NWX-US DEPT OF COMMERCE

Moderator: Robin Wyvill
September 9, 2016
1 – 5 p.m.

Coordinator: Excuse me, this is the operator. I'd like to inform all parties that the conference is now being recorded. If you have any objections you may disconnect at this time. Thank you.

Katharine Abraham: We have a full afternoon's agenda so I'd like to go ahead and get started. We'd like to welcome our visitors - guests to the second full meeting of the Commission.

As you know we're approaching our work on the Commission by gathering information on a range of topics and issues that are relevant to our work.

The agenda for today is focused on issues related to privacy and data protection that are obviously important for us to be considering as we move forward.

We actually spent the morning in closed session, getting up to speed on some of the underpinnings of these issues, with presentations among ourselves, from some of the Commissioners who have expertise in this area.

Part of the reason that we're starting by talking about privacy is that we know that there have been past efforts to think about trying to make data mare accessible for research evaluation and other purposes. And they've in some cases, run aground as a result of not fully considering these issues related to privacy. And we don't want that to happen with our recommendations.

So, we've organized this meeting early on in the process of our deliberations to make sure that privacy is something that we have in mind, you know, at every point as we move forward.

We will of course, in other sessions - future sessions be turning to other topics. For example at our next meeting on - it's tentatively scheduled for November 4, we are planning to focus on important issues from - related to program evaluation and research.

So, we have a lot on our plates and we're tackling these topics in turn. But today we're focused on privacy and look forward to hearing from our panelists. Ron.

Ron Haskins: I'll make a very brief comment which is, I've been impressed by the fact that I think every single person on the Commission realizes that this is a crucial issue. It could really disrupt the response to our report if we don't get this right.

So we're very pleased to have such a great panel. And we're going to reach out to lots of other people as well. And of course, we have very good expertise on the Commission itself.

So we're aware that this is a big deal and we're going to treat it as such. And it will be a major part of our report.

Katharine Abraham: Great. So why - I'd like to go ahead and get started. Our first two panelists are from the - speaking to us from the Federal perspective about privacy. And I'm really happy to have both Marc Groman who's the Senior Advisor for Privacy at the Office of Management and Budget.

And Katherine Wallman, the Chief Statistician of the United States, also from the Office of Management and Budget.

So if I could turn the floor over to you, I think you were going to start Marc?

Marc Groman: Well good afternoon and thank you very much for inviting me to participate this afternoon. I'm really excited to be here.

I also want to thank the Commission and compliment the Commission on holding a full day to explore issues related to privacy and the work that you're doing. I think that's very important.

And it is my hope that I can be a valuable resource to you as you explore these issues, both today and in the future.

And so I'd like to also point out about my background. Currently, I am serving as Senior Advisor for Privacy in OMB. But I've also been a Chief Privacy Officer for many years where I had to grapple with what I call, privacy on the ground. So, real world experience, grappling with real world challenges.

And I've done that and worked in both the public and private sector. So, I have both of those perspectives.

So in my time this afternoon I'd like to start with some higher level principles and then get to some more concrete approaches to how we address privacy throughout the federal government, in the time that I have.

So first, you know, one thing that does distinguish privacy and confidentiality - and when I think about privacy, it is that I focus on, and our profession focuses on, information about people. And it's that's fundamental.

And it's about the collection, creation, use, sharing, transferring, disposition, security, and destruction of information about people. And in my case it's about when the federal government engages in those practices.

Often privacy is viewed as a roadblock for information sharing or, a roadblock to success. And that is not this administration's perspective and I would suggest respectfully, that that is completely wrong.

To the contrary, privacy when properly implemented through good programs and early on, promote innovation in every case, privacy allows for and encourages the wide scale adoption of new technology. And in fact privacy, when implemented correctly, fosters more information sharing; not less.

And it is the administration's perspective - and this administration has been very focused on, technology, digital services, and open data.

And in creating my position and establishing the Federal Privacy Council, what the administration and the President has stated is that privacy is not a roadblock. To the contrary, privacy is fundamental to the long-term success of every initiative that's a priority.

And it is fundamental to the long-term success of the issues that you are exploring here today. And so we really believe that and it's reflected in the Executive Order where the President stated that.

Notwithstanding our focuses on innovation and technology, privacy has never been more important in this country than it is today. And privacy is fundamental to our democracy.

And then he stood up for the first time ever, the Federal Privacy Council which I have the honor and privilege to Chair.

I want to point out also that there’s been a lot going on in privacy for some time. But really what we’re seeing now is that the administration is doubling down on how we are addressing privacy. And we are leading what I call, a fundamental shift in the approach to privacy across the federal government.

And I want to highlight in particular, the way we’re looking at it is sort of a three-pronged approach. Looking at people; looking at policies; and looking at accountability. And we need all three to make it successful.

