context of what the government does within that. But the reality is, the law doesn't change.

Rather, we continuously -- and, certainly, going back even to the times of the initiation of our legal process in Great Britain, technology has continued to move forward, and so has the law. It, perhaps, is a little bit slower than technology. And we need to, of course, guide it along the right path.

But I think about very interesting argument out there that has been ongoing for a number of years between Judge Easterbrook and Larry Lessig. Judge Easterbrook wrote a very interesting article, called "The Law of the Horse," in which he argued that law must be viewed in the context of law. And one would have not, in the 19th century, gone and taken a course in law school that specifically dealt with the law of having and owning and using a horse; but, rather, you would look at it in the context of contracts, within the context of torts, within the context of our legislative system. Yet, while I believe that that is entirely true for what we're talking about here, Larry Lessig's argument saying that the technology, certainly in cyberspace, has created this new realm, there are some new aspects of it, and these aspects are what we've just talked about, in terms of the aggregation of information, the ability to know more than we used to know because of the inefficiencies going away, I think have an impact on that.

But I do believe that what we've talked about here says that technologies exist that, while they're not perfect, while they don't solve everything, given good minds thinking together, we probably can come up with the solutions that will let us be able to complete the mission.

And I still go back to what I said, which is that homeland security and privacy are not mutually exclusive. They can exist completely together as long as we recognize the reasonableness of our argument in dealing with privacy.

At this time, I'd like to open up the floor to questions. If you have a question, there is a microphone down here on my left, to your right. I would ask that, for the transcript purposes, if you could please state your name and state your question, and, hopefully, we'll give you an answer.

MS. ACKERMAN: I'm Linda Ackerman, with Privacy Activism.

I asked a question yesterday afternoon of the technology panel about building privacy into systems while the systems are being put together, rather than trying to

retrofit after the fact. And one of the answers I got was that if you build a system around a specific privacy policy, and then that policy changes, you have an obsolete system. So, I'd like to hear the response from members of this panel to that conclusion.

MR. GRAVELL: I'll throw out a portion, but only one portion, of the answer. I think that contributory to that goal, which I agree with, I would say that when the question is adaptation of privacy laws, the emergence of perceptibly new cyber-attack threats or capabilities, the discovery of holes and weaknesses in the system that are architectural defects that need to be resolved -- all of these needs to change the structure and management and function of the system after the fact, based on whatever purpose, go to one issue, and that is the question of the fraction of the system that has been physically distributed versus the portion of the system which is under central control of those persons that are capable of, and responsible to, maintain and manage it.

I think that's one of the fundamental design and conceptual trades in this whole debate, and it goes to data architectures and its intention to the question of data aggregation, because it says, ideally, if a brilliant and passionately patriotic freedom- and privacy-centric computer scientist could control the security and privacy architecture from this imaginary central console, which policies would instantly radiate throughout the vast architecture and instantly provide the benefit that is needed, that would clearly be the best way to go. It would be instantaneous, it would be efficient, it would be consistent.

But, unless one thinks in terms of the implications of giving that person, if poorly motivated, improper access to too much data, well, then we're back to what we spent the last hour and a half talking about. The good news is that there are technology remedies that we've spent a lot of time talking about. So, I vote for the notion of centralized control, centralized architectural management ability, minimal physical deployed components, because those are the most expensive, inaccessible, and difficult-to- operate parts of any such system.

MR. DACONTA: I'd just like to add to that. I don't know who answered that yesterday, but I totally disagree with their answer.

You certainly can build -- I mean, Engineering 101 -- Software Engineering 101 talks about the modularity of systems. We certainly can build modular systems. I think it more addresses the aspect that Bill was mentioning, in terms of then thinking from, sort of, that producer-centric viewpoint that if I have to tightly control aggregation and access, I don't know if the rules are going to change under me. As we have mentioned here, I think that's the wrong point of view.

So, it's certainly -- we certainly can be -- can build privacy into the systems. And, in fact, let me just go back to a point that John made relating to this idea that the risks for

aggregation, of course, on the government are much more severe than in the commercial world. I absolutely agree with that. And, in fact, I do believe we need to match those higher risks, like somebody getting arrested, somebody getting shot, with the -- really delivering on this idea of greater need for transparency, formal definition and formal modeling of things like probable cause. And that needs to get wrapped in to the law, because the law needs to understand, and the law does need to change, because our scope of reasonable expectation has changed, as has our ability to affect that. You know, we're going to start getting into areas of, you know, videotaping of all public access areas. You're not going to have a reasonable expectation. Now, what is done with that information is something that will be -- is very important. And we need to be able to assure the public that privacy is built in from the get-go. And I do believe it will take some launching.

