



1 Attendance:

2 Toby Milgrom Levin, Senior Advisor, Privacy Office, U.S.

3 Department of Homeland Security

4 Hugo Teufel, III, Chief Privacy Officer, U.S.

5 Department of Homeland Security

6 Jay M. Cohen, Under Secretary, Science and Technology

7 Directorate, U.S. Department of Homeland Security

8 David Jensen, Associate Professor of Computer Science,

9 University of Massachusetts

10 Martha Landesberg, Senior Privacy Analyst, Privacy

11 Office, U.S. Department of Homeland Security

12 Fred H. Cate, Distinguished Professor and Director of

13 the Center for Applied Cybersecurity Research, Indiana

14 University

15 Greg Nojeim, Senior Counsel and Director, Project on

16 Freedom, Security and Technology, Center for Democracy and

17 Technology

18 Christopher Slobogin, Milton Underwood Professor of

19 Law, Vanderbilt University Law School

20 Barry Steinhardt, Director, ACLU Program on Technology

21 and Liberty

22 Peter Swire, C. William O'Neill Professor of Law,

23 Moritz College of Law, Ohio State University

1           John Hoyt, Chief, Knowledge Management Tools Branch,  
2 Command, Control, and Interoperability Division, DHS  
3 Science and Technology Directorate

4           Stephen Coggeshall, Chief Technology Officer, ID  
5 Analytics, Inc.

6           Stephen Dennis, Technical Director, Homeland Security  
7 Advanced Research Projects Agency, DHS Science and  
8 Technology Directorate

9           Chris Clifton, Associate Professor of Computer  
10 Science, Purdue University

11           Dr. Anant Jhingran, Vice President and Chief  
12 Technology Officer, IBM Information Management Division

13           Rebecca Wright, Associate Professor of Computer  
14 Science and Deputy Director of DIMACS, Rutgers University

15

16

17

18

19

20

21

22

## 1 P R O C E E D I N G S

2 [Convened at 8:39 a.m.]

3 Ms. Levin: Good morning. The Department of  
4 Homeland Security Privacy Office is pleased to welcome you  
5 to our workshop, "Implementing Privacy Protections in  
6 Government Data Mining." I especially want to thank all of  
7 you who have traveled from far, and I understand a number  
8 of you had delays yesterday as the result of the storm, but  
9 hopefully everyone who wanted to attend has been able to  
10 make it.

11 My name is Toby Levin, I'm Senior Advisor in the  
12 DHS Privacy Office, and I'm co-coordinator with my  
13 colleague, Martha Landesberg, who you'll meet shortly, for  
14 this workshop.

15 Before I introduce our welcoming speakers, I have  
16 just a few housekeeping announcements to make. First, you  
17 should have a packet for the workshop which includes the  
18 agenda and the bios -- we will not be doing biographical  
19 introductions -- as well as copies of some of the key  
20 slides from the presentations that you'll be seeing for  
21 today and tomorrow. We will post a transcript of the  
22 workshop on our workshop website at [www.dhs.gov/privacy](http://www.dhs.gov/privacy),











1 the Science and Technology Directorate so that it could  
2 foster the development of vital technologies for protecting  
3 the nation. Jay deserves tremendous credit for his efforts  
4 to transform the Directorate into an efficient and  
5 responsible organization that makes vital technical  
6 contributions to the DHS mission to protect against and  
7 respond to catastrophic events. The S&T Director provides  
8 technology solutions to help the men and women who face  
9 risk every day on the front lines of Homeland Security to  
10 do their jobs more quickly and safely with greater  
11 accuracy. And with that, I'll stop. Jay Cohen, Under  
12 Secretary of Science and Technology Directorate. Thank you  
13 all very much.

14 [APPLAUSE]

15 PRESENTATION OF JAY COHEN, UNDER SECRETARY OF  
16 SCIENCE AND TECHNOLOGY DIRECTORATE, U.S. DEPARTMENT OF  
17 HOMELAND SECURITY.

