



Homeland Security

DEPARTMENT OF HOMELAND SECURITY
DATA PRIVACY AND INTEGRITY ADVISORY COMMITTEE
FULL COMMITTEE MEETING
WEDNESDAY, SEPTEMBER 20, 2006
Transportation Security Administration
Town Hall
601 South 12th Street
Arlington, VA 22202

AFTERNOON SESSION

CHAIRMAN BEALES: All right. Welcome back to our afternoon session. Let me ask you again to please make sure your cell phones are turned off. We had a ring tone-free morning, and it was pleasant. And, hopefully, we can make it through the day. Again, bathrooms are down the hall to the right, if you're just joining us for the afternoon session. At the end of the day, the last item on the agenda is public comment from anybody who would like to address the Committee, and we would love to hear from you. All you need to do is go see Lane Raffray over here on my right and sign up. And please do that at some point before we get there.

This afternoon, we have two panels. Our first panel is to discuss data integrity. And I think what I'll do is to introduce the speakers one at a time, and then if we can hold questions to the end and ask questions of everybody. Our first speaker will be Jennifer Barrett, the Global Privacy Officer of Acxiom Corporation. She's responsible for public policy, privacy, and information practices as the Global Privacy Officer. She also provides direction for Acxiom's information use policy across all its global operations and for internal compliance. She's a speaker on privacy and customer relationship management and has published numerous articles and participated in writing books on these subjects.

What we're hoping to find out from this panel is how you people who really have to worry about data integrity for a living do it to see to what extent we can learn from that experience and apply it to the homeland security issues that we're looking at. So Jennifer, thank you for being here.

MS. BARRETT: Thank you, Howard, and the rest of the Committee. It's delightful to be here. I see a lot of familiar faces, and I'm happy to be invited back. I guess that means the last time I was here, I said something, so hopefully today I'll repeat it.

What I've prepared, and I'm happy to answer any questions you might have on this subject because, obviously, we deal with a lot of data in a lot of different ways within the business, is a little bit of how we approach the issue of data integrity and some of the things we have learned with a rather extensive study that we've done over the last six or eight years on the subject.

Obviously, the cost of bad data, inaccurate data, or unusable data to a company like Acxiom is very high. But it's also true, and some of our sales guys cringe when I say this, is that none of our data products are 100 percent accurate. That's a fact of life. No database in the world today exists that's 100 percent accurate, so the question becomes how do you balance and how do you manage those kinds of issues?

There's some common pitfalls that I think we need to recognize when we talk about this area, particularly because we're typically talking about automated use of data. Data quality, it doesn't seem to be that big of a concern, even though I'm not so sure we don't do some of the same things in the people world that we do in the automated world.

But the first is is that the data quality issues are often recognized too late in the system development process, resulting in kind of a fire-fighting mentality to the problem. They focused on the system, all the processes, how the data gets in and out, but not the quality issues as the data flows through the system.

So we start with a fire-fighting mentality and, because the system is already, you know, in some level of development or maybe even implementation, we'll never get out of it. And that's not the way we deal with it. Complex data dynamic environments for a company like Acxiom, and I think it's probably true for most companies or agencies that are actually bringing data in, putting them into some systematized process, and then using them for some decision making, have four key points of failure, and these key points of failure need to be recognized. In some cases, they're very, very low risk; but, in some cases, they can be very, very high risk. So it's appropriate to look at each of these points of failure.

The first is the source of the data. How often is the data coming in to the enterprise? There are varying degrees, and I'll talk a little bit about what we do with search data.

The second point is, in most systems, you are taking the data in the form it's coming in, and you are transforming it in some way into some standard representation that's used within the system. That transformation process can introduce errors into the

data, particularly when you're dealing with stream data or text data, as opposed to, you know, fixed numbers and characters.

The third is almost all data applications today have data that's integrated from multiple sources into some kind of database or data warehouse, and that integration process can introduce errors into it if data is not integrated when it should be or data is integrated when it should not be. So we're getting them both ends of the spectrum.

The fourth point of failure is in accessing the data. Have I retrieved what I'm really looking for, or have I gotten back in accessing that data something that's in error? We deal with all of these, and we deal with these across a vast number of products that we deliver both domestically and internationally.

What we did a number of years ago, because we kind of found ourselves in the fire-fighting mentality, even though we put a lot of time and energy on it, was we hired an academic, a Professor John Talbert. I have some material on what he is doing at the moment. He worked for us for a number of years and conducted a research project with MIT's Information Quality Program to actually study this and figure out what are the dynamics, what are the points that need to be looked at. They modeled it after kind of a total quality management program.

What came out of that, and I've got some brochures I'll pass around to you, is a data quality management program, and it's actually become a Master's Degree in science in conjunction with these programs. So I'll pass these out at the end of the session, if anybody would like a copy.

I know you are forming a data integrity subgroup, and I think it might behoove the subgroup to have Dr. Talbert come in and talk to you about the specifics of exactly what they learned through this process. We found it extremely valuable, and I think you will, too.

I want to share at a very, very high level in the short time I have just a little bit of those lessons, and I'm going to summarize this into three main areas that we found were absolutely critical. The first that there are four dimensions of data quality, some of which are probably pretty obvious but some of which are not so obvious.

The first one is the intrinsic aspect of it, and this includes the accuracy; the objectivity; the believability, which is sometimes we tend to overlook; and the reputation.

The second characteristic is the accessibility of the data, and this has to do with access and security and how hard or how easy is it to actually have the data to use in whatever decision process you need it for.

The third dimension is contextual quality, and this is the relevance of the data. Does it add value? What is the timeliness and completeness of the data?

The fourth dimension is the representational quality, which is the interpretability, the ease of use or understanding, how concise is it, what's the representation, and is there consistent representation of the data over time?

These all, some of these, I think, we all, if we'd say what are the dynamics, we probably all read six or eight of these pretty easily. But there are some of these that I think tend to fall in that category that we might not think of.

Lesson number two was it's critical to recognize that the data quality needs vary with the application. And this is kind of, this builds on the total quality management concept of fitness for use, which talks about -- and I'll give you a very simple example in our world. I'm sure you can think of lots of them that are appropriate for the homeland security arena. But if we are sending out a campaign for marketing purposes and there's a three or four percent non-deliverable rate. In other words, the address is bad or it's missing an apartment number or whatever, that's okay. You know, they can live with that. We're done sending out bills to that same audience. I cannot live with a three to four percent non-delivery rate because that person didn't get the bill, and I'm not going to get paid.

Now, that's an absolutely, kind of maybe the simplest example that I can provide you. But it shows you in the context of what the use of the data is going to be applied for, what the tolerance for error can be and the variations that can happen.

Lesson number three, and this is where I think a lot of companies and probably a lot of agencies fall short, is an area that we were not doing near as much in prior to Dr. Talbert's study, and that is that you can't improve what you cannot measure. And if you're in the tactical approach or in the fire-fighting approach of discover a problem, analyze it, and fix it, and then wait until the next problem surfaces and analyze it and fix it, you probably don't have a very good measurement process. And so taking a more of a strategic approach to this and saying, look, which is where a lot of the quality management concepts come into play here. You know, determine what your tolerance for error is up front, analyze it, measure it, fix it, analyze it, measure it, fix it. You know, it's that continuous loop. And it just puts a different perspective on it than dealing with somebody found an error so how are we going to resolve it? And in that mode, the adage of think big but act small or start small is always really good advice.

So what have we done with all of this? We've taken and created a data quality index for every single different product we have. We measured all of those four points of failure along the way, some of which we had made a conscious decision to say we don't need to measure this point because we've analyzed it and we feel like the chances of introducing error are very slim. But when we do measure it, whether we think that or not, we do measure it.

I want to just give you an idea before I wrap up on some of the concepts that we'd embed into that. Will the end-use of the data be by an experienced person who makes the ultimate decision or by a lot of people who may not be as well trained or by a system that has to say yes or no? Obviously, the tolerance for error, if it's by a person who's going to factor some human logic and maybe some other known quantities or known facts into it, can tolerate a much higher degree of error than a system that's going to say go or no go.

How will the unknown or not-sure category of the answer be dealt with? And I always like to use the traffic light example here. You know, we don't just have green and yellow, stop and go or warning or it's clear or safe or whatever the metaphors are that you want to use there. The big question is what do you do if it's red or, more importantly, how big is the yellow? And can you identify the yellow? Can you say, you know, I don't have enough data, the data is contradictory or whatever all the factors are to say I can't make a decision with this, or I have to do something else? So those are, you know, obviously considerations that had to be defined and analyzed up front.

You need to know about the quality of the data you're bringing in and all along the way. As I started out saying, there will always be errors. How you handle those errors when they're not caught, when they don't fall in to the yellow, is critical.

I think the last time I came for this group, we talked about redress and some of those areas. You know, your tolerance for errors may be higher if you have a really strong redress program. And I don't know that those two aspects of data quality are always coupled. If it's easy to fix whatever the bad data resulted in, then I may be able to tolerate a little bit higher level of it than if it's really hard to fix. So these are all factors that play into making a decision about is the quality of the data high enough for me to say let's go forward with it?

Lastly, I'd just like to suggest that Dr. Talbert is an excellent resource on this. He has been studying these for five or six years, and he's been back in academia now a couple of years, engaged a lot of experts across the country in this subject, and is a resource that I would highly recommend you reach out to. Thank you for the opportunity to come today, and I'd be happy to answer questions at the end of the session.

CHAIRMAN BEALES: Thank you very much. And I'm sure we will have some. Our next speaker is Linda Lopez, who is the Director of Customer Information Services for TransUnion. She works directly with the data furnishers, who provide consumer account information. She's been with TransUnion for nine years and is responsible for the overall leadership and data acquisition and data formatting. Linda?

MS. LOPEZ: Thank you. First of all, let me begin by saying thank you to the Chair and Vice Chair and the Committee for the invitation to be here with you today to speak.

I'd like to begin by sharing with you just some background statistics on TransUnion. TransUnion is a credit-reporting agency.

We update over 2 billion consumer account records a month. We work with approximately 85,000 data furnishers that are supplying us with information. Those sources range from financial institutions to mortgage companies, credit unions, student loan providers, and we even gather public record information.

Reporting is voluntary, and there is no fee to report information to TransUnion. We do, although, purchase public record information, and that information is judgment information, lien information, and bankruptcy information. And on our repository, we maintain both positive and negative credit data on consumers.

I'd like to share with you some of our processes and procedures that we have in place to ensure the overall integrity and accuracy of the data which is coming into TransUnion. It starts at the first point when a data furnisher contacts us wanting to report their information to us. We have a very stringent membership process, and we make sure that we do our due diligence to ensure that that data furnisher is who they claim to be, that they're located where they claim to be, as well as that they have permissible purpose to be reporting that information to us and utilizing that information.

Once they have passed that membership process, we work very closely with that data furnisher to ensure that they have a clear understanding of the industry reporting guidelines and that they understand the responsibility that, once they do begin to report credit information to us, how important it is to continue to update that information. So we become very intimate in the sense with the customer, ensuring that they have a clear understanding and we have a good familiarity of the type of information that they're going to be supplying us.

At the point that they're ready to, they've passed those checkpoints with us, we secure what we call a test file from that data furnisher, at which point we do a ton of verification and validation up front. We review their name formats, their addresses, that they're passing on consumers. We make sure that the terms on the accounts make sense. We make sure that, if it's a student loan data furnisher, that they're not passing us mortgage information. So we do all types of validation and verification at that point.