We could have the most current policies in the government, but we must have the right talent in the federal government to implement those policies. And if we do, it will dovetail perfectly with the initiatives you are working on. Meaning, privacy professionals trained to handle tomorrow’s challenges which includes the challenge of de-identification of anonymization and the like, because that is the job of the privacy professional in the federal government.

This is important to get right, because when we talk about information about individuals, real harm can result when it is processed improperly. When it is released in some cases. And so there are real consequences and real impact to real people. And beyond that there is real harm and real impact to our nation, our government and our agency if these issues aren’t approached properly and thoughtfully.

But I am actually a believer that we can do it. Nothing about this work troubles me. To the contrary, I am quite excited about this commission and the work you're doing.

So, I talked about a shift in what we think about privacy that's codified in A130 which I would commend to all of you. It's a must read for anyone in the privacy profession today either in the government or otherwise.

And the shift is away from viewing privacy as a compliance or check-the-box exercise, to viewing privacy as part of a comprehensive, continuous, strategic, and risk-based program.

So to be sure, there are laws we must comply with and that is critical. But privacy needs to be viewed from also a risk lens. As we approach a world of -- I don't want to use too many clichés -- but whether it's the Internet of things or the new digital services or new ways you want to interact with the public, we have to explore privacy from a lens about risk.

And we have to do that intelligently. It has to be based on fact and science and technology; not on conjecture and not on just stakeholder interest and political deal making. But it has to be informed by fact.

And that is where, I think, we're headed. To ensure that every agency is taking a proactive, not reactive approach to privacy. And that we will be engaging in a risk-based approach.

We're not here to eliminate risk. The only way to eliminate risk is to have no data and, no one has any interest in that. But rather as A130 makes clear, what we want people to do is to identify risk. And that means not just risk to the agency but risk to individuals and risk to the agency and the nation and the government. Identify the risks then take steps to mitigate those risks.

Where you can't mitigate all risks, you have to acknowledge that is a residual risk. And then you need to be accountable and own it. And that's what we're looking for through a privacy program.

And it is that last part that troubles many people but, it's unavoidable. When you're dealing with data and information about people, we're not asking for eliminating risk. But I'm asking you to understand the risk and be accountable for that risk. And that's part of every privacy program and it needs to be.

I want - a few other points to highlight is I was asked I view different ways of releasing data has different potential risks. And so there is a distinction when you want to use data within the government. So it's collected for one purpose and you'd like to use it for research within another agency.

And there's a difference as to releasing it to the public and those two scenarios create very different risks that need to be discussed in detail because they are very different scenarios.

Privacy as a principle level, I think I won't spend time on that. But A130 reflects the Fair Information Practice Principles which include transparency which means, as the United States Government we have an obligation to be transparent to our citizens about what we collect and how we use it.

We make representations about use and we have to honor those representations throughout privacy.

One of the difficult issues that is teed up by this and other discussions, indeed it comes up in almost - most all my work each week is, addressing the issue of de-identification and what does that really mean.

And that has to be explored. There’s work taking place in other places. But for most conversations in privacy, and for most things to succeed or initiatives, we need to get a handle on de-identification to have a common understanding; a common language and nomenclature.

And to make decisions again about de-identification based on fact and science and the state of technology today and tomorrow to understand things like the risk of re-identification.

The mosaic effect; how can two data sets be combined and used in a way we didn’t contemplate? And what are the potential consequences to people when that occurs?

This is not about risk elimination. But if you don’t have a common language on de-identification, anonymization, synonymous data; identifiable data based on actual fact and technology, then it’s impossible to do what we want to do here. Or in contexts like healthcare and precision medicine and other areas.

I think I’ll just close by referring again to A130 because A130 was issued in July of this year. It was last issued in 2000. And this 88 page document is the core document around information government for the entire federal government. And it is mandatory.

It is not best practices, it is required. And it addresses information security. It addresses privacy. It addresses the release of data, and it talks about data in the full lifecycle and therefore, it’s highly applicable.

And even in the - and I worked very hard on this and I don’t believe any of these concepts are mutually exclusive or rather, even inconsistent.

So A130 in the basic consideration says, an agency - a basic consideration is making federal information discoverable, accessible, and usable because it can fuel entrepreneurship, innovation, scientific discovery that improves the lives of Americans and contributes significantly to national stability and prosperity, fosters public participation in government.

But also in those basic principles it says that, protecting privacy; protecting an individual’s privacy is of utmost importance. The federal government shall consider and protect an individual’s privacy throughout the information lifecycle.

And then it goes into detail about how agencies are to address that. And I read those two basic considerations because as a senior advisor for privacy, I don’t think they’re inconsistent with each other. Particularly as we enter -- not enter -- but move forward in this digital age we have to grapple with all those issues comprehensively. Not in silos.

We have to know data governance. This document talks about data governance a lot. Having inventory sounds basis. Know where your data is. You need to know where it is to protect it. You need to know where it is to know what you should release or not.