18 Mr. Cohen: Well, good morning. And thank you so  
19 much for sharing your most valuable asset with us, and that  
20 is your time and also your thoughts, at this workshop.  
21 It's a real pleasure to work with Hugo and his team. I  
22 don't know anybody who has a tougher job in Homeland



1 did it to eliminate or minimize seams because terrorists  
2 and criminals will always take advantage of seams. And  
3 anything that eliminates or minimizes those is good for  
4 security and bad for those who would do us harm. So that's  
5 my little shtick here; I'm not a Mac person, but I'll do my  
6 best with this computer.

7           So what are the goals in law of the Science and  
8 Technology Directorate? And I can tell you, as Chief of  
9 Naval Research for six years of a three-year tour and the  
10 Office of Naval Research was established in 1946. Half a  
11 page in Title 10, it says there will be an Office of Naval  
12 Research, it'll be led by a Navy Admiral, report to  
13 Secretary of the Navy, and it'll do good research. In  
14 2003, of the 183 pages creating the Department of Homeland  
15 Security, 17 pages describe the S&T Directorate. You know,  
16 a camel was that animal created by committee, so we could  
17 have ended up with a camel. We didn't. It was very, very  
18 thoughtfully done. And so half a page in 1946, 17 pages in  
19 2003; it shows you the impact of word processing on the  
20 legislative process.

21           But to synopsise in the law what are the goals  
22 and what do I follow, number one, is to accelerate the

1 delivery of enhanced technological capabilities to my  
2 customers. Who are my customers? In law, they are the 22  
3 components: TSA, Border Patrol, Coast Guard, Secret  
4 Service; and in law, first responders -- the police, fire,  
5 emergency, medical, bomb disposal -- our heroes. And I had  
6 no appreciation for the scale of our first  
7 responders in America. We have 35,000 fire departments in  
8 America -- 35,000 fire departments, of which 80 percent are  
9 volunteer. When I go and visit them and I say, 'Hi, I'm  
10 from Washington. I'm here to help.' They say, 'Great.  
11 Buy a raffle ticket or a muffin because we need a new  
12 pumper.' I mean, this is America. So it is a federal goal  
13 with a local execution; I can tell you it's a great  
14 challenge.

15           Second is to establish -- in my words -- a lean  
16 and agile government service -- world-class S&T management  
17 team. Ladies and gentlemen, I don't do S&T and my people  
18 don't do S&T; we are a venture capital fund, we are a  
19 mutual fund, we invest in S&T to de-risk it to give  
20 capabilities to our customers. And when I say government  
21 service -- because some political appointees -- people like  
22 me come and go, but the half-life of Science and Technology

1 is such that there must be a continuum, and so that's where  
2 government service is so critically important. And in my  
3 experience in Navy and in Homeland Security, is that  
4 Science and Technology -- unless I do something stupid and  
5 Hugo works very hard to help me from doing  
6 something stupid -- is bi-partisan, non-partisan, and I  
7 believe that that is how it should be.

8           And then finally -- and this is a labor of love  
9 for me -- is to provide the leadership and opportunities  
10 for the next generation of our workforce. This is STEM,  
11 Science, Technology, Engineering and Math. Ladies and  
12 gentlemen, we're in crisis in this country today. In fact,  
13 we're in crisis in most of the western countries. People  
14 in middle school, young people, are turning away from  
15 science and math, and when you ask them why, they tell you  
16 the truth -- it's too hard. They're the Playstation  
17 generation; they want instant gratification. If we don't  
18 turn this around, ladies and gentlemen, in my opinion, in  
19 fifteen or 20 years we will not be a first-world  
20 economy. So that's a little bit of the background.