MS. WRIGHT: I couldn't agree more that you -- of course you want to think about privacy from the beginning.

When you design a system, you want to have mechanisms in it for supporting privacy. It's much harder to add it after the fact. And I think -- and, again, I wasn't here yesterday, either, to hear who answered the question. I agree with them that if you build in a particular privacy policy, your system will be obsolete. I think the point is, you want to -- in fact, that's what you get when you don't think about privacy from the start; you have a built-in obsolete system that can't address privacy. I think what we want is systems to be built with a built-in privacy framework, so they have mechanisms for supporting various kinds of -- classes of privacy policies. And then when you actually run it and instantiate it, you run it with particular privacy policies that make sense for the actual use for the current laws and the current goals.

And, you know, whether it's handled from a centralized or a decentralized point probably depends on the specific system. But I think -- I think it's exactly right. If you build in any particular privacy policy, including not building one in it at all, then your system will be obsolete.

MR. MORTENSEN: And, actually, I'd like to answer, because I was here for that, and I heard you ask that question yesterday. And I disagreed, as well. And my engineering background has said to me that we cannot build anything without using a System Development Life Cycle.

Although I think that name is not -- doesn't disclose, really, what we're talking about, because people read that as "system development." But to anyone who's an engineer, development is ongoing. We never stop development. Even once we've deployed the system, even once a system is in steady state, there's an ongoing review of that.

And in a properly done cycle, you'll have gateways at which you're checking the privacy. And, of course, you're going to have changes that will occur. I mean, I'm sure anyone here who's a technologist will remember times when, earlier in their career -- I used to get what we used to call "engineering change orders" that changed the aspects of the attributes I was dealing with, so -- but I think you can build a system that -- we didn't anticipate this privacy aspect when we built the system a year and a half, two year, five years ago, yet we can incorporate that into the design, because, if you build it, as Rebecca was just saying, to have a framework, then you can incorporate privacy even as you go forward in the future, even as technologies change, even as the law changes.

MR. GRAVELL: And to the extent that these are not, I think we would agree, settled areas of either law or public perception, intuitively there will be more, rather than fewer, such changes needed over time.

MR. MORTENSEN: Next question, please?

MR. SCANNELL: Yeah, hi. My name is Bill Scannell. I'm just a concerned citizen.

I've been listening to this mantra of 9/11 in the beginning, and it feels almost quaint, in the sense that we've seen, in the past couple of weeks, that DHS isn't capable of even delivering food or water in times of crisis. I don't know they'd be possibly capable of responding in the event of a terrorist attack. In addition, we're talking about the exchange of information, databases, and data aggregation, and we've seen, repeatedly over the past couple of years, that DHS, in general, and TSA, in particular, isn't capable of being trusted with this data. Whether we're talking about CAPPS II or Secure Flight, we have seen, time after time again, where DHS has been, at best, economical with the truth or hiding behind legalisms, and worse, as we saw quite recently, outright lying on the part of specific members of the Department, that I don't know how we can possibly even consider trusting DHS with this information. But let's assume, for the sake of argument, that we had competent, intelligence, honest, decent people running the Department of Homeland Security. And I realize that, for some out in the privacy community, that might be a bit of a stretch right now. But let's assume that, for the sake of argument. The first question is, when it comes to identity, we're really talking about -- after all is said and done, we're talking about relying on ID. And, as we -- as we all know, identity theft is a pretty simple thing to do. And aren't we really talking about punishing honest people when we're looking at these systems of exchange of information? Because dishonest people will just make up a fake ID or steal an honest person's identity. And the second question goes back to this -- the fact that DHS simply cannot be trusted to police themselves. And we know that the Privacy Act is pretty toothless. And we know that the -- even the DHS Privacy Office really has no teeth. Can't major firms, like IBM and Northrop Grumman, that spend millions of dollars a year on lobbying, lobby the legislative branch for very

stringent privacy laws, with teeth, with sanctions, so that the products that you come to the government to sell might be more acceptable to the citizenry? Thank you.

