DEPARTMENT OF HOMELAND SECURITY

- - -

PUBLIC MEETING OF THE

DATA PRIVACY AND INTEGRITY ADVISORY COMMITTEE

- - -

Monday, September 22, 2014

Suite 800

1331 F Street, N.W.

Washington, D.C.

The meeting was convened, pursuant to notice, at 2:01 p.m., LISA S. SOTTO, Chairperson, presiding.

COMMITTEE MEMBERS PRESENT:

LISA S. SOTTO, Chairperson, presiding

JIM ADLER	SHARON A. ANOLIK
CRAIG W. BENNETT	ALLEN BRANDT
ALAN BRODER	JAMES BYRNE
JOSHUA GALPER	MELODI M. GATES
LYNN GOLSTEIN	JOANNA L. GRAMA
LINDA D. KOONTZ	DEBBIE MATTIES
JOANNE McNABB	SARAH MORROW
GREG NOJEIM	JULIE PARK
CHRISTOPHER PIERSON	TRACY ANN PULITO-MOCHALEK
RUSSELL SCHRADER	C.M. TOKE VANDERVOORT
MARJORIE S. WEINBERGER	RICHARD WICHMAN

BARRY STEINHARDT (via telephone)

JEEWON KIM (via telephone)

CHARLES PALMER (via telephone)

ALSO PRESENT:

SHANNON BALLARD, Designated Federal Official

STEVEN RICHARDS, DPIAC Staff

P R O C E E D I N G S

MS. BALLARD: My name is Shannon Ballard. I am the Designated Federal Official for the DHS Data Privacy and Integrity Advisory Committee. I want to welcome you here today. Thank you very much for participating. We are now on the record, so I want you to know that all of the information that we are discussing here today is on line at the URL as originally offered to you and you can find more information on it later on after the meeting is over.

So I'm going to turn the meeting over now to our Chair, Lisa Sotto. Thank you very much.

CHAIRPERSON SOTTO: Thank you very much, Shannon.

I am delighted to welcome everybody to our second committee meeting of fiscal '14. I am particularly delighted to welcome Chief Privacy Officer Karen Neuman, who was ill at our January meeting. So I hope you're feeling well today.

MS. NEUMAN: Much better, thank you.

CHAIRPERSON SOTTO: I'd like to remind everyone, including myself, to silence cellphones, please. For those of you who are viewing remotely, there is a couple of second delay when pages turn, so just be aware of that and please be patient with us.

To help facilitate questions and comments from those on the phone, this is an operator-assisted teleconference, so please follow instructions that are given when you ask a question or make a comment.

We will take questions from the members that are in the

room first, and then we'll turn to the phone for additional questions.

We have a very full agenda today. We're very tight, so please forgive me if I try to -- if I cut you off. We're going to keep to our schedule.

We are going to hear about cybersecurity and Big Data initiatives. Then we will be talking about our draft recommendations from the papers that many of you have worked long and hard on.

We've reserved time, as we always do, for public comment. The public comment period is on the agenda to begin at 4:40 this afternoon. If you're interested in addressing the committee, please sign up at the table in the back of the room. For those on the phone, we will ask you for comments after those in the room have commented. Please keep in mind that our public comment period may begin earlier if we end the session earlier, although I have a sneaking suspicion that we will not end early today.

I'd like to remind you -- we heard this morning, but for those of us around the table -- push the green light, push the button on the back of your microphone, please, and you'll see the green light go on. Then when you want to turn it back off, push it again and you'll see the red light go on.

Also, as we always do, to ask questions please put up your name card as a tent, your tent, put it upward, and I will call on you in, I hope, the order that you put up the tent, but forgive me if I don't look around quickly enough.

All right. Let's begin with our first session. We're really delighted to welcome Karen Neuman with us. Karen is the Chief Privacy Officer of the Department of Homeland Security. She was appointed in October of 2013. In her role as Chief Privacy Officer, Ms. Neuman is responsible for evaluating Department-wide programs, systems, technologies, and rulemakings for potential privacy impacts and for providing mitigation strategies to reduce any privacy impact.

Together with her DHS Privacy Office team, Ms. Neuman is responsible for privacy compliance across DHS and also serves as the Department's chief Freedom of Information Act officer.

Karen, we're eager to hear about your activities, your office's activities since January. Please proceed.

(Screen.)

DHS PRIVACY OFFICE UPDATE

MS. NEUMAN: Thank you, Lisa. It's a pleasure to be here to open my very first DPIAC meeting since I became Chief Privacy Officer, going on almost a year now, and I can tell you being here addressing you today is a lot more comfortable than two weeks into the job. So it's a pleasure to be here.

I notice that we have the agenda up there and in the interest of time it's been slightly modified, but I will hit all the essential food groups when I go through my briefing and overview of what the Privacy Office has accomplished over the course of the last year and certainly since you all met as a body.

But before I do so, I would like to welcome the new members to the DPIAC and express my gratitude to them for stepping

up and participating in the very important work of this important committee and for their patience going through the process to come aboard and join the group.

I'm going to ask the members as I call your name, if you'd please stand and be recognized. First, Sharon Anolik, President of Privacy Panacea from San Francisco, California. Thank you.

Alan Broder, who is a Fellow and Chairman of Novetta Solutions in McLean, Virginia, and an Adjunct Clinical Professor of Computer Science at Yeshiva University in New York. You could stand. Thank you. Thank you very much.

Josh Galper, Chief Privacy Officer and General Counsel of Personal.com in Washington, D.C. Good to see you. Thank you.

Debbie Matties, Vice President for Privacy at CTIA-The Wireless Association in Washington, D.C.

Sharon Morrow, Chief Privacy Officer for the State of South Carolina. Hi, there. Thank you.

C.M. "Toke" Vandervoort, Vice President and Assistant General Counsel for Technology, Privacy, and Security, and Chief Privacy Officer for XO Communications in Herndon, Virginia. Thank you.

All of you -- and I can assure you that I say this emphatically -- all of you are absolutely excellent and outstanding additions to an already very high caliber team and I look forward to working with you and getting your feedback on these increasingly complex and evolving privacy challenges that all of us are faced with in our day to day activities. So I thank

you again.

Just so people know, biographies for all the new members are on our web site and they are also provided for you in the back of the room at the registration table.

Please also note that a Federal Register notice was published today seeking applicants to serve on the DPIAC. So I would encourage everyone here to consider applying and to share it with those who you think would be interested in working on the DPIAC and who would contribute to really the remarkable and excellent work of the committee.

I'm going to preview the meeting a little bit and then I'm going to move into the specific briefing. During the meeting I'm going to -- we as the Privacy Office will provide insight into the cybersecurity activities, and I'm also very pleased to host Assistant Secretary Andy Ozment as he addresses the committee on the work that his office is doing and his team is doing on the Department's cyber activities.

From that, we will also continue our discussion of the Department's use of Big Data. We have expert panelists who will update us on implementation of the DHS Data Framework, including three recently updated Privacy Impact Assessments, to be followed by what I expect will be a very robust discussion of the subcommittees' research on this topic.

Big Data, as you know, has been an important project for the Privacy Office and indeed the entire Department. My office actually spearheaded a briefing about the Framework for John Podesta's lead of the White House study of Big Data, and we

focused particularly on helping educate Mr. Podesta on government use of Big Data, particularly how DHS is using Big Data for a whole host of functions associated with its mission.

My staff also made a significant contribution to a chapter in the Podesta report in the broader context of embedding adequate privacy protections, in fact cutting edge privacy protections using technology, into government use of Big Data.

So I look forward to receiving the committee's final recommendations on privacy best practices for notice, transparency, and oversight in the Department's use of Big Data.

Now I'd like to give a brief overview about the Privacy Office. A number of exciting things have happened since we last met either in the subcommittee or as a group, and there have been a number of other recent developments. Our annual report, which details all of this, will be published soon and will provide you more information about our activities, and I will spare you the functional equivalent of essentially reciting that report to you. It's fantastic reading, but I will just provide a few highlights here.

I think it doesn't come as a surprise, or it shouldn't come as a surprise, that the entire government is operating in a very constrained fiscal environment and the Privacy Office has certainly not been immune to that. So I have initiated a process of reviewing the office, the Privacy Office's strategic plan early in the summer. The objective of this strategic review has been to evaluate and assess our goals and our objectives and whether or not and the extent to which they should be adjusted to reflect the

environment we're currently operating in, as well as our priorities in the coming years.

The purpose of this exercise is really to make sure that we are able to agile as an office in carrying out our mission to help the Department implement its mission, including clarifying what goals are realistic and furthering our role as a true leader in the privacy community while implementing our statutory obligations.

This is no easy task and it's not for the faint of heart. I have been very impressed with and encouraged by the Privacy Office, who has contributed enormously to this effort. We've sought their comments and we've gotten a lot of feedback that we have taken to heart and will continue to do so as we continue moving through this process in the coming weeks.

I expect that we will be in a position to implement a new plan by the first quarter of fiscal year 2015.

Many of you know that my office also conducts a lot of outreach to the privacy advocacy community and others so that we can preview some of our activities and hear what their concerns might be and generally have an ongoing conversation with them about privacy issues associated with the Department's activities. I really believe that this type of outreach is very beneficial. It's interesting and useful for us to hear what their concerns are.

It doesn't mean that we will shift direction, but it is very valuable to hear what their concerns are, and especially to the extent that their feedback demonstrates that we've anticipated

a number of privacy issues and we're moving forward in the direction that all of us think is adequate and appropriate.

In terms of those meetings with the privacy advocates, in January my staff and I met with the advocates to discuss the Big Data Framework and Unmanned Aircraft Systems and the work that my office has done on these systems and making sure that privacy protections are embedded into these systems.

In May we met to discuss facial recognition technologies and the work of the Department currently going on in that context.

Next month we're planning an advocates' meeting to discuss the Department's cybersecurity activities and Andy Ozment again will be participating in that meeting to engage in the discussion with the privacy advocacy community.

My staff and I are also very active in terms of our outreach to the Hill. We have spent some time briefing congressional committee staff on a number of topics since January, including Unmanned Aircraft Systems, the DHS Data Framework, and the general oversight role of the Privacy Office and how we function really as an independent entity within the Department.

On the international front, that work is performed by staff in the policy and advocacy team, and we've had a very busy year on the international front. In Canada, we continue to work on U.S.-Canada Beyond the Border Action Plan implementation, and as part of our implementation of the plan we developed flexible training materials on the Beyond the Border Privacy Principles that our office and the privacy professionals help craft. We've also reviewed all of the Beyond the Border information-sharing

agreements and provided expert guidance on agreements for Beyond the Border projects.

We are also assisting with the continued implementation of phase III of the Beyond the Border Entry-Exit Program, and we are in discussions about the exchange of biometric information between the two countries.

In late June, I traveled to Canada and I met with a variety of government officials, our counterparts, to discuss DHS privacy policy and implementation of the Beyond the Border Privacy Principles. While I was in Ottawa, I met with the Canada Border Services Agency, Public Safety Canada, as well as the new Privacy Commissioner, and I'm really encouraged by the work that we're doing with the Canadians. We have a very robust and good working relationship and they've been a good partner, and on the privacy front we have a lot that we are able to work together on to achieve our common goals.

I also at the end of that meeting, I led a discussion with the Conference Board of Canada with a number of privacy industry experts, academics and government officials, generally on information-sharing, security, and privacy. The conversation did wander off into sort of a broader conversation about privacy in general and how technology is really changing the way we all communicate and what our expectation of privacy is.

That was a really important conversation and I look forward to continuing to engage with those people. In fact, I will be traveling to Canada next week on October 3rd. I will participate in Canada 2020, which will be a conference convened in

Ottawa to address security and privacy issues, and I look forward to doing that.

Our office has been also very active in the U.S.-EU Data Protection and Privacy Dialogue, which has recently been focused on trying to conclude a Data Privacy Protection Agreement. In fact, I literally just got back from Rome from one of our negotiating sessions. The Privacy Office staff continues to support the U.S. interagency talks with the European Commission on this agreement.

The DPPA, just briefly, would serve as a binding umbrella agreement with baseline standards for protecting PII exchanged for law enforcement and public security purposes. You may be aware and you may have read in the press that one of the key discussion points recently has been how the United States can satisfy or meet the EU's requirement that EU citizens enjoy certain rights of judicial redress equally as U.S. persons enjoy under the Privacy Act. As you can imagine, that's been a very vibrant conversation.

My office continues to support engagement with the governments of Australia, Canada, New Zealand, and the United Kingdom under the Five Country Conference, to improve information-sharing in immigration and border security.

The Five Country Conference created a new Privacy Task Force to assess the privacy and information-sharing laws and policies of the member countries and how these policies and practices might impact information-sharing goals.

Privacy Office staff act as the lead for the Department

on the task force and were successful in obtaining agreement among all five countries to promote transparency by proactively sharing with one another their full PIA's for all Five Country Conference projects.

I'd like to spend a little bit of time on national security. Our policy and advocacy team continues its leadership role in the Department in the national information-sharing portfolio.

In May of this year, Ken Hunt, whom many of you may know, began a detail at the National Counterterrorism Center, or NCTC, as DHS's first on-site oversight representative. He's reviewing NCTC's compliance with its information-sharing agreements with DHS and working on an overarching PIA to provide greater transparency about DHS's information-sharing relationship with the NCTC. I've been really delighted to hear how he's been very well integrated over there and he is achieving quite a bit in the short amount of time he's been there.

We also are continuing our participation in the DHS Records Working Group, where we review sensitive national security information-sharing activities or policies, and our office has been very successful at influencing the privacy approach to those activities.

With respect to information-sharing governance, we maintain, as you may well be aware, an active leadership role in DHS's internal information-sharing and management governance processes, which are complex, and we do play a vital role in developing those processes. Through our participation in the DHS

Information-Sharing and Safeguarding Governance Board, or ISSGB, and the DHS Information-Sharing Coordinating Council, or ISCC, we supported the development of the DHS Information-Sharing and Safeguarding Strategy and its implementation plan.