Once we're ready to load that information and allow it to process through our system, we're still able to get very specific customer statistics. And we get what we call informational statistics to let us know that that customer is passing us all of the required fields. It also gives us information on that file to let us know how many consumer records are not being updated. At that point, we're able to generate what we call an error-listing report, and we're able to analyze that information and go back to that data furnisher to let

them know what were the findings of our analysis and provide them with some sort of resolution and how to fix that record to ensure that it is updated to our repository.

Also on the front end, we also have established thresholds. And if those are exceeded at any point in time, for instance if the data furnisher provides us a file that, let's say, 50 percent of it is flagged to be deleted, the records are flagged to be deleted, that file, once it's processing through our system, is stopped. And at that point, it requires manual review from someone on my team to go in and analyze what is the problem with that file. At that point, we engage the data furnisher, because obviously they're the supplier of the information, to find out is that really a true representation of what they're trying to send to us, or did they make some sort of a mistake?

We also trend or do pattern recognition. We become very familiar with the data that our furnishers are supplying to us. For instance, if ABC Company supplies us typically 100,000 records every month, and this month we only received 10,000 records, that would be flagged in our system because our system starts to trend their data and says, "Wait a second, this looks differently." Or if, for some reason, this month all of the accounts, instead of having a mix of current accounts and delinquent accounts, this month the entire file is flagged as charged off. That, again, would be flagged in our system because of that pattern recognition.

At the point that the file is ready to be processed and the records are ready to be updated to our repository, obviously our goal is to make sure that we link that process record to the correct consumer file. And although we evaluate all of the data elements that are passed to us, I will tell you that, the vast majority of the time, we're able to update that record to the correct consumer file, just utilizing some key account identifiers.

For instance, I've had the same credit card for the last five years. My name hasn't changed. My address hasn't changed. All that's happening on a monthly basis is that my account is being updated to reflect that I've made a payment or that I've missed a payment or what have you. So as far as our matching process, typically, that's a pretty smooth process.

The last step of it, and this is on a very macro level, even once the data goes out there to our database, we still continue to monitor. We monitor our database just to ensure the overall quality. If we're seeing increases in new subjects or some sort of deterioration, we're able to go in and find out who was the source of that information and work with that data furnisher to find out what's happening.

An example of that, I believe this was about a year ago, there was some sort of a law change impacting bankruptcy, and we saw an increase in bankruptcy filings. Well, obviously, we went right back to our public record vendors that supply us that information, and they confirmed that there was an increase. And once we're armed with

that information, it just makes us feel more comfortable that what we're seeing, the changes that we're seeing on our database makes sense, and we're able to communicate that information to our furnishers.

So at a very high level, those are some of the things that we have in place to make sure that we're always trying to maintain the accuracy and the integrity of the information. We realize that we rely heavily on those data furnishers supplying the information, but we feel that we have reasonable procedures in place, from our vantage point, to try to continue to make sure that that data is accurate.

With that, if there's any questions at the end of the panel, I'll be more than happy to address those. Thank you.

CHAIRMAN BEALES: Again, thank you very much. It was very informative about how you deal with these problems and much appreciated. Our final speaker is Xuhui Shao, hope I was close, the Vice President of Analytics at ID Analytics. He joined ID Analytics in August of 2002. Dr. Shao leads the analytics team that's responsible for developing core technology and advanced analytical solutions for identity frauds. For the past ten years, he's been doing R&D and machine learning and computational intelligence both for the academy and for industry. Before ID Analytics, he was a key member of the Falcon Technology Team for credit card fraud detection and a lead scientist in the technology development group at HNC Software, that's now Fair Isaac. So welcome, Dr. Shao.

DR. SHAO: Thank you. Thank you for inviting me to have this opportunity to speak in front of this committee again. Data, obviously, have tremendous value and, therefore, mishandling data can do a lot of harm at many different levels. At ID Analytics, our business model focuses on identity risk. So, today, I would like to take this opportunity to share our experience about data integrity and the related issues from the identity risk management perspective.

ID Analytics is the leading solution provider for the identity fraud prevention and identity risk management, and we have been very successful working with a broad number of industries, including the banking, the retailing, the telecom, as well as some of the new members in healthcare utility, e-commerce, and direct-to-consumer business, to confirm this trusted network we call ID network. So over the years, we have built a tremendous amount of data in this data warehouse we call ID network that contributed mainly data through our clients, we call it IT network members, but also enhance with third-party datasets, for example from credit bureaus, such as TransUnion and Equifax, and they gave us their entire credit header files and also other third-party data vendors and also government agencies, such as Social Security Administration and so on.

That worked pretty well. We have just passed the 500 million ID score mark last month, meaning that we have screened 500 million credit applications and payment and other transactions since the inception of the company. This is done without sharing any piece of personal data, even to our network members. So I think this is a big distinction with our business model, compared to the business models of some other solution provider in the marketplace that we provide solutions that work very effectively. Obviously, it's a very healthy business model for our business, and we're out here to make money but without selling data and without sharing any personal data. That is a very important point. I may come to more about that later.

From my understanding, the data integrity is very important to the Department of Homeland Security, which is composed of a wide-ranging responsibilities across anti-terrorists, border security, immigration process, and passport, and so on and so forth. So I'll address later, but, before I go to the detail of some of those topics, I'd like to first define what is the data integrity and then what is the data quality issues.

My understanding is that data integrity is really the true nature reflected by the data for a given purpose. So I echo what Jennifer has said early on that when you look at data quality you have to look at what purpose you want to, what you try to do with the data, what the purpose of your solution or your application. And I gave you some examples, just because, for example, we're dealing with identity fraud problems. So just because an application is submitted by a fraud maybe stealing my identity and changed some of the attributes in filing an application, even though the individual behind that application is false, is fraudulent, but the data actually reflects the true intention of the fraudster and, therefore, is of value to us when we try to detect the behavior changes of those identity risk events.

So it really depends on, for example, on the other hand, if you want to create a dossier on somebody, then you obviously want to be very careful what data to include and what data not to include on this dossier. Another example is, for example, in credit bureaus, in the credit risk model, the credit risk management, the model is that all credit information about an individual should be rolled into one credit report filed by name and SSN, Social Security numbers. It is quite effective for credit risk management. However, it may not be if we are trying to solve the identity fraud, identity risk management, knowing that this credit report is, you know, very high credit worthiness is not going to be useful if we cannot be sure the person we're dealing with is who they say they are.

In the identity risk management arena, from our experience, it's far more effective to build a broader model that studies the relationships between different risk events and the identity information associated with those risk events and study the patterns of those relationships. In this case, the integrity of data is to really preserve the source of information and not to throw away any information.

So, next, you know, I want to talk about the definition of data accuracy and data quality. There are different sources of data inaccuracy. The example I gave early, when a fraudster files a credit application, even though the information is being manipulated, data being manipulated, but it reflects the true intention of the fraudster. Therefore, it is still of value if we try to prevent identity fraud.

On the other hand, for a normal consumer, we know that many people move throughout the year or change their job or change their phone number or change their address. Therefore, when we look at the historical record, we may see discrepancies between their residence, between their employee, employers information, or phone numbers, and so on. So in this case, not only the static information but also the timeliness of the information, timeliness of data that really defines the quality of the data. And, of course, if the source of accuracy is introduced by the system, for example typos, when people entering the information or insufficient techniques in normalizing the data, then, of course, these are the source of inaccuracy we should try very hard to normalize, to clean out.

So the next point I want to make is that the data quality and integrity also depends on the analytics technology and the model behind it. So analytics exist in data mining and machine learning and risk management, and we do that with identity risk management and other people do it with other different purposes, such as credit risk management and fraud detection and other areas.

Now, there are models that are more sensitive to noise and outliers, and there are models that techniques can be developed that are less sensitive to the noise and outliers in the data. For example, simple linear regressions or the old fashioned networks might be more susceptible to the noise and outliers in the training data, but some of the more advanced machine learning techniques we're looking into and that's also being studied by researchers in this space, such as support vector machine or Boosted Tree models. In various applications, they are more robust against noise and the data. So the point I want to make is it's not just the data itself but also the algorithm and the analytics solutions to deploy on the data that makes a difference.

And we, obviously, we not only get data contributions from our network members, but we also purchase commercial data sets. So, obviously, we're facing the same problem of trying to validate the quality and the effectiveness of the data. And from our experience, it's very important to validate the quality of data in a network, data-sharing network setting.

In our ID network, we have the capability, we have the data, as well as the technique, to cross- validate a different source of data. For example, one credit bureau may be more accurate on the adverse information, and maybe a different credit bureau, a date of enter is more accurate on the phone number information. And sometimes a data

vendor has broader data contributions, but another data vendor on the phone number and address information have more timely updates to the information.

So through this network of data sources, we're able to optimize and evaluate the effectiveness and the quality of data before we even purchase it. And after we purchase it, we continue to monitor the effectiveness on particular applications. And so for example, if you want to match recent movers, you want to validate their address and phone information, you don't want a seven-year, ten-year history, but you want more up-to-date recent updates to their address and their phone information.

The next point I want to make is the value of data sharing. We strongly endorse the idea of sharing information, and we believe the data integrity issue can be managed through a trusted network. For example, we work with multiple companies across the industry. We work with the data vendors. And through the sharing of data, there's a subtlety there that I mentioned early on: even though we share a lot of data in the ID network, we put it into a highly-secured environment, and we never expose the data. We never sell them. We never even share the data back with our members. And that is important. I want to separate the usage of metrics and token from the usage of personal data.

For example, ID score is a metrics of the likelihood of an identity being involved in fraud or the riskiness of a piece of identity. Having this ID score, even though -- for example, if someone get the ID score on an identity, it's worthless and useless to the fraudster. On the other hand, the personal data, if we don't secure them and it's very hard to, it's very valuable to individuals and it's very hard to or even impossible to replace. On the other hand, a metric or a token, it's very easy to replace and worthless by itself.

So we can solve the problem without sharing the personal data, without exposing the personal data but sharing it in a highly-secured and trusted network environment. Because of today's world, there's no single source of truth, and personal data, such as Social Security numbers, are not secure and confidential at all. You can easily obtain these personal data with little cost or, you know, sometimes even free from many different sources. And so it is very important to share but without exposing information to harm.

And also to share data with a clearly- defined, a strictly-defined purpose, in our experience, we have, in our ID network for example, close to 400 million identity records. But, you know, obviously, there's a lot of usage of the data, but we limit ourselves to focus exclusively on identity fraud, identity risk management. We will never use the data for any other purposes, and that's why we gain the trust of the members that contribute the data. So I think that's also very important.

The last point I want to make is that, obviously, across different agencies, for example in the Department of Homeland Security, there should be broader information

sharing, broader data sharing, and one of the questions, I think, in people's mind, how do we integrate those data and then further insure the integrity of the data in the process. So the point I want to make is that you don't want to lose track of the source of the information, source of data. When you integrate a data, you want to keep track of the data sources. Therefore, you can update and correct any mistakes or inaccuracies from that data source because when you merge, you may run the risk of losing the information if you lose track of the data source and the purpose of the data collection.