21           Now, what are the threats that we face? This is  
22 a PowerPoint presentation, we'll leave copies, you can move

1 the boxes around however you want. I view the threats from  
2 terror -- and oh, by the way, DHS is responsible for all  
3 threats. In the  
4 law, it's not just terror threats, it's also natural  
5 disasters, like earthquakes and fire and flooding,  
6 tsunamis, et cetera. But I view the threats as bombs,  
7 borders, bugs, and business -- those are the original four  
8 b's. It turns out I've got six divisions; two of them  
9 didn't have b's originally. I think last spring they saw  
10 the Bee Movie, but the division directors came to me and  
11 they said, 'Hey, we're without b's; we're b-less.' So I  
12 added two b's and that's bodies -- that's human factors,  
13 and buildings, which is infrastructure protection. You  
14 understand bombs, you understand borders, you understand  
15 bugs; what's business? Business is the underlying cyber-  
16 backbone that enables everything we do, and it is a very  
17 new area, and very threatening and scary area, of warfare.

18 So if you look across the bottom left to right,  
19 you see consequence of occurrence low to high, and then  
20 likelihood of occurrence. We're always going to have  
21 physical attacks; that's the reality of the world that we  
22 live in. If you look in nuclear, that's a nuclear device -  
23 - that's a nuclear bomb. The consequence of occurrence of

1 that going off are unimaginable; it's far off the scale to  
2 the right. But today, today a terrorist would have to  
3 either buy or build a bomb, and I would tell you -- you can  
4 disagree -- that I think the probability of that is  
5 somewhat low. Maybe not tomorrow, but today. But the day  
6 after 9/11, ladies and gentlemen, we were delivering death  
7 by 37-cent stamps in the U.S. mail -- anthrax, biological  
8 attack. And so you can see while it may not be as much of  
9 a weapon of mass destruction as nuclear, its occurrence is  
10 more likely. We have seen it, we will see it again.  
11 Biological warfare is the poor man's weapon of mass  
12 destruction. Because today, with the internet, with  
13 genomics, all it takes is a brain, a basement, a  
14 microscope, and you can create a pathogen that will give  
15 you a pandemic.

16 IED's -- they're weapons of mass influence, not  
17 weapons of mass destruction. Tom Friedman said IED's are  
18 coming to a theatre near us, and I believe that.

19 But the tactics, techniques, and procedures that  
20 we use so well overseas, many of them don't apply -- don't  
21 apply in the United States because the Constitution,  
22 because of the Fourth Amendment -- many of the things that

1 you're going to be discussing here. Before a bomb squad  
2 can actively jam a bomb and its trigger device, they have  
3 to get a license from the Federal Communications Commission.  
4 It's a very interesting challenge; not what you're going to  
5 be addressing today.

6 But what you are going to be addressing today is  
7 up in the upper-right, high and to the right, and that's  
8 cyber, because every three seconds someone's losing their  
9 identity. And you have Estonia, and you understand if your  
10 background, the challenges of what a cyber-attack could do.  
11 Those of you who have children or grandchildren in college,  
12 you understand they live from ATM swipe to ATM swipe. And  
13 if we can't do that, in my opinion, there will be panic in  
14 the streets. So you can agree or disagree, but that's sort  
15 how I see life.

16 So Hugo has already talked about the enabling  
17 legislation, I think very well thought out, well debated;  
18 it has been modified, we've had a change in the Congress in  
19 the ensuing years. We get to testify a lot. Everything I  
20 do -- I'd contend 99.9 percent of what I do is  
21 unclassified. We invite the Congress to our processes, we  
22 invite the Inspector General; and Hugo has workshops like

1 this, which I know will be the first of many to come. So  
2 the authorizing legislation for me, I have summarized it,  
3 in the first, telling you what my goals were. I think I --  
4 I'm too fast.

5           So as we look at data and we look at the threats,  
6 and I looked at what is unique in Homeland Security, I  
7 settled really on two things. Because the enabling  
8 legislation is very thoughtful, it tells me not to recreate  
9 the National Institutes of Health and not to recreate the  
10 Center for Disease Control and not to recreate the  
11 Department of Energy or Department of Defense labs -- and I  
12 think that was very thoughtful -- but in exchange, it  
13 allows me to leverage everything they do. I can't tell  
14 them how to invest their billions of dollars in research,  
15 but they give me full disclosure. And it really does work.  
16 And then I take my precious dollars, our precious dollars,  
17 and apply it to the things that are unique in Homeland  
18 Security and the missions that we have.