I will pause for a moment to tell you that I noticed somebody on my flight back from Rome watching the movie "Mary Poppins," singing the song "Supercalifragilisticexpialodocious," and I feel as successful going through "ISSGB" and the "ISCC" titles as the people singing that song in that movie.

(Laughter.)

But I don't want to digress from my overview here, and I will just let you know that the implementation plan includes a priority objective for privacy, civil rights, and civil liberties compliance processes to promote enhanced privacy oversight of DHS's information-sharing agreements and is co-led by the Privacy Office and the Office for Civil Rights and Civil Liberties.

As part of the Information-Sharing Coordinating Council, the Privacy Office also participated in the Data Access Request Process Working Group. This working group seeks to memorialize and automate DHS's internal clearance processes for information-sharing agreements and to ensure that the oversight bodies within the Department, the Office of General Counsel, the Privacy Office, and CRCL, are able to review DHS agreements with external entities.

Through these boards and councils, the Privacy Office gains really critical insight into the information-sharing needs of the DHS law enforcement operators and is able to engage in a

truly collaborative dialogue on how to address those needs in a privacy-consistent manner.

You will not be surprised to know that training is also an enormous part of what we do here in the Privacy Office at DHS. It's one of the ways we bake privacy into the activities of the Department. We provide a variety of training to DHS personnel, but I do want to briefly call your attention to some relatively new specialized training that we've developed.

One of which is training for the DHS Office of the Chief Human Capital Officer. So we have trained all of the Headquarters CHCO staff on best practices for safeguarding PII. We also continue the international privacy policy module for new DHS attaches stations at U.S. embassies worldwide to raise awareness about U.S. privacy law and DHS privacy policy and practice. We trained over 125 newly stationed DHS officers this fiscal year.

In June 2014, 196 personnel from 42 federal agencies attended the DHS annual privacy workshop. This workshop is a one-day event and it provides in-depth training on all of our privacy compliance processes and best practices, as well as on other important privacy topics. It is one of the best attended public sector-focused privacy events.

I'd like to take just a little time to talk to you a little bit about the work of our compliance team. As you can imagine, as the Department's activities evolve and expand so too do the compliance team's work. They have been very busy since I've been aboard as well.

Since the January 30th meeting, the DHS Privacy Office

has published 33 PIA's and 11 SORN's. Some of the highlights of this work include: having updated the DHS Data Framework PIA's. As you know, the Data Framework is DHS's Big Data solution to build privacy protections while enabling more controlled, effective, and efficient use of existing DHS-related information across the DHS enterprise and with other U.S. government partners as appropriate.

We'll talk more about this shortly, but after the successful completion of the pilot and the prototype phase DHS intends to mature the Framework by entering into the next phase, which will be limited production capability.

The compliance team at DHS updated the Data Framework, Neptune, and Cerberus PIA's to reflect this transition.

We've also published the CPB Border Surveillance Systems PIA. This was published on August 29th, and it's a combination of surveillance systems, eight FISMA systems to be precise deployed to provide comprehensive situational awareness along the U.S. border to assist CBP in detecting, identifying, apprehending, and removing individuals who illegally enter the United States at and between ports of entry or otherwise violate U.S. law.

CBP owns and operates this system, which includes commercially available technologies such as fixed and mobile video surveillance systems, range finders, thermal imaging devices, radar, ground sensors, and radio frequency sensors.

There's a new system of records titled "E-Authentication Records System of Records," which allows DHS to maintain and retrieve records about individuals, including members of the

public, who electronically authenticate their identities as part of DHS's programs. The information in this system of records includes data collected by programs and applications for use when DHS or a trusted third party performs some or all of the functions that are required to enroll, issue, or maintain a credential on DHS's behalf that can be used by an individual to electronically authenticate his or her identity to the DHS system.

On the Unmanned Aircraft Systems front, there's not a lot to report on our work in this area, but this is a topic that's been of interest to the public and most people seem to follow it, so I'd like to give you just a few updates.

Secretary Johnson approved the Unmanned Aircraft Systems Privacy Civil Rights and Civil Liberties Working Group's draft best practices for the UAS document on September 2, 2014, and it's now ready to submit for OMB review. Our office played an extensive role in the development of those practices.

As reported in the press, the White House is working on a UAS-related executive order that includes privacy elements. Again, my office has been deeply involved in this effort, and we believe the executive order will address many of the issues previously raised by the DPIAC members.

The Government Accountability Office completed a review of DHS's oversight of CBP's use of these systems late in August. We expect that it will be published shortly and that it will be mostly positive.

On the FOIA front, as Lisa mentioned, I am also the Chief FOIA Officer. We continue our push to be proactive in our

transparency, and this year alone DHS posted 3.8 million pages of records, which really eclipses the combined total of the previous five years. I like to think that with greater transparency comes an uptick in FOIA requests, and I certainly don't expect that to change. The total pages proactively posted since 2010 is really staggering. It includes 3.9 million pages.

With respect to reporting and guidance, we continue to receive, DHS as a Department, the largest amount of FOIA requests within the Federal Government and, not surprisingly, with an associated backlog. In March 2014 we reaffirmed the Department's commitment to openness and transparency you issuing a new policy memorandum entitled "Freedom of Information Act and 2014 Sunshine Week," which highlighted some of the Department's accomplishments over the past year in furthering openness and transparency initiatives.

There's been a lot of FOIA training in my office as well or conducted by my office. In July 2014 the Privacy Office trained 61 DHS participants in how to document the FOIA records search and provided an overview of the revised FOIA search form.

The training was tailored to those who are responsible for gathering records in response to FOIA requests and for FOIA processors.

I want to focus now a little bit on privacy oversight. As you know, we do perform privacy compliance reviews of a lot of our systems and programs. In August we completed an update on the EINSTEIN Privacy Compliance Review report, and we conducted PCR's in a collaborative process designed to assess highly privacy-

sensitive programs' compliance with existing PIA's and SORN's, and to work together with the components toward improvements where necessary.

EINSTEIN capabilities are the subject of various oversight bodies, and in addition to our review the program recently underwent a thorough review by the DHS Office of the Inspector General.

The final August PCR report found that all of the findings of the 2012 review were compliant and we didn't issue any new recommendations.

We are now in the process of conducting a PCR on the Enhanced Cybersecurity Services Program, as we noted in our assessment of the Cybersecurity Executive Order. We anticipate the results of this review will inform our upcoming assessment under the executive order.

I think what I will do at this point is just share a few parting words. As many of you may recall, the Privacy Office recently celebrated really remarkable achievements during its first ten years at a "Decade of Excellence" event that I hosted. DHS Deputy Secretary Alejandro Mayorkas enthusiastically recognized the work of the Privacy Office and emphatically underscored how our office has created a "stronger and more effective Department."

As I reflect on these achievements, as well as the accomplishments of the Privacy Office since the last meeting, it is exceedingly clear to me that my office continues to work hard to influence the way the Department responds to a complex range of

threats in a manner that is privacy friendly, and the office is widely recognized for this very challenging and important work.

Today's diverse topics, Big Data and cybersecurity, are just a sampling of the evolving and complex privacy issues that we encounter and we can expect to encounter for many years and decades to come.

I looking forward -- looking forward, we have to continue, as my office has done well before I came aboard, asking the very tough questions, the increasingly tougher questions, with limited resources, while knowing that DHS must adapt and use the tools at its disposal to counter these Hydra-headed and ever-evolving threats.

It's certainly an exciting time and a challenging time to be working on privacy and transparency at DHS and I very much in this regard appreciate all of the work of all of you to our efforts and your contributions to our efforts.

With that, I thank you.

CHAIRPERSON SOTTO: Thank you so much, Ms. Neuman.

We are very tight on time. I will allow one question. Joan McNabb had her card up first.

Ms. McNABB: Before I got cut off.

Thank you, Karen. You've been busy, of course. I was wondering if, with the various forms of best practices guidance that you have developed and are developing on the use of drones, if you provide any of that or intend to provide any of that to the recipients, state and local recipients of grants to purchase such equipment?

MS. NEUMAN: First of all, Scott Matthews in my office is sort of the subject matter expert on this and I would direct you to him. We do not sort of rent out our unmanned aircraft.

Ms. McNABB: I mean the grant money that DHS provides to state and locals, which is often used to purchase those things. Do you also give them advice?

MS. NEUMAN: To the extent we would make systems available we do condition our grants on compliance or adherence to privacy best practices. And there are a whole bunch of other areas where we are involved in grantmaking activities, for example in the context of FEMA awards, where privacy best practices are an integrated component of the grant to states, state and locals.

But I encourage you to contact Scott Matthews in my office if you want to drill down into some of the details.

Ms. McNABB: Thank you.

CHAIRPERSON SOTTO: Thank you so much.

MS. NEUMAN: Thank you.

CHAIRPERSON SOTTO: May I ask Andy Ozment to please come forward.

(Pause.)

CHAIRPERSON SOTTO: Thank you very much, Mr. Ozment. We're delighted to welcome you. Mr. Ozment is the Assistant Secretary in the Office of Cybersecurity and Communications and previously served as the President's Senior Director for Cybersecurity at the White House. Mr. Ozment will bring us up to speed on the Department's cybersecurity activities. This is a topic of continuing interest to the committee, so we're excited to

hear you.

Please proceed.

DHS CYBERSECURITY OVERVIEW

MR. OZMENT: Wonderful. Thank you, and it's good to see everybody, and I see some folks I know well and have interacted with in previous efforts. So very glad to be here today.

My plan today is to talk first a little bit about our office in general and the approach that we take, and then about some programs that I think are of particular interest to this group, to include the EINSTEIN program and the ECS program and then some of our other work in information-sharing under EO 13-636. And I want to leave some time for questions at the end, and of course if we have questions en route, particularly if there's any uncertainty or lack of clarity about something I'm saying, I'd love for you to stop me and I can make sure it's all clear.

First, let me say -- let me talk about my current organization, Cybersecurity and Communications. We fit within the DHS Office called NPPD, the National -- National Programs and Protection Directorate -- sorry, Protection and Programs Directorate. It's one of the worst acronyms in DHS. We're working on that.

But essentially, our broader organization looks to protect critical infrastructure from all hazards threats. Then within that organization, my organization focuses particularly on cybersecurity and communications.

Now, our mission as we view it, then, is to make cyberspace and communications more resilient. In part, we choose

the word "resilient" because we are not after perfect security here. This organization is very familiar with the fact that you don't get there. So we recognize that we can't get perfect security. What we can do is help our customers manage their risk and be resilient in an environment where the interconnectedness of infrastructure and cyperspace are increasingly resulting in fragility, if you will, of that critical infrastructure.

So we view our customers as critical infrastructure in the private sector, as other civilian federal government agencies, and as state, local, tribal, territorial governments. If you've been on this committee a while, you've probably heard the terrible DHS acronym "SLTT" to stand in for "state, local, tribal, and territorial" governments. I may use that acronym. I'll apologize in advance.

What do we do with these customers? How do we help them? Broadly, we have three ways that we help them. First, to understand and manage their risks; second, to reduce the frequency and impact of any incidents that occur; and then third, to build their capacity -- all vis a vis threats against cyberspace and communications.

Strategically, our organization has two guiding principles, and I'll tell you that these principles are one reason why I left the White House to come lead this organization, because I think these principles are truly what make our organization unique in federal cybersecurity.

The first one is customer service. We are not a law enforcement organization, we're not an intelligence organization.

There are many things that we are not. Our sole goal when engaging with our customers is for them to be more secure and resilient than before we started working with them. So we have no other goals in that engagement and I think that's important.

I'm not saying that we don't need those other functions in government, but I'm saying it is also important to have an organization whose sole purpose is to leave their customers more secure and resilient.

Perhaps more relevant to you is the second principle, which is the need to protect and, when possible, enhance privacy and civil rights and civil liberties. I view this as critically important in cybersecurity as a field and truly one of the strengths of my office, and that's why I'm here today. That's why I'm going to meet with the privacy advocacy community, which Karen is graciously hosting me for, next month. And that's why, among other reasons, that's why you'll see a lot of engagement between my office and the Office of Privacy at DHS and also the Office of Civil Rights and Civil Liberties.

Frankly, we're incredibly lucky to have these broader DHS offices. Working from the White House, I interacted with privacy offices across government and I think you all recognize that the DHS offices both of privacy and CRCL are far stronger than any other offices, equivalent offices in the rest of government.

I did actually steal somebody from PROF, but it was before Karen got here, so I'm hoping she's not holding it against me. And then I came to DHS, so I shot myself in the foot on that

one.

(Laughter.)

I also want to highlight that we also have our own excellent team at DHS, led by Emily Andrew, who is right here as well.

So that's within NPPD, my parent organization. Then we have -- their team has provided staff that sit in various offices within CS and C, my organization, to help us out.

I think my final point on that, and I won't belabor it any further, is from a cybersecurity perspective I don't believe that our nation can be resilient against cyber threats unless our citizens and other people have confidence in the actions that we take in cybersecurity. That means that we have to do it in such a way that they understand what we're doing in they're comfortable with what we're doing. That to me is just simply fundamental, and that's one reason why I view the efforts that we have in privacy and civil rights and civil liberties as really a core part of our efforts.

I'd like now to speak to a few of the activities that I understand are of interest to the group. Actually, are those two stood up new since I started talking? Do we need any clarification?

CHAIRPERSON SOTTO: No, keep going.

MR. OZMENT: If I start seeing a wave, I'll know that I've really sort of meandered off topic.

Some of the things I want to talk about. First, setting this within context. I mentioned some of our customers. Our

customers include state, local, tribal, and territorial governments, the federal government, and privacy sector, particularly critical infrastructure. I think for the remainder of my comments I'm going to focus on dividing those into two groups, the federal government and critical infrastructure. State and local, SLTT, are actually a little interesting in that sometimes they act like the one and sometimes they act like the other. So in different instances they'll fall under either what I describe about federal government or what I describe about critical infrastructure.

Speaking first to critical infrastructure, Executive Order 13-636 came out about a year and a half ago, February 2013. That Executive Order focused on information-sharing, raising cybersecurity best practices across critical infrastructure, and then privacy and civil liberties in conducting those activities. I think there's probably less to talk about vis a vis the cybersecurity practices here, so I'm going to skip that, the best practices.