MR. MORTENSEN: I think that the simplest way to answer those questions is to say, in terms of certain policy aspects of that, we are trying to focus on the technology with regard to answering those questions, and looking at issues of the identity and trying to understand how best to use technology. The conversation here was open to the fact that we want to look at ways technology can enhance privacy to the best possibility that we can do. Certainly, I don't think anyone on the panel will say that technology is the only solution. There are other things that must be done, and it does require an organizational and, certainly, to a certain extent, a societal understanding of what privacy is.

So, I would like to defer that to more of the policy folks. And, certainly, yesterday's panels probably would have been more appropriate to ask some of those questions to.

Next question, please?

MR. BARNES: Good morning. I'm Anthony Barnes. I'm a presidential management fellow at the Federal Bureau of Investigation.

And let me apologize, first, for not being here yesterday. But, lady and gentlemen, we are looking at some really difficult issues between yesterday and today. And there are two huge obstacles that are in our way. And I don't know the gentleman that spoke earlier, but he did steal some of the thunder out of part of what I wanted to talk about.

John, you spoke of the willingness of the citizenry to give up some of what they have as privacy, if they can see some reward to it, but you also said that the government isn't capable of communicating to them the success of the work that's happening.

I don't know that that is really true. Certainly, we can't tell everything, in terms of national security, and so on. But if we could establish some sort of communication link with the public, we might be able to do that. But when we have such awful distrust of the government, it's tough for the government then to come and to say, "You know what? We did really well, since 9/11, to stop some more tragedy. We did exceptionally well. As a matter of fact, look around. Nothing happened." And that's -- it's tough for the -- for the -- for the citizenry to lean back and then say, "Well, that's good." Because somewhere in the back of their mind, even though there has not -- never been another 9/11 since 9/11, they're thinking something in there that this government is not telling us.

So, if we could, in some form or fashion, establish better links of communication with the citizenry, I'm sure that aspect would be almost thoroughly dealt with. And, in terms of privacy and technology, you could develop all the encryption you want, getting to whatever you need to, work on finding the bad guys in different places, but you can

say to the citizenry, "We're doing something. Obviously, we can't tell you everything." If this trust could be built up, we'd be able to do that.

That's one issue. The second one is, we have an ingrained culture among us in the Federal Government that is going to kill us, in terms of working and really sharing information. I feel, personally, that unless two generations of us, as federal workers, die out or retire, we're never going to get to the point where we have real, productive, solid information-sharing. And, without it, we are way behind in the fight against terrorism. So, while not knocking the movement in privacy -- in technology and in privacy in the government, we have to nail these obstacles down, because we're thinking strategically. And, as you think strategically, you have to look at the obstacles before you. Try to knock these obstacles out, and that would be a tremendous boost to working through technology and privacy.

MR. MORTENSEN: Well, John, I just want to begin answering, and then pass it to you. I totally agree with the questioner's comment about the need for trust. And I think the next panel will discuss something of what needs to be done in terms of outreach. Obviously, we have talked about the aspect of what you do if you -- after you've gotten the information from the citizen. And we're talking about the ways to use technology to enhance that privacy.

But, certainly, there is the gathering of the information to begin with, and there has to be a level of trust with our citizenry, and also with our friends around the globe, about whether or not they should provide that information to us. I think that that is where, many times, people come and talk about the balance of privacy, in terms of whether or not -- the question of whether or not they should provide the information.

And I think a lot of that can be handled through outreach. There are programs within DHS that have had excellent outreach and explain the privacy issues and the question issues. US-VISIT is one that I can think of that has done an excellent job of making sure that they have spoken to their stakeholders and the folks that are involved with providing information, why they need that information, how it is being used, and what they're doing with regard to that information.

I think, going into your second issue, in terms of the culture and information-sharing, it's certainly something that will have to be dealt with, in terms of breaking down the stovepipes that exist organizationally. This is something that the private sector learned that they had to do 20-25 years ago in order to be able to compete, globally. Different situation, but, strategically, I think it's the same, in terms of being able to operate and deal with these issues that are not going to be subject-matter specific, but, rather, go across the breadth of the government -- will require the need to create a culture of sharing and integrate with that -- and this is important -- to integrate with that a culture of privacy.