In our experience, even though our ID network composed of many different data sources doesn't have a contributor to the network, as well as third-party data we acquire, we put them together in a network but we always keep track of the data source, the purpose. Different data sources may have different levels of confidence and completeness and timeliness of the data, so it is very important to keep track of them separately and preserve the information and preserve the overall structure of these data sources. And, therefore, we can study the relationships and the integrities using analytical techniques. And we can observe the patterns across these data sources, and that, to me, defines the integrity and the quality of the data.

So that is the points I want to make, and I would be happy to answer any questions.

CHAIRMAN BEALES: All right. Well, thank you all very much. And although I have too many questions to know where to start, so I will start with someone else. Joe Alhadeff?

COMMITTEE MEMBER ALHADEFF: Thank you. This was touched upon mostly in the last presentation, but it was a question that I thought of during the presentation for Ms. Lopez, so I guess it's for anyone to pick up. But the concept of having 85,000, I think was the number, of furnishers of information that you were talking about, which led me to try to think of how do you maintain source information because, at some point or another, if there's a unification of that information in the records, there's got to be duplication among the furnishers of who's giving you what. And if you look for correction of that information so that you don't continue to sweep it in, how do you keep and maintain and understand where the source was?

We heard from the last speaker they maintain a second record of source so, even in the combined record, they don't lose that. But I was just trying to figure out if maybe you or any member of the panel could address how those source ideas are kept separate so that you can actually understand the truth of the originating information if it may, in fact, be flawed.

MS. LOPEZ: I'll try to answer that. Every data furnisher processes their file. They have a unique program that identifies their information. And without getting into too

much detail, because some of this may go into some proprietary information, but those unique IDs that tie it back to that data furnisher. And I'm not sure if I'm answering your question.

COMMITTEE MEMBER ALHADEFF: From any single data furnisher, I understand how you understand where that data came in. But at some point, are you creating a unified record of the data? And if you are, do you use meta tags to then map back to where the source of the data was? It's the mechanism, which if it's proprietary I understand that you can't discuss it, but it's the mechanism of how you manage source once you start to unify the record. Or is every record 85,000 potential data points?

MS. LOPEZ: No, every record is not 85,000 potential data points, and I really can't get into the other part of your question. I apologize for that.

MS. BARRETT: Do you want me to talk about it from an Acxiom perspective, which is not certainly 85, we don't have 85,000 sources. But we have several thousand, so it's a little bit of a complex issue. We deal with it in a couple of different ways, depending on the application. In some cases in the application, you want one representation of the data and you need it to be the most accurate or the most current or the most complete. Again, it may vary with the application. And so we would expect data from various sources with rules, sets of rules, when we're bringing the data together to say this source has been rated as more current, for instance, than another source. If I have a conflict in data between two sources, and one of my objectives in this application is currency, then I'm going to take the data from that source. And we do use meta tags to track that data as it comes in.

In other applications, though, the data is actually stored in all its original source structures but linked in such a way that when it's accessed, we scan those structures, and we can apply rules to that access a little bit, maybe along the lines of the last comments that were being made or the algorithm that says when I'm making this kind of query into the database, here's the rule set I use to create, if you will, a composite record on the fly that may be created differently based on the inquiry and the need. Because sometimes, again, if I'm looking for history, old addresses are really important. If I'm looking for where you live today, old addresses are not important. So, again, it depends on the application. Does that help a little?

DR. SHAO: Let me make a few comments. Obviously, we do that. We keep track of different data sources. I want to give a little bit of a reason why we do that because, to us, data quality is both statistical, as well as temporal. You cannot say, just give you one phone number, the quality of that phone number. You have to validate across a data set, and that's why the only way to keep track of which data set, the quality of it, is to keep track of source of it. Once we've merged them, if we don't keep track of it, we lose the ability to discern the statistical quality of the data set.

And, also, the quality of data set is also a temporal. Like Jennifer pointed out that older data may be less useful if you want to verify the current information that's constantly changing.

One other example is that we're currently working with VA, as well as Department of Transportation, in that data breach and vindication of the largest and second largest data breach of the government. And in order to do that, obviously, we have to keep track of these data sets very separately and to measure the harm index on the data set as a whole. The harm index is defined as how many identities are being misused and the estimate of the financial impact on each identity being misused. So the only way to measure not only the quality, the integrity, also the exposure to risk and fraud and to harm, is to keep track of them as a statistical set.

CHAIRMAN BEALES: Neville?

COMMITTEE MEMBER PATTINSON: Thank you, Chairman. For Jennifer and Linda, as far as the information that you're collecting, you talked about the quality and the integrity and coming at what index and so on, what checks and balances are put in place for the data that you hold and maintain, as far as any manipulation or covert modification while they're in your possession? Obviously, the reporting that you're doing is giving value to some client at some point. If somebody can manipulate the information while it's in your systems in such a way that it's undetectable and perhaps up to their detectable checks and balances there. So it's really how do you maintain the integrity and then the consistency of that integrity over the clearly daily updates that are going on?

MS. LOPEZ: To answer the first part of your question, we don't manipulate the data in any form or fashion. We don't want to get into a position. We really hold the data furnisher accountable. They're the ones that have the relationship with the consumer, so they know that account better than we can. So once we receive that information, we really don't do any manipulation on that data once it's on our repository.

COMMITTEE MEMBER PATTINSON: How do you know it's not manipulated?

MS. LOPEZ: I'm sorry. What was your question?

COMMITTEE MEMBER PATTINSON: How do you know it's not been manipulated in your possession?

MS. LOPEZ: I don't have any way to manipulate that data. Our information security people would probably be able to address that, but, as far as I know, once that information comes in, we don't do anything to manipulate it and change it in any way.

CHAIRMAN BEALES: I think that is the question, right, Neville? What do you do to keep a hacker from changing it?

COMMITTEE MEMBER PATTINSON: Or an inside quark.

MS. BARRETT: Obviously, it's a security issue, from our perspective, if I understand the question. The issue of access, you know, inappropriate access, goes the full gamut, from somebody using a legitimate access or something to do something that they shouldn't, versus someone actually hacking through whatever your safeguards are to keep, you know, inappropriate people from accessing it.

From the legitimate access standpoint, if it's a system that involves sensitive information, we track all the queries. And obviously there's multi layers of password, access, and protection, but it probably is not going to prevent someone from going in and getting access or doing something inappropriately if they're a cleared person that has legitimate access and has a legitimate password.

Most of the products that we use, and I think this is true of other data vendors, you can't update the data in the product. It is a retrieve-only kind of application. So by building in retrieve-only applications, you eliminate users from having update capability. And so you narrow the scope of who has authority and responsibility for actually changing the data. I hope that helps a little bit.

MS. LOPEZ: Yes. And if I can answer that, although I work very closely with that information coming in the door, I have no maintenance capabilities. The only folks within our organization would be the individuals that are handling consumer disputes, and it's a little bit outside of my area of expertise. But, obviously, if a consumer is disputing a record and, once it's verified, a change has to be made, there is a select few individuals that do have capabilities to change that information on our repository. But myself or the folks that I work with, with the data coming in the door, we have no way to manipulate or do any type of maintenance once it has come in.

MS. BARRETT: One other thing, and this is getting kind of granular, but it's probably worth mentioning because it's one of the techniques that we kind of eluded to throughout all the discussions is statistical analysis of the before and after of any kind of sanctioned update that takes place. Linda discussed, you know, recognizing there were too many bankruptcies and that was out of the norm of the number of bankruptcies that typically come in.

So I think all systems do some analysis of the database before and the database after. And in some cases, that might uncover, it wouldn't necessarily prevent but it might uncover the fact that a database has been tampered with and so forth.

CHAIRMAN BEALES: John Sabo?

COMMITTEE MEMBER SABO: Well, it's a pretty complex subject for short Q & A and for the briefing, but it's really interesting. So focusing away from integrity and a lot of issues around that and bringing it into the world of DHS and PIAs, you know, from a public policy perspective, the programs that DHS administers have huge consequences. I

mean, so do credit reports for getting an auto loan, but nothing like what we're talking about in terms of liberty and impact on freedom of movement and privacy and fundamentals.

And looking at a lot of the programs, like screening programs and existing systems records notices, what we typically see would be the data elements that are collected: name, SSN, DOB, etcetera, etcetera. Maybe a set of 16, another program may be a set of 15 or 10.

But the point Jennifer was making, and I think Mr. Shao talked about this, as well. These dimensions of data quality, particularly contextual quality, meaning relevance, timeliness, and completeness, I'd like to get your views about that from the perspective of a governmental, the necessary for both the data provider, that is you guys, as well as the government, to more clearly think through those specifications.

For example, the timeliness of the data not with respect to how quickly you get it to the government but the timeliness of the data with respect to how current it is, versus historical, would clearly impact the utility and the value and the effectiveness of that information with respect to a decision made by an algorithm. And I'm wondering if, when you work with customers, you have a framework or guidance for saying to the customer, "Well, we can help you with this, but if you're using it for purpose X, then for purpose X, it really should be within X parameters of timeliness." I'm just using this as an example. And so we would recommend this type of data controls around the use of the data.

So that's part of it. Do you have a framework for working with customers to help address those questions and, B, what would your views be about having a privacy impact assessment or a systems of record notice or some other documentation around those parameters for use in a particular programmatic system. See the distinction I'm making, as opposed to just a clutch in the data elements that are used in any random way. So I guess I'm looking for some views about what you do today and frameworks, and, B, what you'd recommend government do with respect to data quality, as you defined it, Jennifer.

MS. BARRETT: Well, I think you hit on a very important issue here, and that is not all data is appropriate for all applications. But you don't know that if you can't measure the quality and the timeliness and the completeness of the data. We do play a very consultative role with our clients about data from us. In some cases, we say we won't sell you data for that application. And so it's important that a data vendor always understand the application that data is going to be used in or they will never be able to do a good job of advising the client, whether it's a government agency or private sector company, on what they should or shouldn't use the data with.

I mean, the hard answer to that question is, in most cases, the data is not quite as current, not quite as complete, or whatever those other metrics are, as you would like. So that's when you kick into is it so incomplete, is it so out of date that I can't use it all and have to look for another alternative to solve whatever the agency's security problem is. Or are there ways that I can litigate against that?

Sometimes we go through tiers of decision making. Let's say it's a timeliness issue and I'm trying to validate, for example validate a telephone number or an address or something, and I know that my data services, the primary data source I use is only updated every 30 days, but people change their phone numbers every day. Well, you know, some of this factors into cost, too. Sometimes the government doesn't have some of the same cost constraints that the private sector does, but our private sector clients certainly have cost constraints.

The data that's 30 days old may be priced at a lower price than the data that's 30 minutes old or 24 hours old. So the question is if I can't validate something with a 30-day old data, then is there a second way to go use a second data source to validate it, you know, in the 30-minute or 24-hour mode and only pay the premium when I need to? But, again, there's not a simple black or white answer to it.

COMMITTEE MEMBER SABO: No, but I guess what, that's where I was going, that, you know, either through meta data of some kind and/or other controls, would you see this, in order to get to where you were suggesting with respect to data quality, that these are types of, I don't know what you'd call them, specifications that would be almost mandatory for very sensitive programs? That is, you can't just send in the latest DOB, you'd have to have a specification that's developed based on some standard. I guess that's where I was looking for some advice.