But the result was the cybersecurity framework, and I think there was a lot of interesting discussion there about the tools given to company privacy officials or private sector privacy officials and how widely understood those best practices were when you took it from a level of the FPS and try to take it several levels more concrete and build it into the practices of an organization.

So I think there is interesting things happening in that Cybersecurity Framework space, and one of the decisions in that

Framework was that a follow-on topic was really to further flesh out principles and best practices in privacy. So I do flag that for your attention. That was a NIST, National Institute of Standards and Technology, -ed effort, and they actually have an RFI out right now -- I hope it's still open -- looking for more feedback on it. So if folks have not seen that, I recommend you take a look.

Now, with respect to information-sharing, there are really three things that the Executive Order tried to accomplish. The first was to issue more clearances to private sector operators in critical infrastructure. That is an important thing to do, but is really a drop in the bucket of raising awareness of cybersecurity threats and empowering owners and operators of critical infrastructure to manage those threats, because there is no way we can or should give a clearance to everybody in the country who is in that role.

So the second thing that the Executive Order focused on -- let me lean back here and move the mike. The second thing that the Executive Order focused on was increasing the quality and timeliness of the classified information or really the cyber threat information that the government passed to private sector entities. We do have a role in that. In particular, implementing this Executive Order, the government has focused on notifying private sector companies when either the government knows that they have been particularly targeted by a cyber threat or are already a victim of a cyber threat.

So in practice what this means is the government is

doing a much better job of going out and knocking on the doors of companies and saying: We've got bad news; you are either being directly targeted by a cyber threat actor or you've actually already been compromised; here's what we know about this compromise; how can we help?

That is I think largely good news. The part of my organization that is focused on that effort is the NCIC, and that is our incident response organization and also our more broad operational organization. So almost everything that we talk about vis a vis information-sharing or incident response, it is our operators in the NCIC who are handing the information or going on site at times for incident response, and so it is those individuals who have a key role vis a vis privacy-relevant information.

I think it's a good time to pause and say that we have very strong standard operating procedures and training for those operators with respect to privacy-relevant information, how to recognize it, what to do if they inadvertently are sent such information, how to purge it from the systems, how to report it, etcetera, and then an oversight and compliance regime as well. Greg got excited about that one.

So let's say we go on site for information or we just find information about a threat, we pass it on to a company and we say: Here's a problem, here's the information we have about it. We help them remediate the problem. Life is good, right?

More broadly, though, the third information-sharing activity under this Executive Order was called Enhanced

Cybersecurity Services, or ECS. ECS has some similarities with the EINSTEIN system, and so what I really want to foot-stomp for this group is that ECS is focused on the private sector, the EINSTEIN system is for the federal government. In this case, federal government is federal government; it does not include state, local, tribal, and territorial governments. So the EINSTEIN system is federal government, ECS is the private sector.

So ECS we view as a form of information-sharing, because what is happening in ECS is that we the government, and in particular DHS, are taking classified threat information -- say, this IP address is malicious or this domain name is malicious, like www.badguy.com. When you see that one --

(Laughter.)

There's probably a business there and I really have no idea. So maybe there's a perfectly good business. Don't quite me on that.

Passing that information to private sector service providers. So imagine a company, Acme Security, starts up a service provider, a service offering. We give them this information and then Acme Security goes and sells to critical infrastructure this security service. Now I'm Andy's Power Company, and Andy's Power Company goes to Acme Security and we say: All right, we want to buy this security service. I, Acme Power Company, am going to route some of my network traffic, so some of the information coming to me on the network via the Internet, to you, Acme Security Service, and you will use this classified information to detect and block threats, and that way

I'll be protected.

Now, there's three parties engaged in this: the government, Acme in the middle, Acme security provider, and Andy's Power Company as the customer being protected. The government gives information to Acme, but doesn't know what is triggering for Andy's Power Company. So I don't know that Andy's Power Company went to www.badguy.com as a government person here. What I can know is over time, hey, from all of the customers we serve, 100 of them tried to go to badguy.com. So we the government get aggregate information back, but not company-specific information back.

That's the ECS program.

I'm watching the clock here and realizing I need to accelerate to give time for all these questions. So let me switch gears and actually go now to the EINSTEIN program. The EINSTEIN program predates the ECS program and it's for government only, federal government only. The EINSTEIN program is essentially what -- in the abstract, think about it as the security functionality that every, I hope, private sector company has, which is essentially perimeter defense. So you can think about it as the wall going around the organization, with a guardhouse checking what's coming in and out.

Now, there are actually three separate EINSTEIN programs: EINSTEIN 1, 2, and 3. EINSTEIN 1 is the most basic of the programs. What it is doing is essentially -- by analogy, it's like the logbook in an office building, where you come in, in fact to this building here, you sign your name, who you're coming to

visit, and what time it is, and you go on up. They have no idea whether you are in fact a notorious criminal when you sign your name in the logbook. They're not checking it against anything. They're just recording, here's what happened.

That's EINSTEIN 1, and again this is for traffic coming in and out of the federal government, federal civilian government, not DOD, not the intelligence community.

EINSTEIN 2 takes it a step further and it's an alarm system. So it's actually, all right, here's who's coming into the building, checking against a list of known bad actors, if you will, and sound an alarm.

EINSTEIN 3, or it's actually EINSTEIN 3A for "Accelerated," is like a guard. So find out, check the name of the person coming in against a list of bad guys, don't just sound the alarm; actually block them from entering the building. And that's an intrusion prevention system, whereas EINSTEIN 2 is an intrusion detection system.

Now, a few things I want to flag. First of all, we're obviously talking about network traffic and not people. So what's actually happening are packets of network information are coming in, they're being examined for sort of known indicators of malicious activity, and then those get acted on, either by alarming or, for EINSTEIN 3A, by preventing the actual intrusion.

The other thing I want to flag is that this is looking at traffic coming in and going out, because often what happens is we detect an intrusion not as they break in, but as they try to go out, either to contact the computer that's controlling them or to

take information and move it outside of government networks. So again, these systems are looking at traffic both entering and leaving the government.

I think with that, let me stop and open the floor for questions.

CHAIRPERSON SOTTO: Thank you so much, Mr. Ozment.

I'm going to take one minute, the Chair's prerogative, and ask a question. We've seen an absolute sea change in the private sector with respect to the sharing of threat information, really night and day from five years ago, when very, very little information, if any, was shared, and now I would say a deluge of information is being shared. Maybe we're in a TMI situation.

What I'm not seeing, however, is really any discussion about privacy issues when there is the sharing of information back and forth, and really we see the one-way information-sharing with government to the private sector, and then the private sector makes the decision, are we going to share information back. Really, I have not seen, and I've handled very, very many breaches for clients, I haven't seen any instruction by the Secret Service, for example, about what information should not be shared with the government or how to protect that information because there may be some personal information or information that needs to be shared from a privacy perspective.

Could you comment on that, please?

MR. OZMENT: Absolutely. I can't speak to the Secret Service. They're a separate organization. I don't know their practices.

But I can speak to when our organization, for example, goes on site and responds. When we go on site, our goal is twofold: first, to get your organization back up and running; and second, to understand or help them understand the extent of the intrusion or attack, both to get them back up and running and then also so we can take whatever lessons that are learned from that and then share them with other private sector companies so that they can not suffer the same intrusion or attack.

I'll highlight. I make a distinction between intrusion and attack. I think that's one area where we often have quite bad terminology. There are very, relatively speaking, few attacks in this area, many intrusions. Generally what's happening is an intrusion for the purposes of crime or espionage, which is crime, but not for sort of disruption.

When we do that, again our people who go on site have privacy training, they have standard operating procedures. In general, their first goal is to not get the data in the first place. So often when the organization they visit is sophisticated enough, our people are working by asking questions of the people who actually are in the organization and saying, well, would you check on this, I recommend that you have this capability, that kind of thing.

In fact, only in the most extreme cases and after signing many more elaborate legal agreements do our people ever put their hands on any keyboard whatsoever. In some companies that's just not even possible, just given the construct.

So I think that's all I can say from a specifics

perspective, is to say that we have really trained our folks to recognize what information is privacy relevant and not to pass it back whenever possible. Now, I think we all recognize that in cybersecurity a fishing email is very security-relevant and it may be that the subject line of a fishing email is what is used to identify it, and that's obviously information that we should be concerned about from a privacy perspective. So I'm not saying that we never see that information. I am saying that we're trained to look at it, to say very carefully, is this actually a fishing email, let's make double sure, here's how we process it, here are the purposes that we're using it for, etcetera.

CHAIRPERSON SOTTO: Perfect.

Mr. Pierson.

MR. PIERSON: Joan pulled her question.

A quick question. The Executive Order of February of 2013, Cybersecurity Framework February of 2014. You have the more publicized launch of C3, even though it existed in different formats prior to that as part of the PCI program. A question for you. It was announced in March, March of 2014, that through a deal with DHS and PPD specifically and the MSISAC, so Will Pelgren's group, that the MSISAC services, which essentially could be seen as a 24 by 7 for third parties, in this case that it was going to be offered to state and local governments, that that was going to be available. DHS, Phil Schneck, had struck a deal between DHS and the MSISAC for the provision of those services to the states.

My question is simply this: Can you update us on how

many states have been eager to move in that direction, utilizing the MSISAC services, how far along it is, and, quite honestly, any positives coming out of that as a part of the outreach activities between DHS and PPD and the MSISAC in terms of furthering cybersecurity within that area?

MR. OZMENT: Thank you. That's a great question. Unfortunately, I'm going to have to say right off the top I can't tell you the number of states. We'll take that as a get-back.

What I can tell you in general is what this program was, is the MSISAC, the Multi-State Information-Sharing Analysis Center, has been set up as the equivalent of the private sector information-sharing and analysis centers that are helping individual sectors be cognizant of and mitigate cyber threats. It is funded each year in part by a grant from DHS or has to date been funded each year.

They have a program called ALBERT. It is separate from, but modeled after, EINSTEIN. Hence "ALBERT," Albert Einstein. The best naming we've got, so I'm kind of proud of it, actually. Everything else is an acronym.

(Laughter.)

So their ALBERT program is modeled after EINSTEIN 1. If you think back to my analogy of the logbook that tracks when you are entering or leaving a building versus the alarm versus the guard, what they're doing right now is the logbook. So that's a service that they offer to states.

What we announced, the particulars of what we announced, are states can receive that service for free if they adopt the

Cybersecurity Framework. So even if they choose not to adopt the Cybersecurity Framework, they can get the service, but they do have to pay for it.

I think it is fairly widely adopted amongst the states, but I cannot tell you the exact number. So we'll come back to you with the exact number.

MR. PIERSON: Just from an "ish" perspective, I mean, are you talking a handful, under ten, or a dozen?

MR. OZMENT: I think we're talking three dozen, give or take.

MR. PIERSON: Okay, so pretty widely. I'd call that pretty widely adopted.

MR. OZMENT: Yes.

MR. PIERSON: Okay. Very, very good.

MR. OZMENT: Really, don't quote me on that. I want to get you a precise answer.

MR. PIERSON: Thank you.

CHAIRPERSON SOTTO: All right. Greg, you're up.

MR. NOJEIM: Hey, Andy.

I've been involved in the cybersecurity information-sharing legislation for about four years now and there's a --

MR. OZMENT: Did you finally get that sorted out?

(Laughter.)

MR. NOJEIM: There's a perception on the Hill that companies are telling Congress that the information that they get from DHS is usually not timely and usually not actionable. From your presentation, I get the sense that that's changed. If that's

the case -- you know, when I go onto Amazon and I want to buy something, I always look at the customer reviews -- where are your customer reviews? Where is the stuff from the companies that says you're doing a great job and, frankly, that you, DHS, with a customer service hat, are a good alternative to the other thing, which is the NSA, which has a surveillance hat?

MR. OZMENT: That's actually one of my goals, is to obtain those reviews. I don't think we have a great recording. We have various sort of customer service surveys that we've done, but not published, as we're just trying to make sure that we're doing a good job. One of my goals is to get exactly that information, because we have to change that discussion on the Hill.

A few things I want to provide some general thoughts, though, on your comment. One of them is I think there's two particularly relevant types of information-sharing, maybe three, that matter in this space. The first one is is that knock on the door, right now you're either targeted or you're a victim already. The second one is threat indicator sharing, so look at for badguy.com.

The third one is the more -- more generic alert, that is not quite as actionable and timely as "look out for badguy.com," but gives you some context and some ability to shape your practices. A good example of that is a bulletin we put out in July with the Secret Service about the malware or the malicious software that criminals are using to infect point of sale devices, like those machines that read your credit card. We've seen a few

incidents lately involving point of sale devices. And we've heard directly from some companies that they took that, they went and looked at their devices, and in fact they found that they had infections they hadn't known about.

Let me talk about the sort of reviews in each of those. One of the challenges is sometimes our greatest reviews are reviews that nobody will ever talk about. That last example, somebody went and looked and found that they had an intrusion they weren't aware of, they're really eager for us not to talk about that. So one of our challenges is often when we do best the companies want to remain anonymous. Word will nonetheless still percolate around from that, but it is a frustration.

I do think there's this interesting second situation here, which is -- some of it is -- I do believe that there is this strong perspective in a lot of the private sector that the government has the crown jewels of information and we're just not sharing it. I think -- that we knew about this bad thing long in advance and then by the time we get it to you it's stale.

I am confident that there are things that we can do to improve the speed with which we share information, but I also believe that over the last five years the private sector's knowledge of what's happening with respect to cyber threats has increased so dramatically that the government really doesn't have the monopoly on really useful cyber threat information. So some of the comments you're seeing are not so much about us as the vehicle for sharing that information and more about the fact that the information that the government now has is less unique than it

was five years ago when there was a less robust private sector market surrounding this space.

That's one reason why -- our argument in information-sharing is that a large part of the value is not just government sharing to private sector, but it is us hearing from the private sector what's going on, but also fostering private sector to private sector sharing. Now, all of those have privacy concerns and considerations, not all the same in each of those scenarios. But I think whatever we do on the Hill, it does have to tackle all three of those scenarios.