I do believe very strongly that this should be done in a way that does enhance privacy. And things that -- such as the Federal Enterprise model that Mike is working on, I think, are moves in the correct direction and provide us with a foundation of that. I'll let John provide his response.

MR. BLISS: To the fellow's point, the Patriot Act says "Thou shalt share." The intelligence community shall share with law enforcement. And it shall share across verticals, between state, local, and federal. The executive order of the President said, "Thou shalt share." I don't know what more you can do. I mean, there aren't any penalties for the failure to share yet.

But the technology that I've discussed today might encourage them to move down that road. If you're anonymizing sources and methods, you might be able to address a deconfliction problem. So, when -- if the FBI is running an undercover operation at a particular block, but so is the CIA, that it would be nice if they both knew about their other's operation. They don't want to reveal their sources. If you anonymized that information and there was a match, there might be some process by which that information could be revealed in a way that didn't damage either of its operations. That's just one potential way in which technology can nudge them along the way of sharing. But I don't know how much further you can take it.

MR. DACONTA: I'd like to offer a brief answer. Of course, right now DHS is certainly a lightning rod for criticism, but I would like to assert, especially the people that I work with, there are -- DHS is a bunch of hardworking, competent people trying to balance very difficult things, as well as trying to do the impossible, in terms of protecting from another terrorist attack.

So, I do take slight offense to the idea that we're all incompetent and not working hard, because we are.

As far as the -- again, a tension between information-sharing, stopping more bad guys, and trust of the American people, I do believe that we can do both, but I also do believe, as the gentleman mentioned, that some of this is that we, in the executive branch, operate under the framework and the priorities, including budget and resources, from the legislative branch. I do think Privacy Act with teeth is a great idea. The stuff that I'm doing in the data- reference model -- and, in fact, a lot of it is focused on information-sharing, so everybody knows we're doing it in a very collaborative, very open process. We have a public Wiki site.

So, we are moving -- and we have a mandate implement information-sharing using that model across the Federal Government in a very open and transparent way. So, I am all for people lobbying, whether it's vendors lobbying, private citizens lobbying the legislator to get Privacy Act -- a new Privacy Act with teeth, and I do believe that because

we have the potential for higher risk, as well as the technology changing, something that was written in the 1970s probably needs an update.

MR. GRAVELL: I am one of those that has concrete, specific knowledge that DHS has saved our nation in the last four years from at least a couple of potentially horrific events. The sad truth is that if I were to call for a show of hands in this room, or any representative room, of the people who could define and describe those benefits, it would be a very thin crowd of raised hands.

Therein lies one of our problems. In this area, where we are speaking about an abstraction like privacy, which is differently understood and perceived by every single one of us in this room, there is no gold standard, there is no base case to establish where "the right answer" is, and how many standard deviations off that you are, and when you will have progressed toward the goal outcome. It is something that will have to be sold to every American man and woman, in their head, through a set of visible- to-them processes.

That'll be a challenge to all of us in this -- in the technology, the vendor community and in government -- and we have to think about not only the unknown successes but the visible small successes. I cited a minor case, the requirement to put a social security number on a paper check. Trivial. But it is something which, if we were to adopt technology and policy outcomes which would back away from that particular little under-the-skin form of what we might view as a violation of our privacy, that will be contributory to the goals we have in all of this.

And so, the ground-sweeping is important. If we accept as a general characteristic of the maturation of any technology that the root function is increasingly buffered from the end user by an increasingly sophisticated beneficial user interface, any technology you could name describes -- or performs that way. Again, it'll be challenging to be able to show you and me that your privacy has been enhanced, and yet that is exactly what we must set out to do.

MR. MORTENSEN: Well, I'd like to close out this panel session and thank all of you for participating. I just have a couple of very quick thoughts.

And being very new to government service, and coming to DHS just months ago, I will say that I have found that privacy is something that everyone that I have dealt with does consider. This is not something which is an afterthought, and certainly not something that people have viewed as an obstacle.

Rather, there is a desire and a readiness to accept this as part of what the mission is. Perhaps it's not always as very visible, but, as the technology is not always visible, it is there as an undercurrent. And I think that what we talked about today on this panel

shows that there are ways to bring to the two together to enhance privacy to the best capabilities that technology and our people can bring together.

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

[Applause.]