MS. BARRETT: I think from an advice standpoint for government agencies that actually ranking all of the sources or having a source tell you in some measurable form what is the quality. If you get an answer of, "I can't tell you that," you may have a problem. And we get a lot of sources that say, "I can't tell you," and we say, "Fine, give us the data and we'll tell you," because we know how to measure it ourselves.

So it doesn't mean to say you shouldn't take a source, but you should do your own measurement on it. I'm not sure that there shouldn't be a requirement for data quality index on all sources and certainly sources that are used for sensitive purposes or that have serious implications to an error in the data, but that shouldn't be part of it. I don't think that that's necessarily something that would go in a privacy impact assessment, other than to say I've done it. I don't know it's something that you publicize. We certainly don't publicize the sources, but I think it's a really good approach to take to try to understand and then mitigating whatever the errors are.

CHAIRMAN BEALES: Lisa?

VICE CHAIR SOTTO: Thank you. This is a really interesting topic for us, and I have to say I will confess really complete ignorance in how DHS deals with data integrity, and I think we're really looking to explore that in some detail and so much appreciate your giving us our real great initial briefing.

Jennifer, you said you can't act to fix inaccuracies unless you can measure them. We have a funny problem here. I think there's a real quandary in that we can't, there's no access to the data by individuals who are in the terrorist watch list, for example, so there's no self reporting of errors. The names that are in the database are all anomalous really because they're anomalous people. Otherwise, they're not mainstream people and they wouldn't be in the database. So how do you do, how do you measure anomalies of an anomalous list, and I'm sure Dr. Shao can tell us that.

But it really is sort of a different animal, I think, in some ways from what you're dealing with. And as John said, the consequences of ours are severe when you're dealing with a watch list, as compared with a marketing list or even a credit file that has inaccuracies.

So the essence of my question is how in the world can you measure inaccuracies in a watch list where people can't access it and where there's very limited access, even by government employees?

MS. BARRETT: I'll make a stab at it, and then everybody should maybe throw in their two cents' worth. You've got to start somewhere, and I think paralysis because I can't get it perfect is sometimes, you know, what causes us not to even take the first step. So, you know, there are ways to measure what you've got and how complete it is, which you've got to step back first. What's the problem with the inaccuracy? Is it the fact that it's -- let's take a watch list for a theoretical example, a watch list. I have a last name, but I don't have a first name. What's it worth? What are you going to do with just that last name? There are things you can do with just a last name, but they're not the same things that you can do with just a last name as with a first name and a last name and a date of birth.

So I think it all goes to taking the data you have, accepting what is there and what is not there. Validating to third party sources is always wonderful or having somebody tell you it's wrong is always wonderful, but that doesn't mean you still can't do some measurement and some indexing of it without those two things that begin to let you understand, put a stake in the ground. And with any quality program, I don't care whether it's data or anything else, you put a stake in the ground and then you decide is the stake where it needs to be or does the stake need to be moved? And experience over time, how many do complain, even though they can't see the data, that somebody goes

back and looks at it and says, "Oh, this person ought to be on the white list," even though we have something similar on the black list or whatever." That analysis process teaches you something, but if you don't have a baseline that you're learning from, then you're shooting over here and you're shooting over there and you're shooting over there, you never make any long-term progress.

DR. SHAO: Well, let me add a little bit more comment to this. Obviously, we're not in the business of anti-terrorism, our firm is not. But we do believe there's a lot of identity risk, identity fraud in this area. I testified last year about this point, as well.

So I think I want to present two things for you guys to consider. The first thing is we can measure the identity risk. We know how to do that, and we have demonstrated it can be done. So if we can eliminate the identity risk, the false identities, if the identity risk in this, I'll say the watch list, then we will make improvement over time on this problem.

The other thing is we can certainly measure the false positives. You know, whenever you have the watch list, the things you're going to observe a lot of times, most of the time even, are the false positives. You match someone, but they're actually a false match. There is a false positive or false alarm. And then we learn from this experience and record these and then try to improve the false positives by either incorporating those incidents as data elements so we know that, oh, this, I don't know, John Doe is not a 20-year-old so and so; he's more a 40-year-old and so on. Or we can improve the algorithm that deployed as part of solution in order to eliminate or reduce a false positive. We have demonstrated that advanced analytic solutions can dramatically reduce the false positive. Of course, this is from an identity risk management arena.

And the other thing that I want to throw out is that when we work with -- first of all, we are not a data vendor. We work with data vendors to create this ID network, and a lot of data vendors may not keep track of how the data set changes over time. So what we're actually doing is time stamping every record that comes in the door when we see them and when they happen. That does two things. One is that we can keep track of how data sets, how data elements, how identity evolves over time. And the second thing is we can keep track of any quality issues, or we can look at the window that certain data elements are valid. If you don't sunset the data elements, what it will do eventually will be creating false positives.

CHAIRMAN BEALES: I would just comment because I guess in a lot of the DHS applications, obviously there are serious consequences, but the consequences also differ across applications. And that I think is important in thinking about this problem. I mean, if the consequence is secondary screening at the airport, Jim actually enjoys that, and it's very different from if the consequence is, "I'm sorry, you can't be a truck driver anymore, or, "You can't work in this port." And errors are much more important in one place than

in the other. But, obviously, a lot of this stuff does have important consequences. Joanne McNabb?

COMMITTEE MEMBER MCNABB: Thank you. I have a question for Dr. Shao. I have a sort of general layperson's understanding of what your system does, but I'm very curious about how you are analyzing the breached data on VA and the other one. How you can keep track of the source of the elements when the key identity elements are Social Security numbers and names, which are present in both? Tracking them out in the world, how do you know where they came from?

DR. SHAO: I can give you quick summaries of this approach, obviously, not to the detail that would reveal a trade secret. We get a data set that's been breached or potentially being breached. You know, there are some private sector data breach and analyzing the paths, and then we're also working with VA and also Department of Transportation. What we do is that we use the identity elements in the breached data set to link to our ID network. In our ID network, we have hundreds of millions of risk events. A majority of them are credit applications submitted through our clients in different industries, and also there are payment activities and other kinds of transactions submitted to us by our clients. So these are the true activities that are happening on a daily basis, on a minute-by-minute basis. And then we can observe any suspicious usages of these identities.

For example, a totally unrelated identity in the breach data set should not be linked in a particular, in a very high velocity or high-density fashion through let's say credit applications or sharing pieces of an address or phone or other piece of a personal data from two very unrelated individuals or multiple unrelated individuals over time. So that is how we can assess the potential misuse and also the size of the misuse. Therefore, we can calculate the harm index on each breach.

COMMITTEE MEMBER MCNABB: Since the key elements in the breaches are names and Social Security numbers and if you see them associated in ways that suggest fraud with various transactions, other than the time connection to the breach, how do you know that that Social Security number got out there for that fraudulent use from that breach and not from some other place?

DR. SHAO: Yes, again, I think it's a statistical measure that we cannot say a number that can be gotten or obtained, a piece of information obtained from different sources. But when we look at the breached data set as a whole and we can assess the statistical exposure and also the unusual anomalous –

COMMITTEE MEMBER MCNABB: Of a number of the different identities –

DR. SHAO: Yes.

COMMITTEE MEMBER MCNABB: -- associated in the breach? It's that –

DR. SHAO: Yes. Statistically, the percentage of identity elements being used compared to the normal general population.

COMMITTEE MEMBER MCNABB: Thank you.

CHAIRMAN BEALES: Richard?

COMMITTEE MEMBER PURCELL: Thank you. I have a question. I'm guessing that, Jennifer, I'll ask you to respond first, of course. Please, if you have a response, I'd like to hear it. But we've been working in our subcommittee and have published one paper, a policy paper, on ways to utilize commercial data to drive down false positives and policies around that activity. We're planning and developing currently a higher level, general policy framework for public agencies' use of commercial data, a very, very important part of what we've been asked to do and provide advice for the Department. But we have a problem, and the problem really comes down to how do you define commercial data? It's very fundamental to answering the question. And so I'd like to ask your interpretation or your definition, if you have one, for commercial data, as opposed to publicly- available data perhaps or public records data. I mean, there are at least three large sets of data, but the bright lines between them are not obvious by any means.

MS. BARRETT: I guess I might answer that with a question, and that's why is it relevant that you create the distinction? Or is the relevancy data that the agency does not have, no matter where it comes from?

Now, there may be legal requirements that are put on commercial data or public record data or different types of data, and certainly those need to be recognized from a standpoint of what you can do with them. So that may be one of the reasons why you need the definition for commercial data. But if you step back and look at it from the agency's perspective, they need data from somebody else, whether it's another government agency; a public record; a commercial entity, like Acxiom or anybody else that sells data; or directly from the consumer, if that happens to be a source.

So each of those sources or types of sources of data may have a set of requirements to place on it, but a definition of what's commercial, I'm not sure I can give you a definition of what's commercial data and I'm not sure I understand why you need one, I guess, is the bottom line.

COMMITTEE MEMBER PURCELL: Our concern comes out of some activities we've observed, the acquisition or the requirement or the demand or the request, however one wants to characterize it, for airline records that, for analysis purposes, to track flights, passenger name records, that kind of thing.

Also, the elements of trust that we believe are necessary between citizens and federal agencies and the transparencies necessary to underscore that trust or undermine that trust in some cases, what right would -- I guess part of it is what reasons government

have for access, but also those reasons can be different. What's the difference, as an example, between a government agent typing a search query into Google and that same government agent going to a major distributor of gasoline and asking for all transactions from the last six months for purchases of gasoline? Is there a difference between those? Is it definable?

MS. BARRETT: I don't know that it is. I think -- let me tell you how we approach it from a corporate perspective, and maybe that will shed some light on it from a homeland security perspective. We don't look at it in terms of classifying sources. We look at it in terms of what's the application we're going to put the data to, and then we work back from that. We look at the quality required for that application, we look at the legal restriction relative to that application, and we put a subjective judgment on the acceptability, the public acceptability of use of data in that application. And then when you start looking at the individual sources, you have to understand what the situation the source is in and under what permission, and I'll use that term very loosely, the source is giving it to you. Either they notified the consumer this might happen; this is the law that says this is going to happen, in the case of the credit reports. The banks don't say, well, maybe they do now after GOBA, I don't know, but they didn't used to say, "We're going to give your data to the credit bureau," because the law said they didn't have to.

So you have to go through that whole what's a legal permission and, in some cases, what is an acceptable use of the data. And in the end, it becomes a judgment call and, from a commercial standpoint, it's something of a soft judgment call. And sometimes you can call it right and everybody's real happy and sometimes you're in the gray area and people complain. But the way you started this was saying classification of commercial data. I'm not sure that's how we would approach the problem.

DR. SHAO: Maybe I can add a little comment. Not directly addressing your question, but addressing, I think, the underlying concern you have. You know, when you're making a serious decision, you want to know what the underlying data source is that supported the decision and what is the consequence if the data source is not reliable or as reliable as some of the other data sources.

From our experience, for example, ID Score is used to rank order the risk of identity fraud, but we don't use ID Score and we don't recommend to use ID Score to make a direct decision. Rather, it is about a risk management. So given your business model, given the purpose of the agency, and given the amount of risk and the cost and how do you want to best, given the limited resource, manage this whatever application it is, you know, our case is identity risk. And so if we see a higher risk supported by the data, then we simply take or recommend the analyst to take extra steps to further validate this application.