I appreciate the fact that you're going to tackle that and get us that legislation in good shape. Thank you.

CHAIRPERSON SOTTO: All right. Mr. Adler, you have the last word.

MR. ADLER: Thank you.

Andy, great presentation. This really question is around the blacklist that you're building for ECS and EINSTEIN 2 and 3, and sort of picks up on Lisa's point around the deluge of information. How often, and are you tracking how often, you get it wrong. So on your blacklist there's a lot of information that kind of helps our deliberations on some of the topics we're dealing with on the committee. Do you track those numbers and try to drive them in a positive direction?

MR. OZMENT: There's two ways of getting it wrong here. One is I tell you that it's badguy.com that's bad, but you actually block badguys.com, some slight variation. Functionally, that happens never. It may have happened a time or two. We do

track that, but that is not really a problem here.

I think the more interesting way and probably what you're referring to is we say that badguy.com is bad and it turns out not at all; it's a tee shirt company and we shouldn't have blocked them. I don't know the answer to how often that is happening.

We are tracking that and, more broadly, the good news is -- so I did a Ph.D. in computer security and what I focused on were where incentives are misaligned, where we know how to secure something, but we don't have the incentives aligned so that we secure it. So I'm always looking at incentives. The good news in this space is our incentives are aligned with not having that false positive happen, one because, frankly, we lose our customers if the service we're providing isn't helpful to them.

One of the things that we track very closely is not just is it good or were we accurate with this indicator, but was it unique. In other words, were we providing value that did not exist in the sort of known list of private sector, that you couldn't have bought from a security company, essentially.

Now, we're never going to get that exactly right and all the security companies have different data, you name it. So I'm not looking for us to be 100 percent unique, because maybe we got it earlier, you name it. But we are looking quite closely at that. It's harder to analyze than you think. But in general we're doing a pretty good job on uniqueness as well.

I can't tell you exactly how many, but I can tell you that essentially our incentives are very well aligned with your

desire to not have a false positive on the blacklist.

MR. ADLER: Thanks.

CHAIRPERSON SOTTO: Thank you very much for a terrific presentation. We really appreciate it. I'd ask the next panel to please come forward.

MR. OZMENT: Thank you.

(Pause.)

CHAIRPERSON SOTTO: Thank you very much. Moving right along, we'd like to continue our discussion of Big Data and the implementation of the DHS Data Framework. This committee has spent several years looking into this issue and now that the pilot programs are moving into limited production capability, we welcome today's panelists to bring us up to speed.

Kellie Riley is the Senior Director for Privacy Policy in the DHS Privacy Office. Kellie will get our discussion started. Thank you, Kellie.

DHS DATA FRAMEWORK - BIG DATA

MS. RILEY: Thank you very much. I have with me today Clark Smith, who is our Chief Information Officer for the Intelligence and Analysis Office here at DHS, and Donna Roy --

CHAIRPERSON SOTTO: Would you use the mic?

MS. RILEY: Sure. And Donna Roy, who's the Executive Director for the Information-Sharing Environment Office. Is that better? No?

MR. RICHARDS: Yes, good.

MS. RILEY: We would like to provide you today with an update on the DHS Framework. Clark is going to start us off with

some information about what we did in the pilots. A lot of you are familiar with that and have been briefed. We understand that there are some new members of the committee and others who may not be as familiar, so just to level-set a little bit he will start off with where we've been.

I will talk a bit about the lessons that we learned in that pilot, what they mean, and where we go; and then Donna will conclude with a discussion of our next phase of this Data Framework project, the limited production capability.

We have some slides. We may not be one to one on the slides. In fact, we may go back and forth a little bit on some of the slides. I'm looking to Shannon to see if that's okay. So we will do our best to move the slides, but mainly we will hopefully just have a conversation with you.

With that, I'm going to turn it over to Clark to start us off.

MR. SMITH: Thank you and good afternoon. Thank you for being here.

What I want to talk -- just to level-set everybody and also to kind of tell the story a little bit, first of all let me explain a couple of terms that we'll be using as we go along here. There's the Homeland Security Intelligence Enterprise. DHS, the Intelligence and Analysis Office is a member of the intelligence community. However, there are also parts of DHS that have intelligence components in them. So CBP has an intelligence shop, TSA has an intelligence shop. And they're part of what's called the larger DHS Intelligence Enterprise.

Coast Guard is also a member of DHS, of the intelligence community. So actually they operate under Title 50 authorities. And there are other parts of DHS that do things that are considered intelligence enterprise type activities, everything from as basic as answering questions from the intelligence community or things that we need answered. But mainly they're focused on their operational missions. If you look at the cargo targeting, if you look at ICE, it uses intelligence to do some of its investigation type work in law enforcement intelligence. They have technical collection mechanisms and stuff like that that are used.

First of all, let's start off with DHS's is this hybrid organization that's got lots of different authorities, and it collects data, as you all know, and pulls data together. There is a Title 50 NIC portion of DHS, and one of the first questions you've got is, well, how do you take intelligence information, many times not classified by the piece of information itself, but by the fact that we know it, the intelligence community knows it or has it, and search DHS's holdings with that piece of information?

For example, we have threat reporting coming in saying that someone is on an airplane and they are this type of person, that type of person, and they are going to do something bad on the airplane. Well, how do you then compare that to what DHS has? Well, you would take that to TSA and say to TSA's intelligence shop: Hey, we've got this threat reporting here; you need to do something with this quickly.

The first thing we need to do is we need to be able to take DHS unclassified data or operational data or DHS Homeland Security data and be able to look at that from the standpoint of intelligence with classified information. Again, most of the information is not classified because of the data itself, but because of the fact that we know it or it's been collected.

So we have this challenge with searching DHS unclassified data with classified criteria, classified information, because we know most of the reporting that the intelligence community sees is not so clear as: The guy sitting in row five, aisle 20, of the airplane is the bad guy or the bad person. It's usually much more unclear than that.

The other piece we have to do is resolve the data into entities. We have lots of data that's very mixed. We have data that's collected in different places, and be able to bring that data together and be able to say, what does DHS know about it? My name is Clark Smith. You can put my name into a database and come up with lots of different matches for my name. So how do you know I'm the Clark Smith that lives in Washington, D.C., that actually works for DHS, those type of things? How do you do entity resolution? Because we don't want to have false positives. You were talking about that earlier, and those different things.

The other thing is to do analytical tools. This is where that Big Data concept comes in. That means a lot, and there's a lot of the world of Big Data. But the very first basic thing that we're talking about is just searching DHS data with classified indicators or classified information and being able to

do that in such a way that we protect the sources and methods of the classified data, at the same time making the connections we need to make to protect the country.

So the Big Data pilot started off in that area, with that kind of goal. So we came together -- and Donna has been working this issue for many, many years, and I have been working this issue for less than she has, but we came together recognizing that to build this type of enterprise to do those type of things you needed to do, you really need to look at it holistically, not look at it simply as, hey, DHS needs to do these type of searchings or I&A needs to do these type of searchings, let's build that one little system.

You really had to go back and step back and look at the whole picture and look at the fact that this data is collected and being maintained for many different reasons. There's lots of issues around redress or accuracy. There's lots of issues around how it's used and what's the authority to collect. There's all these complex things that really aren't technical things. They're very much in the big picture about privacy, civil liberties, all the things that we're worried about here at the panel and this discussion, beyond just simply, I want to build a system and make this thing happen. You've got to think about the larger picture.

So the Data Framework is really getting at those pieces of the pie, is how do you recognize that there's DHS data you want to search from a counterterrorism perspective, but that data's not available for other types of searches? You can't just hand it to I&A and let I&A do whatever it wants to do with it, because it

really is not collected for purposes beyond like a CT purpose, or it wasn't collected for a purpose beyond an immigration decision purpose. So you have to wrap the data with that kind of understanding of what's the use, why you plan to use it, because this concept of hand your data over to the IC is not something we're just comfortable with: Trust us; we'll manage it carefully. I'm being sarcastic.

And then also recognizing we need to be dynamic in that access control, because an analyst at I&A who's doing a CT case one moment may suddenly turn to another package, may have another package and may be doing another type of case at the next moment. So that we're not a big shop of people at INA and they do many different things. So recognizing how do you technologically build into the system this concept of use, this concept of what you're trying to do, and the concept of context.

Then also recognizing that we wanted to -- bulk data sharing, the concept of handing data over and letting people do whatever they want with it doesn't really include a logging component. You can kind of force it, but we wanted to build some kind of system that had immutable auditing and logging tied to it, that was then tied back to our core values of trying to make sure we could identify what people did, why they're doing it, and even in fact build into the system the fact that system administrators can't touch the data, can't see the data.

So it was meant for just the users only, but it was in our systems that we can build in the proper protections inherent to the systems that we built together.

Pulling that together, we built this DHS Data Framework that we feel like inherently in its own architecture and design of the system protects the civil rights and civil liberties. It's coded into the system with some of the tagging that's put together. It also has a strong governance and oversight component to it, and then it still allows us to do our classified searching of unclassified data over authoritative sources. It's not just simply another replication, another box, but it's tied back to the authorized source system. It's a tight tie, so if you want to do a correction of the system the I&A analyst isn't correcting on the classified side, but actually going back to the unclassified side saying, hey, assuming that the correction is unclassified, hey, that's not right, you need to fix that at the source system. That's the other piece that we tied into this.

So there's a couple tenets there, all that bringing us together to a platform that we can use to answer the questions that we need to answer in a way that's actually valuable across all of DHS, to many different missions, many different pieces, while at the same time recognizing the complexity that we're doing our jobs within.

That's kind of a set-up. I know that's a lot of touch points I tried to hit upon. It's not a technical set-up. A CIO talking like this is probably scary to me. But the next concept was, as you can see -- this is what we focused on. It was a lot less about the technology. I could sit up here and talk about routers and systems and what software we bought, but that's really not the problem we were really faced with.

The problem we were really faced with was how do we wrap the data properly. So I'll pass it over to Donna now to kind of get started on that -- sorry. I guess I should talk about the pilots first.

We put together the pilots using three data sets at DHS: Alien Flight School, which is all the folks coming to the country from across the world to fly, to get flight lessons, a 9-11-derived data set, of course; SEVS, which is our Student Entry Visa System, so all the students coming into DHS from across the world who are coming here to learn, higher education; and then ESTO, which is our system that we're doing for the visa waiver program countries that are coming in and out.

We picked across three different -- that's CDP, TSA, and ICE, so we picked different components of DHS. We wanted to make sure everybody got a little bit of a touch on this. We pulled the data together and then, through many different conversations between Kellie, Donna, myself, and all the different staffs at DHS, including the higher, the senior leadership at DHS, basically went into a process of wrapping each of those data sets and the data coming out of those data sets with their use authorities and why they were collected, and wrapping who can do what with the data and how it was collected, and then build a system that we call Neptune to collect the data on the unclassified side, put those policy wrappers around it, and make sure that data is properly tagged and that we had all the error checks and stuff going.

That then fed two places. It fed -- on the high side,

it moved up to a cross-domain transfer and went to Cerberus, which is where we actually do the searching that I just described earlier. The classified search parameter, searching of unclassified data, happens at Cerberus. It also fed a system which Donna will elaborate more, called Common Entity Indexing, on the low side.

So please be aware that Neptune's really a data lake. You get all these different kinds of terms us technologists like to use, depending on what we're branding. But basically it's a data store that either moves up to the high side to go to Cerberus and then is allowed to do the classified searching or it moves across and it's used for the Common Entity Index Tool, which is basically indexing -- back to my point about I'm Clark Smith, there's a lot of different data that, if you thank you typed in "Clark Smith" as a search parameter, you'd get on different people.

So what did we show in the pilot? One is that we used real DHS data. We were able to tag it. I wanted to make sure that the actual system would run, in other words that you wouldn't type a search parameter in there and the whole thing would just crash because all the policies crashed each other or they didn't work. We showed that. We demonstrated that you could search across the three data sets, gave appropriate access and controls. I was worried about if you switched different users or different authorities that things wouldn't work or they wouldn't enforce or the data wouldn't properly manage.

So we showed all of those key things work. We had

performance, no performance problems, great searching. In many cases -- I won't name the systems, but we were throwing search criteria in our test environment that would not have been possible on the original systems without having to do a special query. So you would have had to call up the owners of the system, for example Alien Flight School, and said, can you build me a special query to do this search, because the actual back end technology is so -- it's not strong enough to handle the search queries that we were throwing at it, whereas in this system, the server system, we could.

Not to say we couldn't do those searches. We would just have to make special requests to do those searches because of the technology. Now, the technology upgrade will allow you to do that itself, which is the type of things we're talking about.

It showed value with the Common Entity Indexing. We also showed a tracking of all the user movements. So any time a user put a request in it would record their query string, record the permissions they were operating under with that query string, so recording that, and then also -- it wouldn't record the results, because that would then -- if you do that, let me know; I'll go buy a hard drive company and stay back and make a lot of money. But it could go back and give the same query to the data set and give the same results out.

But the concept is we stored, had autologs on all of what was going on. Then if they clicked on anything off that, like did a deeper search into a query, a file, it would record every movement after, what they got and any movement they would do

after that. So it would not give the results, but it would say -- let's say they looked at the results and they picked the fifth line and did a deeper query and said, give me more of that on that file, it would record that look, too.

It also -- we demonstrated, which is one of my things, is that the system administrators couldn't see the data. So a systems administrator going in and tuning the system and looking at the system, trying to make sure that the system was running right, doing the data uploads, make sure they ran right, wasn't getting full access to the full-blown data itself. They would get told, hey, great, thumbs up, it came across real well. But they weren't actually able to go in and drill into the data itself and just walk across the data, like you do in most of your database systems.

We also ran a governance structure, which was a lot of work, and then we have been trying to keep you all apprised. That was kind of what we did in the pilots. Those pilots were completed earlier this year, back in January. And we've been working through the last nine months, has been doing the announcements, going operational, and then Donna's going to describe what we're teeing up for right now, which is limited production capability.

MS. RILEY: Before Donna gets into limited production capability, I'm going to talk a little bit about some of the lessons learned from the pilots that Clark just alluded to. There are five lessons learned that I'm going to highlight that we've discussed in our privacy compliance documentation moving forward,

but I think they're worth just noting here.