19           So from my perspective, as I looked around at all  
20 of the areas of Science and Technology, all the different  
21 disciplines, the two that I felt -- and I still feel that  
22 way after two years on the job -- that were unique, was

1 number one, the psychology of terrorism. Why do terrorists  
2 do what they do? I mean, you can view them as criminals,  
3 you can view them as armies, et cetera, but why do they do  
4 what they do? It was not clear to me any other component  
5 of government was investing in that.

6           And the second area is hostile intent, and we're  
7 going to talk a little bit about that. Are there ways of  
8 knowing that someone is about to do something bad to our  
9 society? And so these are focus areas that we are looking  
10 at. This is new science. We've gone to the National  
11 Academies of Science to help us define those sciences. You  
12 know, after World War II, the Battle of the Atlantic,  
13 strategic bombing, the science of operations, research  
14 operations analysis, was born. And after Sputnik  
15 aerospace, you get the idea. As time moves on, challenges  
16 change; new areas, new disciplines develop. But how do we  
17 know that what we think is appropriate research, even  
18 vetted by the Privacy Office, even briefed to the Congress;  
19 and of course, the press is very interested in this, as  
20 they should be. I mean, at the end of the day, ladies and  
21 gentlemen, I am a citizen, I value my privacy, I respect  
22 and value your privacy, and when I'm done with government

1 service, I will again be a citizen. I think I'm a citizen  
2 while I'm still in government service, but you get the  
3 idea.

4           So Dr. Sharla Rausch and her people are  
5 represented here today. She's head of my Human Factors;  
6 this is a division that I set up. There's a great ad by  
7 Dow Chemical, it talks about the human element. I love  
8 that ad because it's the human element that creates  
9 terrorism and it's the human element that will solve the  
10 challenges that we have. It really is all about humans.

11           But Sharla went ahead and worked with the Privacy  
12 Office and others, established on her own, the Community  
13 Perception of Technologies Panel. And so these are just  
14 average people from a wide cross-section -- they have a  
15 picture of them here -- and we go ahead and we brief to  
16 them. This is our initiative, what we're looking at, what  
17 our research areas are, how we're approaching it. They are  
18 not necessarily experts in privacy; we go to Hugo and his  
19 team for that, and I've got Jen Schiller on my staff. And  
20 I can tell you, she is very tough on me. This is an area  
21 where an ounce of prevention is worth pounds if not tons of  
22 cure.

1           And it's very interesting to sit down, and I sit  
2 down with this panel, and get their feedback on their  
3 perception on what we are doing, and then we modify as  
4 appropriate.

5           So let's talk a little bit about the areas of  
6 research that we are doing, and then I'll conclude because  
7 I know Hugo does want to get you back on track into panels  
8 and the discussions are so important. So I'll go through  
9 this very quickly.

10           And I must tell you that personal identifier  
11 information was a new concept to me when I came on board,  
12 and so in the last two years I've had a steep learning  
13 curve. And I also understand that we can be looking, you  
14 know, at totally unclassified, totally public information,  
15 but perception of how that is analyzed, et cetera, becomes  
16 an issue on its own. And I know you're going to address  
17 all those things. The Congress enabled the S&T Directorate  
18 with Centers of Excellence. I have two pillars of basic  
19 research: universities and laboratories. And so at the  
20 University of Maryland, one of our earliest Centers of  
21 Excellence was the Study of Terrorism and Responses to  
22 Terrorism (START). In Washington if you've got a good











1 [APPLAUSE]

2 Ms. Landesberg: Thank you, Under Secretary  
3 Cohen. I'm Martha Landesberg from the Privacy Office, and  
4 it's my pleasure this morning to introduce our next speaker  
5 to you. He is Professor David Jensen who is an Associate  
6 Professor of Computer Science and Director of the Knowledge  
7 Discovery Laboratory at the University of Massachusetts  
8 Amherst. Professor Jensen currently serves on DARPA's  
9 Information, Science, and Technology Group, and he was an  
10 analyst in the Office of Technology Assessment from 1991 to  
11 1995. I give you Professor Jensen.