So if we just say, okay, let's reject everybody below or above certain scores, then it has grave consequences to the consumers. Rather, we focus on how to help the consumers to establish a positive identity. So by taking extra steps, we can assure that the accuracy and the robustness of their identity, and then they become paying customers and next time they apply for credit again, apply other type of enrollment or applications, the process becomes much smoother and much quicker. So I think if you focus on the risk management, rather than try to make a black/white decision, you have more power and more flexibility.

CHAIRMAN BEALES: That, too, is an interesting idea that occurs to me because there may be circumstances where DHS can structure programs in a way where the consequence is relatively inconsequential, we want more information, rather than something serious, and that may be an attractive way to try to structure some programs where there could be bad consequences. It's to try to create a middle ground. It's not yes or no; it's yes, no, and tell me more.

DR. SHAO: Yes. In the end, you want to help the good guys.

CHAIRMAN BEALES: Right. Ramon?

COMMITTEE MEMBER BARQUIN: First of all, thank you very much. It's been extremely, extremely interesting. What I wanted to ask if we could touch on the audit mechanisms, both internal and given the, like HIPPA and SOX, the external compliance framework that you use in the context of your business and how it relates, I think, to the data integrity piece.

MS. BARRETT: Well, I'll start, and then we'll work our way down. We do a tremendous amount of internal auditing because we've got all these measurements in place and these benchmarks that we and all of our people are highly incented to drive the number up, maybe a little maybe a lot, but it's always got to go in the positive direction. So we have a whole team that analyses external data, and that's all they do, and they get really good at it. We have created a random-sampled Truth File that's triple verified, and a portion of that is actually verified directly with the consumer. So we can take a data source that we've not ever worked with before, bring it in, run it through a whole series of things, and come up with, if you will, a rating on that source.

So specializing people that really get very knowledgeable about this, that know how to look at data, that know some of the idiosyncrasies of certain types of sources, and there are idiosyncrasies in different kinds of data sources that you get to know over time, you know, is a knowledge base, if you will, that can be built up and I think can pay a lot of dividends over the long haul.

We actually do external third-party audits of our products, but we do not do audits of our quality index. It's just not been -- we're so motivated to make it happen from an

economic perspective that I don't think an auditor would improve our motivation, quite honestly.

MS. LOPEZ: The types of audits that we do is we basically take a random file from a data furnisher and we dissect that entire file. And from that, we're able to extract different key statistics that we share with the customer. Plus, we also just take random examples from our repository and go back to that furnisher to say this is what the data looked like before that file updated, this is what it looks like after; is that truly the picture that you want it to paint? Is that the information that you want represented out there in our database? And we do those audits randomly. We don't have a set schedule in which we do them, but we randomly just take that data and dissect it and take it back to the source to make sure that they're validating that information and they are ensuring that it is correct out there in our database.

DR. SHAO: We're not a data vendor, but we do work very diligently with the data vendor to ensure all the compliance to the law, to the regulations. You'd be surprised some of it. Even though they're not in the room today, so I wouldn't be offending anybody, but, in some cases, are very lax in handling data and the security, and we actually point it out and help them to ensure the security and all the compliance.

CHAIRMAN BEALES: Jennifer, to what extent are your measures sort of absolute measures, sort of truth-based, if you will, as opposed to a benchmark that's a line in the sand where I can see whether I'm moving, but it doesn't necessarily have any objective reality about how good the data is?

MS. BARRETT: They're a combination of those. I mean, our objective is not 100 percent because we know we've set ourselves up for failure. Our objective is an adequate level of accuracy, of quality, however you want to define it, based on all of these variables that relates to the use of the data. And it's perfectly fine to have a five percent non-deliverable factor in the marketing file because people will accept that. And guess what? They're not going to pay twice the price, which might be what it takes to get the quality from five percent to one percent.

So there are economic pressures to say, "I'll live with error." I mean, it's a little bit like the risk management side of it. You know, I'll live with some risk if I can begin to quantify and understand it and manage it down over time. So it really is a set a benchmark as best you can with the intelligence you have and constantly learn from it.

I think sometimes we make the mistake of saying, "Oh, we need to get all the quality right in the beginning, and then I can go away and work on other things." That's not reality because these are all dynamic systems, the feeds are constantly coming in, like she was talking about. You're going to get data, and it may be bad one day. You've gotten a thousand feeds from this vendor and they've all been good, and the thousandth

and one is bad, and you've got to have processes to catch that, throw that out, and correct it. So it's a constant, ongoing measure it, measure it, measure it, learn from it, learn from it, learn from it, and fix it, fix it, fix it.

CHAIRMAN BEALES: We have time for one last question, and that's Lance.

COMMITTEE MEMBER LANCE HOFFMAN: Thank you. I feel like I've been to a fine restaurant, and I was served the appetizer and the appetizer was great. But now one last question, and I'm waiting for the main course, and I'm afraid I may not get it. I'm fearful of that. So let me ask you. I very much, I heard you can't improve what you can't measure. I get it, and I like that. That's very important, it seems to me.

Do any of you have or do all of you have written information you can share without revealing secrets with the committee that goes into more detail? Because that's what I mean where's the beef? I hear a lot of good stuff but, frankly, I'd like to hear more in terms of details. They may be models, they may be -- what are the metrics? What are the indices? What's the usefulness index? What's the data quality index? What's the harm index? To the extent that you can talk about it because that's the kind of thing that would really be helpful to us. Knowing that you have a great proprietary system, to me at least, is less helpful. So if you could share any of that, that would be extremely useful.

Also, the other thing is to what extent do you deal with systems of systems? I know you deal with this to a large degree. How do you handle the interoperability; the indexing, if you will; the kind of data file, where it makes sense, either operationally or other ways, to lash these together? Because that's the kind of things we're going to be dealing with long-term, it seems to me.

DR. SHAO: Well, certainly, we'll be happy to provide any additional information, maybe through Becky, to address the mode of performance metrics, harm index, and probably will appear in the final reports of VA, but I guess we can probably share with you our paper on that and also data quality measures.

And the second question is, operationally, we look forward to the opportunity to work with the agency to apply our technology to the data applications. I don't know how that can be, that kind of discussion can be furthered, but any suggestions and we can move forward on that, too.

MS. LOPEZ: I'd be happy to continue working with the Committee. And as far as additional information, I can definitely take that back and see what additional information we can provide to you. I definitely would work with Becky to see what information I can bring back to the Committee to share in more detail some of the things that I've talked about here today.

MS. BARRETT: I'll answer that in two ways. First of all, I think you really need to have the main course and desert from Dr. Talbert. I think he can, in a non-proprietary way but with a lot of knowledge from what he did for us –

COMMITTEE MEMBER LANCE HOFFMAN: He's an academic and publisher.

MS. BARRETT: He's an academic and publisher, and he understands those boundaries, but I think he could really help you in that regard. As it relates to the details of our system, the question I would put to the Committee is whatever we would provide you of a public nature or not? And, obviously if it is, then that somewhat limits since we view this as a competitive advantage for us to have better quality data than our competitors, so that would limit what we could share with you directly.

But we do consult with agencies on an individual basis and open the kimono, if you will, at that level. I don't know if that helps this com's efforts, but that is also available.

CHAIRMAN BEALES: I want to thank all three of you. I think this has been a fascinating discussion. And if I could just twist Lance's analogy a little, I think what we really have gotten is a bunch of, we've heard a bunch of great recipes. Unfortunately, now what we have to do is figure out how to cook it and actually make it come out in the homeland security application. But thank you very much. We really appreciate you taking the time and effort to be here with us today.

Our last panel is to discuss, this afternoon, is to discuss redress. Our first speaker is James Kennedy, who's the Director of the Office of Transportation Security Redress. Welcome back. I think you spoke to us on your first day on the job. Was it something like that?

MR. KENNEDY: Two weeks.

CHAIRMAN BEALES: Mr. Kennedy joined TSA in September of 2003 as a program manager with the Office of Information Technology. He's held positions as special assistant to the Chief of Staff and special assistant to the Chief Operating Officer and Acting Deputy Administrator for Compliance Programs. And he was a key contributor to the strategic direction of OTSR during the start-up phase.

And then we'll turn to Debra Rogers, who's the Customer Service Officer of the U.S. Citizenship and Immigration Services. Before coming to Washington, Ms. Rogers was the District Director for USCIS in San Diego. She currently works at headquarters in the Customer Services Office. And I think we will start with Mr. Kennedy and go to Ms. Rogers, and then my troops will be back to ask questions.

MR. KENNEDY: Okay. Good afternoon. Thank you for allowing me the opportunity to come back and, once again, discuss redress efforts at TSA. Like you said,

when I last appeared before you in December, I had been on the job for about two weeks. But since that time, the nine months since that time, we've made a number of changes that have significantly improved our program performance. And if you give me a quick few minutes, I'd like to take a moment to highlight a couple of them.

Through our contact center, the traveling public has had an opportunity to provide feedback to TSA regarding our redress processes. And many travelers actually stated that the application process was too cumbersome. They told us it took too long to complete, and it was too expensive to gather the documents that we required. We listened, and what we did was, in response, we developed three modifications to our redress form, which is now known as the Traveler Identity Verification Form, or TIVF for short.

What we did was, the first modification is that we provided the traveling public with an opportunity to only submit one document, the U.S. passport, for identity verification. If a U.S. passport is not available, the traveler can still submit other identity documents that we required before, such as driver's license, voter registration card, things like that. But we found is that over 70 percent of the travelers who actually do request redress relief from our office do have a U.S. passport.

The second modification that we made involved eliminating the need to have the identity documents notarized, and we did that. A simple signature of the penalty of perjury statement has replaced the requirement to notarize the document. And the penalty of perjury statement is actually attached to our TIVF, and so that now will suffice.

Lastly, we've modified our rules of submission so that we can now accept documents electronically, email and fax, of the form, the associated identity documents, as well as the penalty of perjury statement.

The other major change I'd like to highlight is the fact that, like I said before, the traveling public told us it took too long to complete the process, and so to address this concern we actually looked for ways to automate our process. And our solution is the Redress Management System, or RMS for short.

With RMS, travelers are able to submit, as well as check the status of their applications electronically via the internet. And to ensure that the personal data and privacy rights of the traveling public were protected, we conducted a privacy impact assessment, which was reviewed not only by the TSA privacy officer but also the DHS privacy officer. And we also went through a comprehensive system, IT certification process, for RMS.

Our PIA was actually approved by DHS in August. The IT certification and accreditation process is pending and scheduled for approval for this week. So with that in mind, RMS should be launched very shortly.

With the changes I've highlighted, OTSR has actually seen some dramatic improvements in our performance. For example, last time I appeared before you, I told you that our average response time was 45 to 60 days. With our improvements that we've made, even thus far, we've been able to reduce that down to an average response time of less than ten days today.

In closing, I'd like to say, while we've made significant improvements in the last nine months, we know that we still have a long way to go. TSA is an active participant in the working group level and the government support level for the one-stop redress process that Kathy Kraninger described for you earlier today. We look forward to working with this committee to help TSA accomplish our goals of protecting our country while defending our freedoms. And I look forward to any questions that you may have.

CHAIRMAN BEALES: Thank you very much, Mr. Kennedy. Ms. Rogers?