Among them are governance. In developing the Data Framework, we knew this going in and it was validated going through it, that we need to establish a strong governance process in order to evaluate the integration of new data sets as the Data Framework develops new missions, new uses, and new analytical tools. We need to have something in place.

It worked very well. As many of you may remember, the Common Vetting Task Force -- and Michael Frias sends his apologies. He was going to be here today and at the last moment had to -- had an emergency he needed to deal with. But Michael is the head of our Common Vetting Task Force.

They, that group, which Privacy and other oversight bodies and I&A were all members of, was the initiating body for this project. But we knew that it was a task force and that eventually we would move out of CVTF into a governance process that would live with the maturation of the Data Framework.

So we're in the process now, and Donna may talk a little bit more about this when she talks about limited production capability, of figuring out exactly what that will look like and how that will function. But the general concept is that we would have an executive steering committee or something called either an executive steering committee or called something else, that that will include oversight, including privacy and civil rights and civil liberties, and have a body that can make determinations about prioritization of data sets, which data sets are appropriate to go in, because not all data sets will be appropriate to go in.

If one of our goals, as we've stated, is to stop aggregating, having many, many aggregated data sets all over the Department, then we need to take a look at what are our authoritative systems, what are our source systems, and how will that play out, and there will be priority issues that we need to look at, because we can't do all of this all at once.

The other lesson learned, I think, that comes out of this is that the incremental development of this is very important. It is very complex, and we I think had great success in taking a small portion first and figuring out what the rules are and how this plays out before we start adding everything in. It doesn't always happen that way. Sometimes we are faced with a project that everything's going in and the whole project needs to be done now or yesterday or two weeks ago.

But I think that we -- and I think my colleagues would agree, we've been very successful. Really, we want this to work from an oversight perspective. Clark's group wants it to work from an analysis perspective, as done Donna's group in OCIO. And the incremental work on this has been very valuable in that regard.

We also recognize that redress and refresh are very important things that need to happen as the Data Framework matures. The pilot was a static data set and we moved it from the source into Neptune, and in the pilot project there was no effort to have a refresh because we needed to understand that the access controls and the tagging was all going to work. But we know that that's a critical privacy and civil rights, civil liberties issue,

that we have the ability to refresh the data in a timely manner, and what the timely manner is is something that we will look at in the limited production capability. We need the mission stakeholders involved in that determination and it will be very important to know that we can refresh data from the source system in Neptune and in Cerberus wherever it goes. That will be important for people who want access to their data.

We need to know where we're looking and what they get when they make a request of us for the data and ensure that if there is a need to correct the data that it gets corrected -- Clark alluded to this a bit -- in the proper places.

Stakeholder engagement, which I have mentioned, it's critical. It's critical that the mission users and the operators understand the value of the Data Framework and are willing to use it, because we can do a lot of work making sure that we have folded privacy concerns into the development of the Data Framework, but we need to make sure that the mission users and the operators are at the table, because this has to work for them and it's all a conversation, what works for them, how do we get the privacy controls in, so that they recognize the value of having this thing that we have created, as opposed to the -- "ad hoc" is perhaps not the right -- the aggregation of data sets that they pull in for their own mission use. One place to go, to stop aggregating the data, is very important.

Then transparency is critical and we recognize that. We have made efforts. We've put out privacy compliance documentation, PIA's, we've updated SORN's. And as you know, we

have tasked the DPIAC -- I actually look very much forward to hearing your discussion after this -- on how we can do better on transparency, because I think it's a very important concept from my perspective, I think from my colleagues' perspectives, too.

We've had a very good working relationship and we need to be able to explain what we're doing, why we're doing it, and how this will proceed. I think they're all very important, and we've expanded on them in our compliance documentation.

I'm going to turn it over to Donna now and let her talk about our next phase, which we are calling limited production capability, and she may touch a little bit more on the lessons and how we're going to implement those going forward.

MS. ROY: Thank you. I've enjoyed a long relationship with this body and hopefully you won't kick me out of the room today and we can continue to do our work.

We are entering -- as of tomorrow, we'll have authority to operate on the last piece of the Data Framework, Neptune, which is the last piece to go into production. Cross your fingers. I heard the paper's on its way to me.

That starts a limited production capability, and it's limited by the following ways. We're using the same data sets, the same three data sets we did in the pilot. But we're using all of the data for each of those data sets for as long as we can sort of use those data sets, within the current retention period, all of the fields that are accessible and all of the data. So it's a full data refresh as of, hopefully, tomorrow and going forward. So that's the first restriction, only those data sets.

Only 20 users each from I&A and CBP, so limited users, limited data, but for production use, for the real analytical use.

Now, knowing that there are some restrictions on that, we would need to ensure that before any action was taken on a new system on those three data sets we would check with the source systems and we would do the operational diligence to ensure the data was right. But it is for full correction use, and the last round really was real data, but for testing of access controls, testing of technology, testing of bandwidth, all of those things.

This is the testing of the real thing, for limited production. We are using the same queries that are described in the PIA's, the associated PIA's, the named query, the characteristic query, and the broader query. The same queries, the same audit logs, the same data sets, limited users, and sort of a fixed time period.

What we hope to do in that limited production capability, not only to understand how the users are going to apply their analytical processes and the tools that we have given them to solve real problems, but we're going to continue mission case development. Mission use case development is the heart of the Data Framework, it really is. It's a granular conversation about each type of user and each type of component within the intelligence enterprise, what they can see, what they can do, what types of queries, and what types of information they should get back. Very granular, and it is done on a user by user, component by component basis.

Part of the reason we're limiting the production is

because we could only get through two use cases in the time frame that we got through, which is the CPB use case and the I&A use case. But we're going to continue the use case development to get ready for the initial operating capability.

We're also going to continue getting to the optimum data refresh for each one of these data sets. We think that we can get to a conversation where we can potentially get to near real-time on some of these data sets. We're not there today in the limited production capability, but we're aiming there, and we're going to continue the technology project to get us there.

We're going to deliver a final mission needs, operational requirements and CONOP part of our system development life cycle documentation to make sure we've gotten the critical functions. And we'll do that through the governance structure, the new governance structure we set up for this program.

We will optimize the data onboarding process to get ready for new users and new data. We won't add new users and new data, but we'll get ready to add new users and new data. And we'll begin to understand the Common Entity Index mission use cases in a more granular level. That's not part of the limited production capability. The limited production is just for Cerberus use cases and for the analytical processes there.

Parts of the tech stack were put in place optimistically based on what we had on the shelf. So we'll start an alternatives analysis to make sure that what we've got will scale to the intent of the Data Framework and that we've got the right combination of tech products out there.

That being said, we hope to finalize this with end of the limited production capability being the standup of the new sustained and executive-led steering committee or steering group, whatever name we have for it, that guides the Data Framework. We believe that the representation is key of the intelligence enterprise users or the individuals who run those parts of the agency, the intelligence enterprise. The data providers are critical to that oversight, are critical to that.

So we're expecting a broad set of representation to ensure that we guide the Data Framework and we add the right data, we ensure we have verification of those very critical mission use cases as you add new users or as you add new data, that there's a very thorough review at the departmental level, sponsored by the Secretary and the Deputy Secretary.

So those are the things that we'd like to do. We also think that the limited production capability will get us as good as possible a draft of a quality CONOP or sort of an understanding of how we would handle data quality and data redress as either structural issues where the data are found in the data tagging, such as code lists -- we found airport and port codes that didn't make sense to us on the first round; where do you correct those, how do you correct those, how do you add those back in so they come in near real-time?

Then the data quality issues that you get when you actually use the data, not Neptune, but the Cerberus. We found something that doesn't make sense; how do we update in the source system? Then how do we provide redress for a complicated system

like the Data Framework?

Those are all -- although we might not finalize those in the OPC, those are all work packages that we'll start and hopefully gain a significant amount of expertise as we do desk-side visits with our oversight and our users, as everyone learns and understands the training needed, the orientation needed, and the capabilities and processes that the Department needs to put in place before we get to an initial operating capability.

So I think those are pretty much what we're expecting in this LPC period. At this point I think we'd like to open it for questions.

CHAIRPERSON SOTTO: Thank you very much, and that was really a terrific update to a topic we've been studying for now two years, two years or more, yes, with a lot of depth. So thank you very much. That was really helpful.

Chris Pierson.

MR. PIERSON: Thanks. Appreciate the further details, and thank you also for the separate briefings back in, I want to say it was March and April, April and May, whenever. A lot of good detail there as well. Nice to see that the program's moving on.

A quick question. Can you share with us what are the KPI's for success or perhaps the success criteria for maybe these kind of subcategories, but you could go as broad as you want: for increasing the number of users that have access to -- I'm just going to say -- the systems; increasing the number of agencies in terms of -- because part of testing it is making sure of the

attribution in terms of role and agency. You have to in order to test that out. But what are the KPI's attached to increasing agencies?

And-or increasing the databases outside of the current three that were rolled into the underlying pilot. I might have missed another subcategory there, and please feel free to add one in, but I thought that logically those three make good sense.

So the KPI's or success criteria for kind of moving on and moving along, the tollgates, so to speak?

MR. SMITH: I generally believe that the main KPI is data. That ties back to what Donna was talking about when she mentioned use cases. So basically, as an analyst you need to have the right data to ask the questions you've got. So if threat reporting is coming in that's talking about a student taking flight training, Alien Flight School is a great database to be searching. But if your threat reporting is coming in and it doesn't touch one of those three data sets, you need to be looking at other data sets.

So the use cases are kind of tied to the threat stream, number one. They're tied to the questions that come out of the, that threat stream, number two.

So the first piece in my mind, key performance indicator on data usage, is looking back at are you able to answer the questions that are being raised by the threat? With the threat the way it is, are you able to answer the questions that are coming in the system that you've got?

That's a hard metric to quantify, but it's the metric

basically, can I answer this question? If I can answer -- if there's 100 different questions coming out of threat streams over the last two weeks and I can't use any -- and the three data sets, with Alien Flight School, don't help me answer but ten of them, what is the data set that I need to have to start answering those questions? Is there one or two data sets more that would greatly increase what I could answer of the questions?

So if my threat is related to flights, there's people taking flights, what data do I need to answer the flight questions. If my threat is related to unaccompanied minors at the border, what are the data sets that would help me answer those questions? So you look at your threat stream.

I think that ties in very clean to the users. What analysts aren't able -- how many analysts do I have? 20 is a good number at DHS I&A. Do I have an analyst who can't answer questions they've got to answer, or using other analysts to answer questions they've got to answer by teaming on things -- because usually they work in teams. They're not individual; this isn't a single group -- one person doesn't do one thing. They all work in teams together.

Are we saying, well, he's out of the office, but we really would like these questions answered over here by this group over here, because they're now asking questions of the data that we should answer questions on.

So I think you're back to your user set. It goes back again to what's the use case, what are you trying to answer, and do we now have use cases that those analysts could answer that we

don't have enough enrollments or users, user slots actually, to let them answer those questions.

I think the agency question in my mind is much harder, because when you say agency I think outside of DHS. I think inside of DHS -- right now the use cases, again, as was mentioned, we have a CBP use case, we have an I&A use case. That's a really good set of use cases. If there's an ICE use case I think the question is, okay, what exactly is ICE looking at and how does the data fit?

That's a bar I think we can work on. It's when you suddenly say outside of DHS. I think that is not a question right now that I think we've really tackled in the standpoint of starting to say, who else can see this type of information. Again, we're dealing with a lot of mixed information and authorities to collect. That becomes a disclosure if I understand. I'm not a lawyer. You have that question of disclosure, is that now a disclosure and how do we deal with that, and there's a lot of issues inside of that, of disclosure.

Then you'd have to wrap all the policy issues around disclosure into the data, which right now we're not doing in what we're putting together. So I think there's an inherent assumption in the current policy development that it is inside of a DHS conversation.

So am I answering your question?

MR. PIERSON: Yes.

MS. RILEY: I would say the ultimate outcome really should drive all of those KPI's, and the outcome is as simple as

using these analytical tools on DHS data to mitigate the current threat stream. So if we're wildly successful something won't happen, and we're in the business of making sure something doesn't happen. Those are harder things to track.

I would say, though, the expectation to go from limited production capability to an initial operating capability, we would keep the number of users below 100 or so in that LPC period. Anything above that becomes something you can't track the analytical user value in a measurable way, given the resources we've got today.

In the data, we're projecting that the tipping point is somewhere we think around seven to eight data sources before we get that really strong outcome that we're looking for, but that we're targeting 20 to 24 data sources across the Department in a five-year time frame, again pending resources.

The real KPI at the end of the day will be related to the safe use of the data, I believe, given the access controls and the Framework. So it's the capability for our oversight offices to understand the audit logs, to understand measurable controls for abnormal behavior using those logs or to understand, the data providers to understand where they might have to have another conversation about who should use their data given the amount of rejects that are happening and those types of things.

So I think it's the use of the audit logs and the use of that capability that we're most keen to understand in the LPC period, because we think that's where the real KPI's for safe use and safeguarding will be derived.

CHAIRPERSON SOTTO: I'll give the last question to Joanne McNabb, please. And if you could just keep your remarks to about three minutes.

MS. McNABB: I think maybe you've kind of answered it. I wanted to find out more about how you can scale this up, since so much of the bringing systems and users on is person to person, sitting down and meeting and talking and doing it? It's not an automated process. It doesn't seem to lend itself to automation, to a nontechnical person like me. Do you ever see it, the tagging, being kind of automated?

MS. RILEY: I would think that once we get to the tipping point of about -- we've studied this a little bit. Once we get to the tipping point of around eight or nine data sources, the data tagging becomes much more of an 80-20 rule, where you know 80 percent of the tags and you're really handling 20 percent of the exceptions. We've studied this because we think that there's a significant amount of reuse in the rules that we would need to have even to scale and maintain this capability over time.

So we think at about year two and a half or three is when you'll start seeing the acceleration of the use cases and the users and the data sources. I would say that the real driver, though, is how quickly we can get the sources that we are targeting into near real-time sort of capabilities, so that we're providing the freshest data possible. So when we get to that tipping point, we've got a really fresh data source for the analytical efforts.