12 [APPLAUSE]

13 PRESENTATION OF DAVID JENSEN, ASSOCIATE PROFESSOR  
14 OF COMPUTER SCIENCE, UNIVERSITY OF MASSACHUSETTS.

15 Mr. Jensen: Thank you. Thank you very much.  
16 Under Secretary Cohen is a difficult speaker to follow, and  
17 so I hope I can keep this as interesting and relevant to  
18 today's conversations. So what I'm going to talk today  
19 about is at some level somewhat boring in that it is about  
20 definitions. But as many people have said, words mean what  
21 we want them to mean. And I think in this particular case,  
22 data mining means many things to many different people.







1 about knowledge discovery and data mining.

2           Now, I tend to use the term knowledge discovery  
3 because I think it is intrinsically more meaningful and  
4 less easy to mistakenly understand than data mining is. I  
5 think data mining has a clear and obvious meaning which is  
6 wrong; the clear and obvious meaning is that you are mining  
7 for data, and that's not actually what data mining is  
8 doing. If you say gold mining, that means you're mining  
9 for gold. If data mining should be mining for data, you're  
10 not. You're mining for knowledge, and knowledge discovery  
11 gets at that. Although, it did confuse my Dean greatly  
12 when I was introduced to him as doing knowledge discovery,  
13 he looked and he said, "Isn't everyone at a university  
14 doing that?" And I said, "Yes, yes. But we're doing it  
15 with computers." He said, "Oh, well, that's very  
16 interesting," and we went on to have a pretty good  
17 conversation. There are other terms, as well -- predictive  
18 analytics, advanced statistical modeling, machine learning.

19           So, well, I'll stick with the term data mining  
20 even though it's not my preferred term because it is the  
21 term that stuck. So let me give you an example of this  
22 sort of work -- this sort of technology, and it's about



















1 the data, and then out pop predictions out the other side.  
2 So that's what we've got, this kind of filtering process.  
3 Now, this filtering process -- this idea of a filtering  
4 process has been encouraged by some of the most powerful  
5 people on the planet, some of the most powerful image  
6 makers on the planet. Those people reside in Hollywood,  
7 mostly. For those of you who have seen Minority Report,  
8 this is a very persuasive image. This idea that there is a  
9 black box out there that will be producing predictions, and  
10 if the predictions are certain, they are crisp, there is no  
11 doubt in them, and they put them out and that's what then  
12 we go act on as a law enforcement agency. For those of you  
13 who watch television also, there was a short-lived show  
14 called Threat Matrix, which had some similar ideas that  
15 were frequently propounded in the show about data mining.  
16 And as you might expect, these media images are somewhat  
17 simple. They're simple because it's very easy to  
18 misinterpret the definitions which I've given you  
19 previously, which can be interpreted accurately but it's  
20 very easy to misinterpret them. Let me explain some  
21 reasons why. The first is -- and I'll explain more about  
22 each of these in the next set of slides -- the first is

1     that there is only once process. The misperception is,  
2     there's only one process that encompasses what I'll refer  
3     to both as learning and inference. The second is that the  
4     records that come in the left side are disconnected from  
5     each other. Here I'm showing just individual records about  
6     reps. Third, that the inferences out the other side are  
7     deterministic. Essentially we spit out a set of reps that  
8     are bad and a set of reps that are good. Fourth, is that  
9     this is only done once, this single stage, it's a once-through  
10    process. And finally, that this process of data mining is  
11    what I'll call institutionally isolating. That is, it just  
12    sits off by itself in this little box and does its job.

13            Let me explain why each of these I think are not  
14    accurate, and what is a more accurate picture. The first  
15    is that the processes of learning and inference are  
16    distinct. That is, there's not just one process, but  
17    actually two. The learning phase takes in data for which  
18    we know the correct answer, or we have good estimates of  
19    the correct answer, and that puts out a statistical model.  
20    That model is then used in an inference process to take in  
21    data for which we do not know the correct answer, and put  
22    out some kind of prediction.















































































































































































































































































































































































































































