MS. ROGERS: Good afternoon. I think you have me on my first couple of weeks in headquarters this time. I've been here since August, so, hopefully, I'll have a lot of information next time, too. I appreciate the opportunity to address the Committee and also for the public here. It's always a great opportunity for us to talk about our customer service programs at U.S. Citizenship and Immigration Service.

The citizenship and immigration law can be very complex, and our job at USCIS is to demystify the process for our customers and to bring a level of comfort to them so that they feel welcome in our office and also seek out the assistance of community-based organizations, just assist us with that information sharing. The better job we do at educating our customers, the fewer problems we need to resolve down the road.

I've been with the Immigration Service and the Citizenship and Immigration Service for over 20 years and have held positions of inspector, deportation officer, and investigator, and also, over the last couple of years, adjudications and benefits services. In 2003, when DHS was formed, I was appointed as the USCIS District Director in San Diego, and I had the privilege of working with a dynamic team of employees there and a very active and supportive community. So I really have a lot of first-hand information about what kind of problems our customers encounter with our agency.

Last month, in August, I was given the opportunity to manage the Information and Customer Service Division here in headquarters, and I'm very happy to be here and have an opportunity to work on issues at a national level. So I'm very interested to hear what you have to say about our organization.

Just to give you a quick idea of what we do at USCIS, I thought I'd share some statistics with you. The USCIS work force is made up of about 15,000 federal and contract employees. We answer on a daily basis over 80,000 calls at our call center. We assist over 19,000 people at our information counters, issue about 7,000 permanent resident

documents, receive over 30,000 applications for immigration benefits. We take over 8,000 sets of fingerprints and conduct over 135,000 national security background checks on a daily basis. So that's a lot of information.

With that kind of volume, you can imagine that sometimes cases get off track and some decisions are challenged. Service errors occur, misspelled names, incorrect dates of birth, last applications. Unfortunately, even incorrect information is provided at times. These issues need to be addressed as soon as possible. What might be considered a small mistake in the grand scheme of things can have devastating, life-changing consequences to the applicant. A mistake could delay family reunification, jeopardize employment, delay adoption, and result in closing a case prematurely. Quick redress is vitally important to our customer, and I think it defines us as an agency how we handle redress.

Our customers also make adjustments to their cases from time to time. They need to make changes of address, reschedule appointments, and provide additional evidence to us. A good two-way communication system is also essential to meet our customer needs.

We have many ways for our customers and their legal representatives, and I'll go over the most frequently used methods. We have a national customer service center. This is a 24-hour 1-800 line, where a customer can opt to hear information in English and Spanish. We receive approximately 20 million calls into that center every year. Any time during the call, a customer can seek live assistance. Over half of the customers do not seek live assistance.

Callers who request live assistance are transferred to what we call a first tier in the customer service process, and they speak to a contract customer service representative. The contact representatives are trained to provide answers on basic immigration issues, such as how do I file for a relative, how do I apply for citizenship, how can I change my address? All responses are carefully scripted, and the representatives are trained not to deviate from those scripts. A USCIS content expert provides the scripts for the call centers, and all scripts go through a legal review before they are released to the contractors.

If the question to the caller is too complex for the tier one center, the call is forwarded to a well-trained USCIS information officer, and they receive about 800,000 calls a year. If the information cannot resolve the problem, they will transfer that case to the district office that has jurisdiction over the application, and they inform the customer that they will receive a response within 30 days. And the district office is required to provide a written response to the applicant.

We have had issues with the call center. I'm sure all of you are aware of the GAO report that came out I think last year, and we have made some major improvements to the center. Just this year, we re-competed the contract and hired two vendors instead of one so that they kind of compete with each other as far as high performance, and we're right

in the process right now of getting them up to speed. One contract is on, and the other one is gearing up to start in the next couple of months. And with that contract, we have also, it's much more strict as far as what they're required to do and the response time and what the consequence is for abandoned calls. And we have a lot of hope that this is going to work out much better.

Customers can also go to their local office. One feature that we added to the local office a couple of years ago was called Info Pass, and that means the customer can go online to our web site and make their own appointment to go see an information officer at a local district office.

And I personally think this is one of the best things that we've done at USCIS since we started three years ago because, before that, in the district offices, especially the big district offices like Los Angeles or New York, people would wait literally overnight to see an information officer. And can you imagine going in there, seeing an information officer, and maybe not getting the information you needed and waiting overnight? I mean, that is unacceptable. And now they have an appointment to come in, and they're seen within the first 20 minutes or so of that appointment. And also Info Pass prompts them with a couple of questions to make sure they even need to come in. If they just need forms, it directs them to the web site. If they just want to check the status of their application and they have their receipt again, it sends them to the web site. So I think that's a wonderful feature, and it works very well, and they receive an appointment within two weeks of the date that they requested, sometimes even the next day, depending on the volume at the office.

Customers can also access a wealth of information on our web page. We get over 135,000 hits on our web site everyday, as well. Applicants can check the status of their case online, if they have their receipt. They can find out what the average application processing time in each district office, so they know whether they have been waiting too long for processing. They can find out what forms they need to file for certain immigration benefits, and then whatever is new with the agency and press releases, public notices, and that sort of thing.

And in January of this year, we're launching a change of address online feature, which is huge, because that's one of the biggest reasons why cases go off track is that we don't have a system that updates automatically their address change, unless they come into the office and we pull the file. So this is a huge feature, and we're really happy to launch that in January.

Finally, many customers rely on community-based organizations for information and guidance. And we, as an agency, rely on community-based organizations just for information for us. A lot of our customers feel more comfortable going to a community-based organization than they do coming into our office, so it's very important that the

community organizations have the right information to provide to the customers. Almost all district offices have very strong partnerships with these organizations. They meet monthly with the American Immigration Lawyers Association, with adult educators, with employment development agencies, congressional staffers, military. And USCIS provides training to these organizations, host working groups. And a lot of offices also have just public sessions where the public can come in and get their questions answered, and they talk about various topics. I know we did that in San Diego.

And many district offices also make regular appearance on local ethnic TV and radio programs to provide information to our customers. Communication is absolutely essential, and the more our customers know the less they have issues with us down the road or they file an application that they don't really need to file and they've paid the money, and that causes a lot of problems. Some of our new initiatives that we're very excited about is the change of address online, and, as we progress in that process, we'd like to add more self-service features to the internet. And we also are getting better at reporting features with our automated system, so that we can identify trends and systemic problems so that we can fix them with the reporting features.

Finally, I wanted just to share with you some of the information that I provided to all of you. We have in here, this is some of the information that we provide to our customers. This is a fabulous guide for new immigrants, where it gives them lots of information about just about everything a new immigrant would need: how to find a school for their kids, how to get a driver's license. It's really great information. This is provided, we have this in English and Spanish, and we provided them to all of our community-based organizations across the country. Obviously, we haven't given them to everybody yet because they're actually very expensive to make. But we also have other product that people can download from our web site that we give out at our application support centers: the 100 questions that we ask on the citizenship test, how-do-I series, lots of very interesting information and the cards with our 1-800 number.

So with that, I'd welcome any questions. If I don't know the answer, I'll write it down and I'll certainly get back to you. I'll recommend another speaker if you need someone else from our agency. Okay, thank you.

CHAIRMAN BEALES: Well, thank you very much. Before we start with questions, let me remind you one more time that there is an opportunity for anyone who wants to address the Committee to do so. You need to sign up over here on the right. So far, no one has signed up, and we're, therefore, looking at an early exit, which is also fine. But we would love to hear from you if you want to talk and please sign up. And I think our first question was Joanne, if that was really your flag up.

COMMITTEE MEMBER MCNABB: No, sorry.

CHAIRMAN BEALES: Okay. Then Joe Alhadeff.

COMMITTEE MEMBER ALHADEFF: Thank you. I guess I want to pick up on some of the comments that Kathleen Kraninger made to us this morning because, I mean, she talked very much about what seemed to be a customer focus and a perspective to take a look at the final customer, not necessarily the agency involved but the actual traveling public that's involved. And we correctly pointed out that we seem to be a right of passage, so you're working for two months within an organization and you must come and see us.

But the question is that when we look at some of these issues and you pointed out the web site and the 800 number, and I was wondering, when James Kennedy was talking to us about these things, James, you told us that we had, the form has been changed dramatically in response to customer complaints, which is a great thing to hear. But one of the things that we were kind of gripping with the last part was people couldn't even figure out where the form was. And it was putting yourself in the place of the traveling public to find out is this a number I can use because maybe when I find out I'm on a do-not-fly list I'm flying out of the United States. Therefore, an 800 number is no longer a valuable reference for me. Am I likely to be in a position where I can access the internet, or how might I be doing this, and is there a process I can use that it makes sense for someone who's in the middle of the travel, maybe on a multi-hop journey?

So I was trying to figure out have you guys been looking at those issues from that perspective as kind of part of that redress, customer one-stop-shop focus?

MR. KENNEDY: Well, I will tell you that, at TSA, the answer is yes because, when a traveler is at the airport ticket counter, right now they're given basically a two-page document that says, "Here you go. Please contact TSA at this 1-800 number." The thing that we are working with the airlines right now over and the industry associations are, number one, to kind of, I want to say step back from saying you're on the watch list because when somebody is at the ticket counter, a lot of times we will hear at TSA they're told, "You're on the watch list." And what we're trying to do is talk to our airline partners and give them helpful tips on how they can calm the customer down, number one, and tell them, you know, this is really what's going on.

The other thing that we are doing that we have developed within TSA that we are sharing with DHS is we are actually sharing a brochure with them that we developed had we not been looking at this one-stop redress piece, which is an easy understanding guide for the traveling public, and it also would include a copy of the TIVF that they would end up having to fill out anyway. And that's something that says, basically, open up the brochure, there's the TIVF, all the information you've got to fill out, and you can mail it in.

Now, obviously, that's the slow way of doing it and, obviously, we would like to have the internet piece. But we are looking at other ways that we can, better ways that we can communicate. That is something that's on the table. Can I say that we've come to a decision right now? No, we haven't. But it is something that is on the table.

CHAIRMAN BEALES: Mary?

COMMITTEE MEMBER DEROSA: Hi. I hesitate. Well, I guess I should apologize in advance because I'm one of the new members of the Committee and I might be asking you a basic question that you've already covered with the Committee at a different time. But it would be useful for me to sort of get TSA redress 101. What is the just basic process that is involved in I go into an airport, I get a hit improperly, and I want to correct my information? Can you just –

MR. KENNEDY: Okay. I'll do it real quick. What really happens is you go into an airport, and let me really start by saying that TSA really maintains or administers a watch list that, I mean that everybody really is talking about of people that we deem as security threats to our nation's transportation systems. And what happens is when a person comes in, and let's say that they are stopped, and the person feels that they were stopped because of the fact that, in many cases, an airline will misidentify a person, an innocent traveler, as somebody who is actually a person on a watch list. If that person thinks that, basically, what they do is, at the ticket counter, they are told, they're given a document right now that says call TSA at the 1-800 number or log onto tsa.gov and, basically, it gives you instructions on how to fill out the application that you can download electronically. Or if you don't have access to the internet, you can call into our contact center, and they'll mail it to you.

You fill out the information that's on there. There's basic identity information. And what we also ask for you to do is to send us a copy of, let's say it's a U.S. passport, and if you don't have a U.S. passport, other identity information. You send that in. What we do is we verify that information. And what we do is if you are not the person that is on the watch list, what we do is we place your name, your identifying information on a cleared portion of the watch list.