CHAIRPERSON SOTTO: Okay, I think we will end this

panel. Thank you very much for the update and we'll look forward to more updates in future months. Thank you.

DELIBERATIONS ON POLICY SUBCOMMITTEE AND TECHNOLOGY SUBCOMMITTEE REPORTS

CHAIRPERSON SOTTO: All right. On January 27th of this year, Chief Privacy Officer Neuman tasked our committee with two separate taskings. The committee was to provide written guidance about privacy best practices for notice and transparency related to the Department's use of Big Data, including information-sharing with other agencies and the use of audit mechanisms in the oversight process.

Over the next few months, the Policy and the Technology Subcommittees separately conducted really extensive research into their respective topics, and now they will go through for us their respective findings and explain their findings to us, to the full committee and to the public.

I want to thank Joanne McNabb and Joanna Grama as chairs of their subcommittees for their really tremendous leadership and for rallying the troops, which they did very, very well.

Members of the committee have now had time to review the papers and we would like to hear from the members with respect to any comments they might have. I would ask for significant comments, please. This will -- this discussion today could lead to a final vote whereby we finalize these papers and submit them in final form to the Secretary.

Joanne, would you please start us off with a summary of the Policy Subcommittee's recommendations.

Ms. McNABB: Yes. Thanks, Lisa, and thanks to the members of the Policy Subcommittee, and special thanks to Chris Pierson, who is the Jefferson of this document.

The specific task regarding notice and transparency for the Policy Subcommittee was to answer these two questions: In addition to the published PIA's and SORN's and future updates to these documents, what should DHS consider doing to expand and improve notice to the public? Second is: Should the Privacy Act notices provided as a point of collection be revised to address Big Data in some way, including repurposing of data in source systems, or are there means of notice that could be provided other than that specifically required by the Privacy Act and the E-Government Act?

So in order to address those questions, we got some supplemental briefings from Donna and several others that were very helpful. We reviewed the PIA's and SORN's, and we also reviewed the current Privacy Act notices on the forms used to collect the three main data sets that were part of the pilot project.

But our findings and recommendations are really not just to the pilots. They are to the use of the Data Framework as it develops in the future.

Our findings, in summary, were the answers to both questions was yes; that the expanded capabilities and potential for new users and uses that are not clearly specified, although adumbrated, shall we say, in the existing Privacy Act notices, that it would be advisable to develop other ways to disclose more

specific information on other future potential users and uses than what can be disclosed at the time of collection of the distinct sets.

Our recommendations are, briefly, three things: First, that the existing required notices be carefully looked at and improved to make sure that they are effective communications pieces for the intended audiences, which may mean things like multiple languages depending on who the data subjects are -- in many cases it would mean that -- formats that are accessible to people with disabilities, but also that are accessible to people from different cultures. That's one recommendation.

The second recommendation is that the printed Privacy Act notices be supplemented by the use of web information to provide -- to become a living document that can be more specifically updated as new uses and new users are involved, and that this web site resource can be referred to in the printed notices at the time of collection, and that there are many functions that a more real-time source like a web site could provide to assist people in really understanding.

And then third, that we go beyond these communications efforts to actually get people involved and conduct specific outreach activities and make efforts to go to where -- logical places to connect with some of the data subjects, through their organizations, at touchpoints as they're coming in, and provide other ways of reaching them, to give them a fuller understanding of the way that DHS uses their information.

So, in sum.

CHAIRPERSON SOTTO: Comments, questions? We'll take small comments as well?

(No response.)

CHAIRPERSON SOTTO: Anybody on the phone? Barry? Did Barry have a question?

Barry, you're on. If you have a question, please pose it.

(No response.)

CHAIRPERSON SOTTO: Shannon, did he have a question?

MS. BALLARD: Just continue.

CHAIRPERSON SOTTO: We'll continue on, okay. He can chime in a little bit later.

Any questions from folks here? Allen.

MR. BRANDT: Well, two things. I wanted to highlight one of the final pieces in this draft was, because I think it's unusual enough to say out loud, that one of the recommendations is that DHS do some outreach and literally go -- and it's even commented here. When someone has an incident like a data breach, you become very public. Well, we're actually recommending without having an incident, that just putting DHS in the forefront of being transparent to the users of the system and the citizens of the U.S. and those folks who are coming in. So I think that's an important point to comment here.

The other one is a question I hadn't thought of until Joanne was just speaking. Are our systems that exist today, are they accessible to people with disabilities, and ZoomText and the other type of technologies that are out there for someone who are

visually impaired or need something else? Does anyone know? And if not, I think that should be in there.

MR. RICHARDS: It is compliant.

MR. BRANDT: It is compliant?

CHAIRPERSON SOTTO: Allen, did you want to propose any changes in particular?

MR. BRANDT: No, I think it's fine as is. Thank you.

CHAIRPERSON SOTTO: Linda, please.

MS. KOONTZ: I think these are great recommendations. Thank you very much, Mr. Jefferson.

One thing I wondered if --in the conversation it came up -- about a need for more generalized educational materials for the public. I think none of the notices -- the notice that you get at the time of collection, the SORN, the PIA, they contain great -- they're great sources of information. But I wonder if there is a need for some sort of a road map about how to use that information for the general public, who aren't familiar with these documents whatsoever; and also whether there is a need for information that's really basic to them, like what are your rights.

If you go to the HHS web site, it says -- it has a two-minute video: These are your rights under HIPPA. Very crisp, very short, very communicative. So I was just wanting to follow up on that.

MR. PIERSON: I think that's a great point. We kind of placed that in something we're calling over-communications. So on page 6, second paragraph, last sentence, basically stating that the efforts can be supplemented by some or all of the following:

flyers, other notices, etcetera.

The goal was really that, instead of waiting for people to come to DHS with a problem and instead of being in a reaction mode, that we would be recommending that DHS maybe pivot off of that a little bit and take an over-communications stance, so think proactively about the different touchpoints, the different things that need to be out there, and through that type of a living document, which really can be a web site that gets continually updated and has that baseline information.

But I think we agree, and I think the key here is over-communications, so before there's a problem and not just in reaction mode to make sure that you are providing those materials that would be sufficient for many more individuals than just those are interested to be able to digest, receive and digest that information.

Somewhere in here, I don't remember where we put it, there was the view that there be a way to subscribe to continual notices and continual updates on the web site on the different privacy notices. It's baked somewhere in these pages.

CHAIRPERSON SOTTO: It's right above.

MR. PIERSON: It's in the same place, perfect. Good place for it to be.

Ms. McNABB: Line 208, 209.

MR. PIERSON: Perfect. There it is.

CHAIRPERSON SOTTO: Linda, any new language that you'd like to propose, or are you comfortable?

MS. KOONTZ: I think I'm comfortable. I just want to

make sure that we include something about making information sort of public.

CHAIRPERSON SOTTO: Jim.

MR. ADLER: A couple things. Under redress, which is not necessarily in the recommendations area, but it was line 163, we talk about incorrect data, and this is kind of the redress issue. It's incorrect data and processing, I would argue, not just the data, because sometimes the processing tries to correct the data, and sometimes it does well, sometimes not.

It occurred to me that we should sort of -- where we have the opportunity to cross-reference the recommendations, we should -- might want to try to do that. For example, the redress we also hit in the audit recommendation, number 4, for example, we talk to redress, because that kind of brings the document together, that we're not speaking different language in a lot of different places.

So that's one language recommendation, could we reference at least audit recommendation 4 under redress.

CHAIRPERSON SOTTO: Where would you put that, Jim?

MR. ADLER: Well, incorrect data and processing --

CHAIRPERSON SOTTO: Let's just make sure the sentence reads well.

MR. ADLER: Yes, I didn't actually wordsmith it. I know we have many wordsmiths here.

CHAIRPERSON SOTTO: We need to do that today.

MR. ADLER: You could say -- well, you can start the thing: "Finally, as noted in audit recommendation number 4 as in

a positive, the DPIAC acknowledges." Would that be appropriate?

Ms. McNABB: I think the concept is great. This whole section that redress is in is actually an excerpt from the 2011 paper that identified privacy issues that exist in a system like this. And there was no recommendation.

MR. ADLER: I see.

Ms. McNABB: But the idea of it, if there's another place where we talk about -- we just use that as a way to quickly summarize.

MR. ADLER: Yes, I actually have another spot maybe where it can go. Maybe we can put it there.

Chris, you had mentioned that in the over-communications sentence on line 217, I think is where that starts. We reference in a lot of recommendations KPI's, and one way to over-communicate would be to publish these KPI's where appropriate. Obviously, not every KPI is appropriate to publish, but sometimes it would be.

Where you could demonstrate both the precision and sensitivity, say, of the system,that would be a data-driven way to sort of get ahead of how good is this system that integrates all these data sets. I would sort of leave it to the Department to figure out what's appropriate or not, but it's a way to actually put a little bit of teeth into some of this communication and get a little bit more data-driven about it.

Ms. McNABB: I think I'd avoid saying "KPI's" because that's a kind of insider term. Something like, if I'm getting at it, what you're suggesting is that we say that as part of the information provided in this living document that we might also

make available, that DHS might also make available --

MR. ADLER: The metrics?

Ms. McNABB: Yes, system metrics, system objectives and metrics, something like that.

MR. RICHARDS: What line would that be?

Ms. McNABB: It would be somewhere between 212 and 218. I'm sort of writing a little sentence here.

MR. ADLER: You can almost tack it onto 218, "including system metrics where appropriate."

Ms. McNABB: Yes. So it would be on line 215, right.

MR. ADLER: Put it in there as one of those items in the commas.

Ms. McNABB: Okay, except it isn't a notice exactly. I think it's a separate sentence. So the system -- how about this: "system metrics" -- I want to say "objectives and" somehow, "and metrics could also be included," or "also be provided," just as a separate sentence starting on line 218.

MR. RICHARDS: Objectives?

Ms. McNABB: Yes. "Objectives and metrics could also be provided," included.

MR. BRANDT: "Provided where appropriate."

Ms. McNABB: That's line 218.

MR. ADLER: That's great.

CHAIRPERSON SOTTO: I'm just going to read that one more time. "System objectives and metrics could also be provided where appropriate."

MR. ADLER: Could or should?

CHAIRPERSON SOTTO: What do you like?

MR. ADLER: Should. I like should. "Should" is stronger than "could," but "shall" is pretty strong.

(Laughter.)

CHAIRPERSON SOTTO: Thank you, Jim.

Melodi.

MS. GATES: I have what might be a little bit broader question for the committee, but having heard the presentation that we just did where the system is moving into limited production and about to be operational between now and the end of the year, I'm wondering if we should somehow convey a timing aspect in the recommendations that we're making today, that we're concerned that a certain level of operational activity might occur before some of these things are addressed.

So again, I'm not sure that's a language change, but just a question that came to my mind.

MR. PIERSON: I think it's interesting, the timing of things. Many folks have been working on these initiatives in different formats for probably about the past two years in terms of the overall DHS Data Framework and recommendations on advice and guidance thereof. A lot of the briefings took place in the Q1 and Q2 of this year.

So it is interesting that it is moving forward. I don't know -- perhaps, perhaps what we can do is this, is have a separate cover letter that goes on top of these two Policy Committee recommendations, so that when they are made as here are the thoughts of the DPIAC, policy recommendations as well as

technology recommendations, perhaps we can address that point in the cover sheet as to DHS should look to these papers as means by including some of the advice and guidance that's contained herein, especially given the fact that the various programs as part of the data -- not data warehouse, but the data initiative, are in fact moving to LPC. Something like maybe one or two sentences there as a cover sheet would be a good way to wrap it together?

Ms. McNABB: We encourage DHS to implement enhanced transparency measures as the Framework moves into production.

CHAIRPERSON SOTTO: I would personally rather put it into the recommendations than a cover sheet. Cover sheets go away, but the recommendations --

MS. GATES: My suggestion would be the sort of thing that Joanne just suggested, and that we put the same kind of thing in our other document.

MR. PIERSON: So in the alternative, perhaps, just a one-sentence, one sentence in the conclusion in both papers? Where's my notes on the other paper?

MS. GATES: Yes.

MR. RICHARDS: Would you repeat that, Joanne?

Ms. McNABB: Yes. "We encourage DHS to implement enhanced transparency measures as the Framework moves into production."

MR. PIERSON: And maybe in the other paper, under the "Next Steps" the same thing.

MR. BRANDT: I have a question. It's going into limited production now, so where does DHS know when to put this in?

Ms. McNABB: I think it's sort of calibrated, as in more measures as it gets bigger.

MS. PARK: Privacy measures.

Ms. McNABB: Or these enhanced transparency measures. And I think as the Framework moves further into production.

MS. GATES: Yes, as it moves further into production. Time is of the essence here as this thing grows.

MR. PIERSON: Should it be "security and privacy measures" on the screen?

Ms. McNABB: Well, here I think it's just -- these are all -- if we say "these transparency measures," we know that's what we're talking about.

CHAIRPERSON SOTTO: I'm going to read that back again. It would be on line 245: "We encourage DHS to implement these enhanced transparency measures as the Framework moves further into production." Yes?

Ms. McNABB: And our recommendations are do this and then this and then this. They're all kind of this kind of thing.

CHAIRPERSON SOTTO: Steve, you're brilliant. Fantastic.

Greg.

MR. NOJEIM: The charge at line 48 says "Should the Privacy Act notices provided at the point of collection be revised to address Big Data in some way, including repurposing of data in source systems." Then at line 208 and 207, is that the response to this problem? There is very general language, and then the recommendation that this raises now, okay, if you really want to know how we're using the data, more specific uses or further uses,

you've got to go to this web site.

The comment then is a person is asked to give up the data under one set of rules and the rules could change; is that what you're saying?

Ms. McNABB: That's the problem with Big Data. That's the essential challenge, is at point of collection you account for future uses by some relatively general statement that ties you to whatever your legal authority is, but you can't be specific because you don't actually necessarily know precisely.