Now, what we do is we send that information out to the airlines so that they know, okay, this person has been through TSA, this person has been misidentified in the past, and what happens is we assist them. We work with the airlines to reduce the cases of misidentifications in the future.

And you all have heard that is the process that we have now. Is it perfect? No. Are we still improving upon it? Yes, we are working with the airline industry and also with Congress on ways to provide immediate relief. But you also have heard about Secure Flight and that coming, which will actually enhance security with the added

benefit of reducing security-related delays. So that's basically what's there now and what we're working on in the future. And if you need some more information, we'll be happy to provide it to you.

COMMITTEE MEMBER DEROSA: Okay, thank you.

CHAIRMAN BEALES: Tom Boyd?

COMMITTEE MEMBER BOYD: Thank you. As distinguished from a situation in which a person is misidentified, as you've just described, how often are individuals placed on the watch list who should not be there in the first place?

MR. KENNEDY: I can't really say that because I don't really know. The only thing that I can really tell you is when someone -- we have cases where people are misidentified, and then we also have cases where --

COMMITTEE MEMBER BOYD: They're misidentified relative to the name that's on the list, correct?

MR. KENNEDY: Right.

COMMITTEE MEMBER BOYD: I'm talking about how often are people placed on the list in error.

MR. KENNEDY: I can't tell you that because I don't know. That's not something that I know. That's something that, if the Committee would be happy to send us a question, we'll be happy to try to find out, but that's not something I know.

COMMITTEE MEMBER BOYD: I'd be curious to know that. Thanks.

CHAIRMAN BEALES: Renard.

COMMITTEE MEMBER FRANCOIS: Do you all have a way of, like a historical database that keeps track of the changes that you make when an individual is misidentified as having a name that's similar to a name on the watch list, and then they are then subsequently cleared and put on a cleared list. Do you all have a way to kind of, from the period of time or if you do this periodically where you look at what's on the watch list and what's on the cleared list and then, later in time, kind of look at what's been removed from the watch list and what's now on the cleared list so that you can kind of tweak all of the rhythms and adjust the program that identifies names and compares them to the watch list?

MR. KENNEDY: We, right now, because of the fact that we are continuing to improve this process, the short answer to your question is, no, we don't have that capability right now. Is that something that's being considered with DHS on redress? That's one of the ideas that's around -- one of the things that we are trying to do is, to get at to your true question, if I understand it correctly, which is if somebody has gone

through our redress process and may still be having problems, how do we know how many times that person is still having problems. That's a capability we don't have right now. Is that something that we're looking for? Yes. But it's not a capability that we have right now. Did I answer your question, sir?

COMMITTEE MEMBER FRANCOIS: You did, and it was in two facets I was looking at it. One, on that kind of customer side, but also in terms of database or algorithmic improvement side, so that if there is a common feature which leads to a misidentification in the algorithm, maybe you can find that common feature by seeing a pattern in practice of the way operate is working.

MR. KENNEDY: I understand that, and, like I said, that's not something that we have right now. But that's something that when we look at not only just TSA redress but also with One Redress, what we're looking at right now are, number one, our system capabilities but also, number two, looking at ways that we can pick up if someone has had an issue in the past. Not just someone, but people are having similar issues that are causing them to have problems after they've gone through our redress process. What are some of those commonalities so that we can address those, but that's not a capability that we have right now; no.

When we talk about redress management system and I like to say this, we kind of started and we do this with the phase development process, we kind of started by building a Buick, and we're kind of going from there. Sorry, but that's, basically, the way we do it with the spiral development. But that is something that we are looking at, but we're not there yet; no.

CHAIRMAN BEALES: Renard, one of the things we learned in the process of looking at secure flight is that, right now, what happens, the matching algorithms now are mostly in the control of the airlines. Each airline does its own thing, and one of the potential advantages of Secure Flight is the government would be doing the matching and would have control over the algorithms and be able to make improvements in order to reduce problems. But, unfortunately, sometimes for the airline, the easiest thing to do may be to go to secondary screening, but that's not necessarily the easiest thing for you. Lance?

COMMITTEE MEMBER LANCE HOFFMAN: It's nice hearing some success stories, especially remember when you came on a couple of weeks on to your job, and now it sounds like a lot of improvement has been made. Pretty soon, I'm going to be expecting customer service, like Best Buy, where I get a receipt and I can go back two years later and say, "I bought this, I think, way back when. Could you please find the receipt since I lost mine," and they have, and it was great.

But, anyway, leaving aside that, leaving aside Info Pass and case check online and all the good things, I don't want to hear about the good things. I want to learn, I want to hear lessons learned and, in particular, for other agencies within DHS, if you had to pick two things which you could not point to as successes but rather saying, "Whatever you do, don't do this," or, "If you inherited this, get rid of it," for each of you, what would they be?

MR. KENNEDY: Your turn first.

MS. ROGERS: Well, I think one thing we've learned is not to take on more than we can handle. When you have an inquiry system, you really don't want to have too many. I know in a local office we used to have inquire by fax, inquire by email, come in, and we couldn't handle it. And you'd get multiple inquiries for the same matter in all the different venues because people want the information as quickly as possible so they'll try all different ways.

So I think standardizing the inquiry process to one or two things -- like, for example, when we did Info Pass we thought there will be people who won't be able to access the internet and they won't be able to make an appointment. We have found that is not the case. Most people have found a way to come in; and, of course, we accommodate the few people that come to our door that aren't able to make the appointment. That's a big lesson, and sometimes we do learn it the hard way but not answering hundreds of inquiries that come in because we don't have the staff to do it. So that would be our big lesson learned.

MR. KENNEDY: The one lesson that we've learned through all of this, and this is what we're really trying to address with DHS One Redress is, when a person has an issue with flying domestically or flying period and they apply to TSA, we go through the process of determining that, you know, this is a case of misidentification or what have you and we have that information and we pass it to our airlines, the problem that we all have, and this is a DHS problem, you all know that DHS was created and we cobbled together 22 different agencies with 22 different systems and 22 different ways of doing things. And sharing that information is extremely difficult, and when somebody comes to us and says, you know, "I've had a problem, please help me," and we help them, and then they turn around and fly internationally and CBP says, "Ah-ha," and that's something that we're trying to fix with DHS One Redress.

I wish that we could have fixed that from the jump, from jump street. But, unfortunately, with the events surrounding 9/11, we didn't really have that luxury of saying this is the best way that we should have addressed this from the beginning. But the one thing we all know is an issue is sharing that data so that a person will only have to go through this process once and not have to go through this process multiple times because, obviously, from the traveler's perspective, that's a lot of time, that's a lot of

money, that's a lot of wasted effort that, if we had that one face, that's something that we obviously would love to have had from the get-go.

MS. ROGERS: Getting it right the first time, even if it takes a little extra time, I think that goes across both of our agencies. There's nothing more frustrating than trying to address an issue and not getting it right and then having to come back and seek redress again and again and again. So that is definitely something that I'm sure both of us are working hard at including.

And the other thing I think that is an important lesson to learn is don't assume that everybody will seek redress. Some people are very afraid of us and very afraid to come into our office and inquire about their case. I think it's very important that we promote a positive, welcoming image so that people aren't afraid to come in and say, "My case is off-track," because there's nothing worse than not getting a benefit that you're entitled to simply because you didn't inquire about the error. So that's something that we're also working hard at improving and haven't gotten there yet.

MR. KENNEDY: And actually piggy-back on that, redress right now is very reactive. You have the incident, and then you start to complain. And it takes a while to get through that process. One of the benefits that we hope to see out of Secure Flight and we see some now when we find an issue, but it's on a one and two basis, we want to be more proactive, where before the problem occurs or before a person has a chance to really complain to us, we will see a problem and be able to fix it. Now, obviously, that's something that we're working on and, through Secure Flight, we'll actually have some data that we would be able to see a person have an issue with their passenger name record, and, "Gee, this person has been misidentified over and over again. Can we fix this?" That's something that we would like to see, but that is something that it's still in the future for us. We want to be more proactive, and we're looking for ways to do it, versus just sitting back and waiting for the spear, if you will. We want to be more proactive and help people up-front, if we can.

CHAIRMAN BEALES: John Sabo?

COMMITTEE MEMBER SABO: Just a question about the methodology you're following to get there. In other words, Ms. Miller spoke earlier today about special counsel's office, and she talked about redress, and then Ms. Kraninger talked about redress, and you're talking about redress and you're talking about redress and these future enhancements. A couple of questions on your methodology for creating this DHS one unified redress process. I mean, is there someone in charge of it, and what is the working relationship? For example, have you mapped out all the redress systems that are currently in place in a composite way showing the business process flows and then some of the distinctions? Some redress will vary from agency to agency within the Department or system to system, I'm assuming, because of the nature of the program. So the question

would be what's the methodology, and do you have that type of documentation that is leading you to this unified system? Can you discuss that at all or provide that?

MR. KENNEDY: Well, I work for the special counselor, so whatever she said today is what we're going to do. But, no, what we do is, yes, we do have a share of the governance board, and Kathy Kraninger is one of the co-chairs. And what we have done is when we first got together, one of the things that we do, not only at the working group level but at the sub-working group level, we actually have a standard operating procedure sub-working group, where we actually came together and said, okay, what is it that TSA does, what is it the CBP does, what is it the ICE does? And what we did was we said, "Okay, these are the things that we do. Are we at the point of saying, okay, well, we're going to with this one method?" No, we're not there yet.

But what we are saying is what we were able to do thus far, and we had the, I think, a governance board meeting yesterday where we said these are the things that we are seeing that are common, and where are the commonalities right now with what we all do? We all take intake requests. We all process requests, and we all have a determination of solution to get back to the customer. And we said how are we able to use the efficiencies that we have existing within DHS to actually proceed with that process in a standard way that is, in a way that we can either save money, save resources, whatever. Those are the things that we're doing right now, and we're still mapping that. But, yes, we are doing that.

And what we do is when we do see opportunities, synergy opportunities, we take advantage of that. And what we do is, if we can save resources by doing that, then we will apply them to other phases of redress. But that's something that we're doing right now.

MS. ROGERS: And I just joined the Committee yesterday, so we'll have a lot more time together, James and I. I know we have information that we can share and lessons learned that will help with the one- step process.

CHAIRMAN BEALES: Charles?

COMMITTEE MEMBER PALMER: Let me see if I can make this clear. I'm not sure it's clear to me either. First, I assume if someone is not on a watch list and they've never been on a watch list, then they're not on any list? Meaning –

MR. KENNEDY: Okay.

COMMITTEE MEMBER PALMER: What I'm getting at is let's assume, for argument's sake, that I'm not on any naughty list, and I go check in and my name comes up not, just don't know. Some are really on a list. My question is once someone gets on a list and they go through redress, are they ever returned to the state that I think I'm in, which is on no list?

MR. KENNEDY: You're saying that if you're not on the list and someone misidentifies you –

COMMITTEE MEMBER PALMER: Yes. Assuming I'm very low key, I'm not on any list, so you don't have me written down anywhere.

MR. KENNEDY: Right.

COMMITTEE MEMBER PALMER: Okay. And then, perhaps, poor John here manages to end up on the naughty list. He goes through the redress process and successfully convinces you he's not naughty, and then he is put on another list, as I understand, but not the bad list, the used-to-be bad list. The cleared list I think you called it.

MR. KENNEDY: Okay.

COMMITTEE MEMBER PALMER: So my question is there any way for him to ever get back to my state, which is no list?

MR. KENNEDY: That's a good question, and what I will say is that's one I'm going to have to get back to you on. The reason being is what we do is, like I said, we do put people on the cleared list, and the watch list is very fluid. But if there's a way to, I guess you're saying, basically, if I can just sail through and I have no problems at all. What we are trying to do with the program Secure Flight is we are trying to make sure that we have enough identifiers where we reduce the number of misidentifications, so that, if he was bad and he's clear now, that, you know, with Secure Flight, he would be able to go right on through.

Right now, can I say that that's going to happen right now? I can't really say that because there are a lot of different parameters with that. But what we are working toward with Secure Flight is just that, where we can reduce the number of misidentifiers.

COMMITTEE MEMBER PALMER: Okay. And in a follow-up, when John does file the paperwork to get off the naughty list onto the cleared list, as opposed to the non-list which I'm on, how long does that take typically?

MR. KENNEDY: Well, that's what I said in the beginning. If he does bring his paperwork in or send his paperwork in, when we receive it in the office, if he has no issues at all, that's a pretty quick process and that's something that we do strive to make that a very quick turnaround time. And right now, overall, our average time is in ten days. But if there is something that we have to work out, then we have to work it out and there's no time frame.

But like I said, when I first briefed this committee, we were at about 45 to 60 days. And now the average turnaround time is about ten days, so we have reduced that a lot, and that's just through process improvements and things that we've been able to do now.

And also with RMS with all of the things being turned in electronically, we should even be able to see some improved performances even beyond the ten days. Absolutely.

COMMITTEE MEMBER PALMER: Great. Thank you.

CHAIRMAN BEALES: Can I just clarify what I really think is the structure here of how this works? If it's a misidentified somebody that's not really on a watch list, then TSA can handle the whole thing internally and through your redress process and put somebody on a cleared list. If it's a correctly- identified person who is, in fact, on the list, then they have to get sent, directly or indirectly, to Terrorist Screening Center to get that problem resolved. Is that right?

MR. KENNEDY: That's correct.

CHAIRMAN BEALES: Okay, okay. Ramon?

COMMITTEE MEMBER BARQUIN: This question is for Debra and, if you don't know, then that's fine, too. I know that the legislation created the Office of the Ombudsman to sort of watch over CIS, and a lot of what the ombudsman is about has to do with redress. So I'd like to just get your thoughts in terms of how that has worked out, specifically on redress.

MS. ROGERS: I am familiar with the Ombudsman Office and the Ombudsman, Mr. Khatri. We've met several times in San Diego and as soon as I got here because our offices definitely have a lot to do together. I think that all our agencies need an oversight, so I don't think there's any problem with having an ombudsman looking at the processes as long as the dialogue is open and we can bring to his attention issues that he may not have noticed when he wrote up his assessment and the same in return, that we are open to his communication and comments and address his issues. So I think that the process is good in theory because we need to have somebody watching our processes and outside. And it also gives applicants an opportunity to speak with somebody besides us if they have concerns about our processes.

And, of course, we're very much held accountable because his report goes to Congress, and we need to address his issues. So it definitely gives us the ability to reach and be creative in addressing problems and looking at things in a different perspective at times. So I think the process is, it's something that we're required to have. It's written into the legislation. We love it. No one likes to have to deal with problems, but, if there are problems, we need to deal with them. Did I answer your question? Okay, good.

CHAIRMAN BEALES: Neville?

COMMITTEE MEMBER PATTINSON: Thank you. I'd like to address my question to Debra. As you may detect from my accent, I'm a fairly recent product of the USCIS, having got my citizenship about nine months ago.

MS. ROGERS: Congratulations.

COMMITTEE MEMBER PATTINSON: Thank you. My question really is, I came here under an employment visa and then a green card and then became a U.S. citizen, so, obviously, I've been processed with fingerprints and face more times than I can remember, including all my family and so on, as we went through the process. As we go through there, we clearly leave records and have various transitions from change of states.

So the first part of my question is how long is that information retained as far as once you become a U.S. citizen is that purged or is it maintained?

The second part is about when I was a green card holder, I had the misfortune to lose it at one point and had to get a replacement. I couldn't find a redress system to help me because I subsequently enjoyed the secondary screening every time I came back into the country for a short business trip, and that became something I couldn't find a way out of in asking the various officials. It was very much, "Oh, this is just the way it is. We can't tell if you're using your lost or your new or if you're the right person to be using this green card," and I was thinking all the fingerprints and faces and the fact that there were cards involved with different numbers, there should have been a way to, fairly quickly, work out that I was the right holder. But I couldn't find any way of any redress for that process, so that's the second part of that. Is there a redress procedure for green card holders that find themselves in that situation?

MS. ROGERS: Okay. First, how long is the information maintained indefinitely. We have a national records center, and we store all what we call alien registration files, where all your information is kept from the time that you immigrate right through to citizenship is kept there. But once you become a citizenship, a citizen, we send it to the federal archives. And the only reason we would ever need to pull it back was if you had lost your certificate and we needed to issue you a new one or something like that.

And as far as the lost card, I think you're talking about when you go to the airport or you cross and they say that you've lost your card. That's a difficult issue because we want to make sure that you are the appropriate person traveling with the card, so we want a record that you lost the card. But it is an inconvenience to go through that identity process, if the inspectors have a concern. I'm not with the CPB, Customs and Border Protection, so I don't know what kind of process you're going through. But as far as the records, the record is just simply that the card was lost. It doesn't put you on a list, on the watch list. And we have improved those processes, too, of that record keeping.

CHAIRMAN BEALES: And last, but not least.

COMMITTEE MEMBER SABO: Picking up on that, wouldn't the process map need to include consequence? I think that's a valid point that if you've lost the green card

and you've legitimately restored it and, yet, you're still subject to additional screening because somewhere the process hasn't connected that, then that may be something. And as you map your processes and your redress, consequence should probably be addressed. Otherwise, you can't deal with issues like that. Just a comment.

MS. ROGERS: Right. But, also, we do have some applicants that lose their cards more than once, and then there's a concern that they're losing them on purpose or selling them. And we do need to have a system in place that we can also detect those kinds of trends and patterns. So we do need to keep record of lost cards, obviously. Then if someone else is using it, we need to be able to pick it up so that someone is not using his identity. So I agree; we need to have a more fine-tuned process, but I think it's very important that we do have a process where we can identify that the card was lost and it's still floating around somewhere and are you that person or are you the person who has the card, a lost card?

COMMITTEE MEMBER SABO: U.S. visit and ident and fingerprints and so on should deal with the identity issue –

MS. ROGERS: Of course. That's the wave of the future that we'll come to a point where a fingerprint will identify the person instead of the card.

CHAIRMAN BEALES: All right. Mr. Kennedy, Ms. Rogers, thank you very much for being with us.

MR. KENNEDY: Thank you for having us.

CHAIRMAN BEALES: Ms. Rogers, maybe you can come back after six months, too. MS. ROGERS: Okay. Happy to. Thank you.

CHAIRMAN BEALES: We're part of the initiation, the hazing. Mr. Chief Privacy Officer, a few words of wisdom for us at the end of the day?

SPONSOR TEUFEL: I think you're offering more than I'm willing to give, but I have a few words. There were a couple of questions that came up earlier, and one that I was in the process of answering, and then the Deputy Secretary came in, and maybe a question that was posed to the Deputy Secretary, matters of process that I either hadn't answered or I may have discussed with some of you yesterday or earlier today but not in a public setting. And so since we're back live, I wanted to just get to those two questions that were process oriented and give a little bit more in terms of answers.

There was one, and I think Mr. Alhadeff had asked about how do you get in earlier, how you approach it; and Kathy Kraninger had given an answer that I had given yesterday and I now want to make public, which is you get in by figuring out where the money is going to be, and that's through the enterprise architecture of the Joint Requirements Council, the Investment Review Board. You get in there early because

these things are often, these programs have long lead times in order to get necessary funding. And so my sense is that that is a way that one gets in early to make sure that things are built in, privacy is built in.

And the other way that we've seen, with TSA for example, with our colleague Peter Pietra is to have a privacy officer in the component. And there are a handful of components that have privacy officers. U.S. Visit and CIS privacy officers were here today, and I'm not sure that we have any others, but we are looking to expand component privacy officers into more components, certainly those where it makes sense to have them, maybe not necessarily all of the components.

And then the question that I was asked about getting, how do you get the job done, do you need more power, juice, authority to get the job done, and Kathy had answered that and the Deputy Secretary also had answered, and that is, first and foremost, you have to have the trust and confidence of the Secretary and Deputy Secretary. And second is the nature of the relationships that you have, that one has, whether it's the screening coordination officer or the privacy officer, the nature of the relationships that one has with the officials within the department. And if you have a good working relationship and they have a sense that you're there to do the right thing and not play gotcha or something else, you're going to have folks who are more willing to work with you to get to the right approach, the right answer, and people who are willing to listen to you when you say, "You know, I think that's a bad idea, and this is why I think it's a bad idea, and maybe there's some other ways that you can approach that issue."

So, no, I don't think more authority is necessary. Moreover, that tends to lead to an adversarial relationship with the folks that I, as the privacy officer, need to work with in a cooperative fashion. And so that actually would probably deter me from accomplishing my mission. That's all I have.

CHAIRMAN BEALES: Thank you very much. We do have one public speaker, Tres Wiley from Texas Instruments. Mr. Wiley? Under our rules, you have three minutes. I doubt that –

TRES WILEY: I'll make this go very, very quickly. It was a simple request. My name is Tres Wiley. I work for Texas Instruments and their RFID and Contact with Smart Card Technologies Group. I've been watching your organization for about a year now tackle some of these very difficult topics that come before you. And, of course, one near and dear to us is RFID and its appropriate use in identifying human beings.

I listened this morning to you and, judging by the amount of gray hair or lack of hair thereof, experience tells you that you want to get privacy considered early in the design process, and I think you many times gave that admonition to people to get involved early, not try to put privacy protections on after the fact but get involved in the

process of writing the policy, designing the system, and not simply after the fact was it done right or wrong.

This afternoon, we addressed redress: how do you fix it when you do make a mistake? And my request to you is to please recognize the importance that you may have in influencing the DHS. And some of your works that you have started now, one that was near and dear to us, was this paper on the use of RFID for human identification. And we saw the first draft, and I will tell you, of course, we were not necessarily pleased with the conclusions thereof, but we thought it was a timely topic and it remains timely. I was discouraged to hear today that it's going to be yet another quarter before another draft of that is completed.

And we have some activities going on the Real ID Act with the U.S. Visit, all of these things having to do with the potential use of RFID or contact list technologies. And my request is simply have a sense of urgency. Give the DHS a paper that is meaningful, factual, thoughtful, and timely so that they can involve it in the front-end of these programs and not have it come out after the fact and not be nearly potentially as influential as it might otherwise be. So it was a request. Thank you.

CHAIRMAN BEALES: Thank you very much. Well, I think we've had a productive day, and yesterday, as well. Is there a comment from anybody on the Committee? All right. Well, thank you very much. Thank you all for coming, and we'll look forward to seeing you again soon.

(Whereupon, the foregoing matter was adjourned at 4:33 p.m.)