We're not saying that they aren't in compliance with the law. The general statement can make that work. But we think that if we can be more specific and timely as specific uses are developed, maybe by using something that can be amended more easily than a collection notice.

MR. NOJEIM: The problem is that, when I have to give up the data how useful is that?

Ms. McNABB: You have a clue, you have a clue. The notice requires that the legal authority and the limits that are within that legal authority be part of the original notice. So you have a clue. That's a problem out there in the world.

MR. PIERSON: The current context is that at the point of collection or the point of submission -- it depends on which program you're in, but at the moment in time that the data is going to DHS, you are receiving a static notice that is sufficiently broad to cover those activities that are known about and intended perhaps in the future. That's the current status of the program now. There is no notion of this make it a living

document, make sure that as times change, as technology changes, as agencies change, the threat changes, DHS reassesses the risks and reassesses what it is actually doing from an operational perspective to address the risks, but that the notice is still that static piece of paper or notice that was just posted on a wall five years ago or ten years ago or whatever.

So the change that we're recommending here is that there be a value add of this living notice, that when practically things change on the operational side or the technology side that there is the expectation -- and we used a lot of "should" language throughout the document. But what we're suggesting is that DHS takes a look at that and maintains step with the times, as opposed to amending notices every X number of years, that it really take a look at that and that it post the notice, publicize that it's posting notice changes, that it's posting FAQ's, information about whatever the new programs may or might be or the new technology that's being brought to bear, especially because the threat is just changing so much with time.

So when we look at what is being done right now, although it's legally sufficient, it is something that we think can be improved by a continuous living document.

Ms. McNABB: If new uses -- the new uses can't exceed the legal authority, and there could be new uses that would require changing the static notice.

MR. NOJEIM: But when I gave up the data I agreed to certain uses. When the rules change on me, I have no control.

Ms. McNABB: It's like the "and to fight terrorism." It

doesn't exactly say that anywhere, but there is that, and that's one of the purposes you gave it up for. But there could be more information about that that would be also appropriate for you to know, but it was different at one time than it is now. But it all fits in that bucket.

MR. NOJEIM: I didn't know it at the time I was giving up the data. DHS puts up new uses all the time, and the person who has given up that data, particularly when they may have to give up a lot of data to get benefits. So you're changing the rules on me.

Ms. McNABB: We're only contemplating in any of these recommendations new uses that are consistent with the law.

MR. NOJEIM: That are consistent with the uses that have already been identified to the person.

Ms. McNABB: And with the uses as allowed by law. Our understanding is that this system, that's the way this Framework is set up to operate. It's to tie the rules to the data. So that stays even if it's being accessed in a different way. So it's actually providing more information by being able to be more specific in the web site way.

CHAIRPERSON SOTTO: Did you have another comment?

MR. NOJEIM: No.

CHAIRPERSON SOTTO: Your tent is up. Go for it.

MS. WEINBERGER: Two things. One, to jump on the prior comment, in order to be able to move on to the next step. You are providing your data, so it's essentially implied consent, and that's how it's treated, so that we can use it for other purposes.

I don't know that you can give enough notification to tell people how their data is going to be used. That's the point of this living document.

Frankly, my next point is there's a presumption of literacy and audibility. We have to think about how we can use technology for audio enhancements, audio kiosks, so that people, if they want to, can select a language and be able to hear the notification, rather than we presumptively assume that everyone can read. It's very complicated language. I think it could be simplified.

Ms. McNABB: That's what we're intending on page 5, the section that starts at 182. It doesn't say specifically that, but that's what we're talking about on the notice on collection, that the notice on collection be provided in ways that make it accessible and comprehensible, which we don't say specifically put up an audio booth, but that would be one of the kinds of things.

MS. WEINBERGER: But I think we should think about advising on specific technologies, so that we focus on the technologies that are available and are in current use at these entry points. And there could be assistive technology that needs to be built upon to include these notifications.

Ms. McNABB: Line 190 is where we're there, except we don't say "assistive technologies." But we say "accessible to persons with disabilities."

MS. WEINBERGER: It depends how the notice is written.

MR. RICHARDS: Your issue, Marjorie, is literacy, which is not called out.

MS. WEINBERGER: There's a presumption of literacy.

Ms. McNABB: I don't think so necessarily.

MS. WEINBERGER: We talk about web site literacy. We're asking people to be able to know how to sign up to be able to get ongoing notifications. As we want to inform people, we want to make sure that we do it in a multimedia way so that we're providing as much information as possible. And I guarantee you, at some of these data entry points there are already assistive technologies that could be built upon to provide the notification you're looking for in other mediums.

MS. GATES: What if we just said "with disabilities or limited ability to read" or "limited literacy"?

Ms. McNABB: "That are accessible, such as using assistive technologies and making them available."

CHAIRPERSON SOTTO: Put a parenthetical.

Ms. McNABB: We could put it, "that are accessible, such as using assistive technologies, and make them available."

CHAIRPERSON SOTTO: We can say "to persons with disabilities," paren, "including." Would that work for you, Marjorie?

MS. WEINBERGER: I think so. I just want to make sure that we're not presuming literacy in the communities that are coming through.

Ms. McNABB: That's exactly what we're trying to get at in this section, and if you can tell us how to do it better.

CHAIRPERSON SOTTO: And there is discussion about immigration status, for example. That suggests, at least to me,

that this is about other languages as well.

Ms. McNABB: It says "other languages."

CHAIRPERSON SOTTO: It does say "other languages."

MS. WEINBERGER: That's why we need to think about different audio technologies.

Ms. McNABB: How about if line 190, which is on page 5, "Employ a variety of notice types, including using assistive technologies that make them accessible to persons with disabilities."

MS. WEINBERGER: Literacy isn't disability.

Ms. McNABB: "Disabilities or low" -- "low literacy level" or just "low literacy," "limited literacy"? That's good. And we already have different languages in the next bullet.

MR. RICHARDS: "Limited" what?

Ms. McNABB: "Literacy."

MS. WEINBERGER: Can we put in something about the use of assistive technology?

Ms. McNABB: It's there.

CHAIRPERSON SOTTO: It's on the screen.

MS. VANDERVOORT: Marjorie, can we put it in with the language line, as opposed to the disability line? Maybe I'm sensitive. So presentation in terms of languages as opposed to tieing it to disability.

Ms. McNABB: Language is down below.

MS. VANDERVOORT: I guess I'm lumping language and literacy together, as opposed to literacy and disability. Maybe it's me.

MR. ADLER: She's saying move the "literacy" down to the third bullet.

MS. VANDERVOORT: I think that concept is better addressed in 192, but that's -- maybe it's the sensitivity of addressing literacy and disability, as opposed to literacy and language.

MS. ANOLIK: I agree.

MR. PIERSON: Can we just do a separate bullet point? Let's just do a separate bullet point under that, "Employ a variety of notice types." Let's just copy and paste that and just mirror that bullet point by saying "Employ a variety of" --

MS. VANDERVOORT: Presentations.

MR. PIERSON: -- yes, "presentation styles, including the use of assistive technology for those of diminished or limited literacy," something like that. Just in a new bullet point.

(Pause in conversation; drafting on screen.)

CHAIRPERSON SOTTO: We're going to need to move on quickly to the Technology Subcommittee, but there are wording changes?

MR. NOJEIM: Yes.

CHAIRPERSON SOTTO: Okay, let's do it. Go ahead, Greg, please.

MR. NOJEIM: I'm formulating. Come back to me.

CHAIRPERSON SOTTO: Okay.

Linda, did you have something else?

MS. KOONTZ: No.

CHAIRPERSON SOTTO: Sharon.

MS. ANOLIK: Mine is quick and it is to the "Create a Living Document" section. I just wanted to suggest language to augment this with an example. So it would be to -- the concept is, in addition to having a web site where notices and changes can be mentioned, to have a place where the public can influence it, having an FAQ section, sort of a living FAQ section very similar to what the FTC does with a number of its pieces of legislation, that gets used as an ongoing guidance area.

So perhaps to line 202 adding, where it says "DHS can direct persons to its web site for subsequent questions, a living FAQ section."

Then I have one short add to line 208 as well, that would provide a way for the public to submit questions. So on line 208 it would read, "such as letting them sign up for ongoing updates, submit questions for DHS response" -- the idea here being if the public does not understand what we are trying to be clear about, this gives an opportunity for clarity.

Ms. McNABB: I think we want to say for DHS response in FAQ's.

MS. ANOLIK: Sure, yes.

Ms. McNABB: This is not come complain to DHS right here, because it's not going to work.

MS. ANOLIK: Yes. That way we make sure that the FAQ's that are there are actually the ones the public needs answered.

Ms. McNABB: Right.

CHAIRPERSON SOTTO: Greg, have you formulated your thoughts?

MR. NOJEIM: Not yet.

CHAIRPERSON SOTTO: Okay, we'll come back to you.

I'm going to turn, please, to Joanna. Would you please provide the committee with a summary of the Technology Subcommittee's paper?

MS. GRAMA: I will, and I'll try to speak as fast as humanly possible as well.

In January the Technology Subcommittee was asked to address auditing and oversight issues in the departmental Big Data project as follows:

"In developing audit capabilities in DHS Big Data projects, what specific activities should we seek to audit and how can we best build those requirements into the technology? What policies are needed to support the technology?

"Once the audit logs are developed, how do we use them in a meaningful way to ensure robust oversight? For example, should the audit logs contain the responses to queries or not? What processes should be in place for approving new or updated access controls? What mechanisms should be in place to ensure that these controls are not circumvented? Similarly, what mechanisms should be in place to assure access controls are not changed without appropriate oversight? What mechanisms should be in place to identify anomalies in the use of systems by individuals?"

To respond to our tasking request, the Technology Subcommittee was briefed on the Neptune pilot, the Cerberus pilot, the Common Entity Index. We also asked questions and received

additional information about current audit logging processes, access control systems, and data volume and volatility. In particular I'd like to thank Shannon and Jennifer Murray for their assistance to our subcommittee in getting us the additional information required.

Our review of the pilots and additional materials have resulted in our report and recommendations about audit processes and access controls and oversight that are before you, and I'm not going to read through those. Because of the technical complexity of our tasking, we formed two smaller working teams to address each of the issues separately. So one team for auditing and one team for access controls and oversight.

The key aspects of the audit recommendations ensure that the audit logs contain enough information or indicators to address system operational performance and efficacy, as well as enough information to investigate potential use anomalies.

The key aspects of the access controls and oversight recommendations ensure that access controls are in place to restrict access to the systems and information only to users with legitimate needs to use that data and who have received appropriate training, and that any changes to access control systems are recorded and reviewed for appropriateness.

We're happy to answer any questions you might have about our specific recommendations.

CHAIRPERSON SOTTO: Questions, comments?

(No response.)

CHAIRPERSON SOTTO: Greg, you'll have a lot of time.

Where's Greg?

MS. GRAMA: I will note that we will need to, in terms of edits, in the "Next Steps" at the end of the document we'll have to add the language that encourages DHS to implement the audit and oversight recommendations in the DHS Big Data systems as they move into further production.

CHAIRPERSON SOTTO: Good.

MS. GRAMA: Just copy the language from the policy document and, instead of "transparency," add in "audit and oversight," and I would be satisfied.

CHAIRPERSON SOTTO: Sharon, I think we have a comment from you?

MS. ANOLIK: Thank you. First of all, I really appreciated the detailed nature of this report. So thank you very much to the committee.

The one thing I saw that I wanted to ask a question about and see if maybe it's ripe for change, it's around lines 217 on page 5, where it says "DHS may consider allowing users to flag errors for capture in the audit logs." I think we would want to encourage a feature that allows for user reporting of incorrect information. That's certainly something our own users are going to be the ones seeing it. But I just wonder, if it's being captured in the audit logs and the audit logs will be reviewed really on a sort of batch or sporadic basis, whether that's really the place for it, or whether just creating some type of feature for capturing of errors or flags might be the more effective way to go. An error report of some sort.

MS. GRAMA: One thing we tried to do was, we were concerned that we were being very technical, so I wonder if we could capture -- the theme we had was finding errors. If we actually ended the sentence before "in the audit logs" --

MS. RILEY: Yes.

MS. GRAMA: So what if we ended the sentence at "for capture"?

MS. RILEY: Perfect.

MS. GRAMA: That would convey what you're saying.

MS. RILEY: I think that's a great idea, that in 217 and 220, just remove that "in the audit logs" clause and that would do it.

Sorry. 217 and 220.

MR. RICHARDS: "In the audit logs."

MS. GRAMA: And it may well be that DHS decides to implement something in the audit log infrastructure that tracks those.

MS. RILEY: It could be.

MS. GRAMA: But that way we've given you that much flexibility.

MS. RILEY: Exactly.

CHAIRPERSON SOTTO: Why don't we say "for example, in the audit logs"?

MS. RILEY: I wouldn't use that as the example.

MR. PIERSON: I wouldn't either.

CHAIRPERSON SOTTO: Okay?

MS. RILEY: Thank you.

CHAIRPERSON SOTTO: Good. Other comments?

MS. BALLARD: Barry?

CHAIRPERSON SOTTO: Barry, are you there? Barry?

MR. STEINHARDT: Yes. Can you hear me now?

CHAIRPERSON SOTTO: Yes. Hi, we can hear you. Can you hear?

MR. STEINHARDT: Yes, I can hear you.

CHAIRPERSON SOTTO: Perfect. Do you have a question?

MR. STEINHARDT: Really, it's a question for the panelists of the prior session. I don't know if this is an appropriate time to ask them or not.

CHAIRPERSON SOTTO: Donna, do you want to come up? Why don't you come on back. We'll take a few minutes to do that and wait for Greg to return.

(Pause.)

CHAIRPERSON SOTTO: The panel is assembled. So, Barry, please ask your question.

MR. STEINHARDT: First, thank you for the opportunity to do this, and thank you to the panelists for coming back.

My question is about redress. A number of you said that you would build in systems for redress, to give an opportunity for redress in the system that you were building. It's difficult for me to comprehend how you have redress when the data subjects can't know that they are under -- their activities, etcetera, are under consideration. Can you talk about how you build redress in under those circumstances?

MS. RILEY: I'll take the first shot at this from a

redress perspective. You are right that there will be systems that have exemptions under the Privacy Act that will not allow for access or amendment of information. But there are systems that, ESTA being one, where there are some exemptions, but the entire system isn't exempt.

I think of redress in this context as somebody makes -- let's use ESTA as an example. Somebody makes a request, an access request for their data, realizes there's an error in ESTA, requests an amendment that DHS complies with. We need a way to make sure that that redress carries through across into Neptune and into Cerberus, so that when Clark's analysts in I&A are making determinations they're doing it based on data that is accurate and complete.

So you're right that it's unlikely that any individual is ever going to be able to pull out and know exactly what the I&A or the CBP or any of our analysts are doing. But I think that makes it doubly important from my perspective that we have an ability to ensure that any corrections that are made carry through and in that way provide as adequate redress as is possible, even if some of that is invisible to the individual.

MR. SMITH: I was going to say the same thing.

(Laughter.)

At the risk of getting myself in trouble: Some of the times when we look at some of the information, we see -- for example, let's pick up SEVIS for a second. SEVIS is a student system. So as I understand the way SEVIS works, every college submits packages on the students. So if the students apply to

multiple U.S. colleges, they have multiple entries in the data set of SEVIS of all the visas that the multiple colleges have submitted on behalf of that student in the records.

Back to the point that Kellie was making, it may not go to the individual student, but there may be some point where we say back to SEVIS: This is the college that the person is actually at today; this is the record that actually represents truth, because the other submissions were -- the person did not attend that college. They actually went to this college. And those things may come across.

Also, if you look at some of the data, they're really fascinating data sets. Some of the addresses, for example, are at the hotel across the street from the admissions building. Those are the types of things that get entered in.

So some of this is back to the data quality issue. What I would consider part of this is also being able to say back to the organizations or say to the analysts: Hey, that data really is not -- while it's important, you've got to check it; you need to be really careful how you utilize that information. That's part of working closely. It's part of the policy process, which we did not highlight, but we're also working closely with the people, the folks who are actually experts on this data to understand the quality issues and what you can rely on and what you can't rely on.

So some of the things I'm giving you as examples are actually from my conversations with the SEVIS folks saying: Hey, wow, we've got duplications all over the place. And I go: Well,

yes, that's how this works; this is what we do; this is how we collect.

So that's been informing the rest of the community to be able to say, when you look at this data set this is the type of data you're looking at. This is the quality. This only represents what address they had when they applied to the school. It doesn't represent whether they actually came. So don't match this address to somebody else because that really doesn't make sense.

That's actually I think part of the redress. It's not redress officially, but it's part of the understanding of the data that you've got and how you can really -- what decisions you really can put on that data, what conclusions you can draw from that data. Just because it's the same address as something else you're looking at, that address, that really wasn't collected in a way that's authoritative. It's authoritative to when the student presented it, but it's not correct. You have to go back and validate it before you make any decisions at all.

So I think I'm trying to answer the question in the sense of, like Kellie said, it may not go back to the individual user or the person who submitted the data, but it will actually go back to improving the quality of the data so that better decisions are made and better outcomes both in the analysts' understanding of what it is they're looking at because they have a better understanding of the data they're being handed and the results they're being handed, and also back to the original source system to say, hey, this is what we've learned so far about this

information, this is really what you need to be looking at.

CHAIRPERSON SOTTO: Great. Thank you very much. Thank you for regrouping.

MR. PIERSON: I have a follow-up to that. I have a question. You guys recognize that --

MR. STEINHARDT: Could I just make a general comment very quickly?

CHAIRPERSON SOTTO: Barry, we couldn't understand that at all.

Ms. McNABB: He wants to make a general comment.

MR. STEINHARDT: I said could I make a general comment very quickly.

CHAIRPERSON SOTTO: Yes, please. But we have to end this discussion, so I'm going to please ask you to wrap up.

MR. STEINHARDT: This will be ten seconds.

It seems to me that "redress" is the wrong word here. This is really error correction. "Redress" implies that somehow the data subject has had the opportunity to weigh in in error correction. I would recommend that you use "error correction" or some similar language.

MR. PIERSON: That language is in our recommendations, actually. Those are in our recommendations, that you track errors. So with that comment, I'll retract my question.

CHAIRPERSON SOTTO: All right. Greg, if you could give us your proposed change, please.

MR. NOJEIM: What I am suggesting is language after the word "update," "that would inform them of how data they have

submitted are being used as a result of new programs."

I'll repeat it. After "update" --

Ms. McNABB: I guess it's before that, before the other thing you just added. "Ongoing updates that."

MR. NOJEIM: Yes, "that."

Ms. McNABB: And before the comma.

MR. NOJEIM: "That would inform them of how data they have submitted are being used as a result of new programs."

Ms. McNABB: I'm wondering about "as a result of new programs," because the Data Framework in a way isn't a new program. It's a new technological way for existing programs to do what they do.

MR. NOJEIM: "That would inform them of changes in how data they have submitted are being used." "Material changes." "That would inform them of material changes in how."

MS. GOLDSTEIN: I'm not comfortable with qualifying it with "material."

CHAIRPERSON SOTTO: Do you want to respond?

MR. PIERSON: Just a quick note, not to stop Greg's thing. The policies -- the privacy notices that are out there for the ones that we reviewed and all the rest, they are quite broad. They do encompass a lot of -- these programs aren't new uses. They're new ways to use existing data that gets rid of manual, 17 databases. It lets you go to one place.

But the privacy notices are extremely broad. I can't quote them, but include things like anything used to fight fraud, terrorism, cybersecurity, all the rest, terrorism, protect, defend

the United States. They are broad. So it's not like we're -- I personally don't feel as though we're going to come upon a new use and say, wow, that's not related to terrorism or antiterrorism or protecting the homeland.

The notices right now reflect those missions that are under DHS quite broadly and quite accurately when you think about it.

Ms. McNABB: That's why we wanted to have this supplemental.

MR. SMITH: Having the living document covers the, well, you didn't think about this one specific.

Ms. McNABB: "Inform them of changes in how data they have submitted are being used."

MR. NOJEIM: I would like "material" there. The thing is, it is a broad notice. They're giving broad information. But if you had known, would you have given it?

Ms. McNABB: And that's exactly what we're trying to say here.

MR. BRANDT: This is just me. It's not having to log in 17 times. Instead of logging in six times with my access, if I want to look up who Greg is and I go to system A and system C and system D, and I don't know how I write down, on a sticky note or whatever, the data, and then I now have the results of these three, I'm getting it from logging in once. It's the exact same access that I have, the exact same purpose, the same use.

MR. PIERSON: Yes, same use, better controls, better privacy controls, better auditing.

MR. BRANDT: Better auditing.

MR. PIERSON: And better kind of, quote unquote, "fraud, improper use alerting," we'll call it.

Ms. McNABB: The ways in which it's potentially new uses, it seems to me, is because it's easier -- or new user -- because it's easier to do a more comprehensive search, some users who wouldn't have gone to 27, would only have gone to two, now they're able to go to 27 or however many there are.

MR. BRANDT: So it's better.

Ms. McNABB: It's better. It's new in a sense, but it was allowed before. If it weren't consistent with the law they couldn't be doing it.

CHAIRPERSON SOTTO: What are we proposing? What's the final proposal?

MR. NOJEIM: The final proposal is: "That would inform them of changes in how the data they are submitting are being used."

MR. BRANDT: You're leaving out "material"?

Ms. McNABB: If it's a material change, isn't it arguably illegal not to have it in the notice?

CHAIRPERSON SOTTO: So what changes are we talking about, then?

MR. NOJEIM: We're talking about, the notice gives you a very broad notice, terrorism, and what I am hoping we can capture is, well, this is this new thing we're doing with the data that we didn't used to do, and we're going to put this in at this link. And that's what I'm trying to capture.

Ms. McNABB: And that's what we're trying to recommend.

CHAIRPERSON SOTTO: I think this is in keeping, yes.

MS. GOLDSTEIN: It sounds like it's more like added specificity, then; it's not really a change.

Ms. McNABB: It's added specificity.

MS. GOLDSTEIN: Added specificity, very different.

MR. NOJEIM: Yes, but how specifically is it going to let us know?

Ms. McNABB: It says "improved information," "improved more specific information." How about saying that, "improved"?

MS. MATTIES: No, It's included within what was described before. It's just more specific.

MR. NOJEIM: That would give them more particularized information about how the data they have submitted are being used.

Ms. McNABB: How about, look at line 207, "providing a static web site link which would allow for more specific information"? How about doing it right there, "more specific information on how the data they have submitted are being used." I love your plural. "Are being used," period, just like that, "more specific."

MR. NOJEIM: It's so general.

Ms. McNABB: Yes.

MR. NOJEIM: At some level, what you're saying is it's not a new use because we already said we can use it for whatever we want.

CHAIRPERSON SOTTO: Okay, Greg?

MR. NOJEIM: Yes.

CHAIRPERSON SOTTO: We have just a few more minutes. Let's fix the sentence. It's not an actual sentence.

Ms. McNABB: "The web site would also provide" --

CHAIRPERSON SOTTO: We need to fix the first sentence. No, no, no. The first sentence is not a sentence.

Ms. McNABB: It's not a sentence. Just "A static web site." Get rid of "providing." "Provide," yes, good. Provide, provide, provide, but it works.

You know, if we want to get rid of the "providings," we should squeeze it on line 208 and just say "the static" -- the web site should allow" or provide or whatever, because we're just saying in two sentences, we're describing the web site. The web site does this, the web site does that.

"The web site" -- I don't think we need "link" any more because we've mentioned it, it should be linked. "The web site would give."

CHAIRPERSON SOTTO: Now we can do "provide."

Ms. McNABB: Now we can say "provides" instead of "give." It's okay with "give." We'll have a little variety. We've got "provide" in the next sentence.

That works.

CHAIRPERSON SOTTO: Are we set? Greg, are you happy?

MR. NOJEIM: Yes.

CHAIRPERSON SOTTO: Good.

I let this go a little longer because we have no comments from the public, and we have no phone comments, either. Yes?

MR. ADLER: One thing we did in the last one, we talked about KPI's. We called them system metrics. But in our recommendation we called them KPI's. Can we just say "aka system metrics" to tie the two, that sometimes we call them KPI's and sometimes we call them system metrics.

CHAIRPERSON SOTTO: You mean in the other paper. In the other paper we did KPI's.

MR. ADLER: In the technology paper we called them KPI's, but in the policy we call them system metrics. I just think in line 28 we could just say "KPI's, aka system metrics."

CHAIRPERSON SOTTO: You're talking about the technology paper?

MR. ADLER: That's right.

CHAIRPERSON SOTTO: Okay, okay.

MR. ADLER: So we tie them and there's no ambiguity.

CHAIRPERSON SOTTO: Joanne?

Ms. McNABB: That's fine.

MS. GATES: The technology paper goes into it in great detail.

CHAIRPERSON SOTTO: It's very detailed, very detailed.

MR. ADLER: We should tie them together.

Ms. McNABB: You mean putting that in the policy paper?

CHAIRPERSON SOTTO: No, no, no. In the technology paper.

Ms. McNABB: Okay, good.

CHAIRPERSON SOTTO: Okay.

MR. ADLER: On line 28 on the technology paper, see

where it says "KPI's," right after that just say "aka, also known as system metrics"; "i.e., system metrics," something like that, just to tie them. That works.

CHAIRPERSON SOTTO: I will ask for last comments, please.

(No response.)

CHAIRPERSON SOTTO: Let me just note, with respect to the public and public comments, you may submit comments in written form to the committee at any time by emailing the committee at PrivacyCommittee@hq.dhs.gov.

All right. We're going to move to a vote. I'd like to move to close the debate on the two subcommittee reports. May I have a second, please?

MS. ANOLIK: Second.

CHAIRPERSON SOTTO: We have a second from Sharon. Members in favor of adopting the Policy Subcommittee recommendations, as modified here today, please say aye.

(Chorus of ayes.)

CHAIRPERSON SOTTO: Any against, say nay.

(No response.)

CHAIRPERSON SOTTO: Terrific. We have a majority. We have a consensus opinion here. We'll finalize today's edits. We'll send the recommendations, the final recommendations, to the Secretary and to the Chief Privacy Officer.

Members in favor of adopting the Technology Subcommittee report, say aye.

(Chorus of ayes.)

CHAIRPERSON SOTTO: That's as modified today.

Any nays? Anybody against?

(No response.)

CHAIRPERSON SOTTO: Terrific, great. We'll finalize this paper as well and submit these, these recommendations, to the Secretary and to the Chief Privacy Officer. Congratulations to both subcommittees and to Joanne and Joan. Brilliantly done, really, really well done. Thank you very much.

I know, Karen, you had a couple of final remarks. Please, come on up.

MS. NEUMAN: Thank you. I'm going to be extremely brief.

First, I want to thank the subcommittees for the obvious and thoughtful diligent work that you've done on the tasking. I really appreciate it. I also enjoyed and appreciated the discussion. I found it very informative, and we look forward to studying your recommendations. So thank you very much, all of you. I look forward to our continued collaboration.

Then a brief announcement. Today I am going to tell you that Shannon Ballard will be transitioning out of her role as DFO, after these years of incredibly diligent hard work. She's going to be a tough act to follow. I will be announcing her replacement shortly. So I just hope that you will join me in thanking her for her service and the work she's done to make these meetings possible for all of us.

(Applause.)

MS. NEUMAN: I said I would be brief and I am, and I

will turn it back to you, Lisa, and thank you all again.

CHAIRPERSON SOTTO: Thank you very much.

Shannon, we will miss you tremendously. But I hope you're not going too far.

Well, thanks so much to our speakers, to the committee members, and to members of the public for joining us at the meeting today. Thank you for participating.

This concludes the public portion of our meeting and we are very grateful for your interest and encourage you to continue following our deliberations and our work by checking our web page frequently. The minutes of today's meeting will be posted on the DHS Privacy Office's web site, again at dhs.gov\privacy, in the near future.

With that, the meeting is adjourned.

(Whereupon, at 4:53 p.m., the meeting was adjourned.)