So data governance is fundamental to this. Information security, privacy, and doing it all consistent with law best practices. And overarching the whole document is this risk management framework which I believe is the way to responsibly tackle that.

And the final point is about people which is that the Federal Privacy Council is most addressing now which is, just as in cybersecurity, we discovered several years ago we didn't have the right talent pool. Or we had good talent pool but we needed more.

The same is true today of privacy. And that is part of the reason why the President established a Federal Privacy Council. It's in the Executive Order. And one of our directives is to first of all, bring in more privacy talent.

So, to improve hiring and to improve the ability for us to bring people in from the private sector. And I'm thrilled and excited to state that on September 14, we are hosting the first ever Federal Privacy Talent Summit with the Office of Personal Management, to do just that.

But it's not just about bringing in more talent. It's training the talent we have today. And we have a huge emphasis on that as well, including all kinds of training programs on both law policy and also technology for privacy professionals so that our privacy professionals today are prepared for the challenges of tomorrow. And that includes the kinds of issues that you're working on.

So I would be delighted to answer questions today and to be a resource at any time in all of your work.

Katharine Abraham: If it's okay, why don't we hear from Katherine and then have general questions. I'm sure we'll come back to some of the points you made, Marc.

Katherine Wallman: Thank you very much Katherine and Ron. I'm not sure who I'm supposed to (unintelligible).

Let me start by echoing Marc by saying, thank you very much for your consideration of this issue and also expressing my personal and my team's willingness to help in any way we can as we proceed (unintelligible).

Katharine Abraham: Turn your microphone on.

Katherine Wallman: Okay. All I had done so far was to echo what my colleague from OMB said about thanking you - the Commission for taking on this work. And also for our personal willingness - my team's willingness to help out as you proceed.

I also - Katharine, I echo what you said about making the data more accessible, but not running around when we try to do it. You and I certainly have some experiences along those lines and we will take those into consideration I'm sure.

I think it easily fair to say that protecting data and building trust are sort of the raison d'etre of the statistical agencies and surely have been our mantra ever since many of them were started.

I'm going to focus my remarks on just one narrow piece of what we do because I think it's perhaps the most relevant to what the Commission is doing.

And what I was asked to do is to talk a little bit about the Confidential Information Protection and Statistical Efficiency Act which henceforth I will call, CIPSEA, for this conversation in order to save quite a bit of time.

My - next slide please. Whoever has got it now? I'm going to just do a couple of things in the short time today which you can see. And if you'll flip me to the next slide already, that would be great.

This slide says that I'm going to talk about the purpose of CIPSEA. But frankly my preference is always to ask at the beginning, what problem were we trying to solve with CIPSEA?

And I believe there were two problems we were most focused on. One of them was that we had uneven statutory protection for the confidentiality of information that the federal government, in its statistical programs, collect.

It was a surprise to me and I know to many, many people who worked on this activity in the decades past, to learn for example that the Bureau of Labor Statistics, until the passage of CIPSEA, although it had an unblemished, and I will say, reputation for protecting the confidentiality of its information, was doing it based on something that was called a Commissioner's Order.

It did not have statutory protection and lived in some fear frankly, that at any moment the information could be taken by courts and whomever. And that fortunately, was not the case but, there certainly were threats along the way.

So one problem that we were trying to solve was to even out and undergird the statutory protection that all the agencies have. And you note I said, even out. We did have some agencies that had statutory protection. Some that had a little bit and some that had none.

So one of our objectives was to not only provide that statutory protection for those who didn't have it, but also to make it clear, when we were talking about giving a pledge of confidentiality, that we were basically talking about the same thing, regardless of which agency was collecting the information.

In doing this we did not in any way, diminish any existing authority the Census Bureau has long had. A couple of extra provisions in its law that are not the core provisions in my view, but we did not undo anything that was there.

The second major purpose; what problem were we trying to solve in putting together CIPSEA, was to enable greater sharing of information among statistical agencies for statistical purposes. I have to say all those words at least once or I won't earn my money today.

This is coined, efficiency, but I think it goes, in my view anyway, it goes beyond efficiency. At the time a selling point was that it was more efficient if the agencies could share the information. But I think it also thinks to what we're here more talking about today is, improving the information that's available for use by a variety of agencies and ultimately by policymakers in aggregate and other kinds of forms.

I suppose this is a cautionary tale and I'll try to keep it very brief. That despite the fact that you may have incredible consensus in the administration and even good colleagues in the Congress who believe what you're doing is good government, it still takes a very long time and lots of little knits and funny paths that one takes when one is trying to pass legislation that seems as simple as straightforward as we believe that CIPSEA certainly was intended to be, and I think even came out being.

It did take a number of years. It took a number of Congresses to put this in place. We started doing something within the administration in 1994 in the form of Confidentiality Order which was intended to provide the same sorts of protections, in particular.

But we really were going for the law. And you’ll see that this law was finally passed several years later and that does not reflect any - the minutia of the effort that was put into it during that time. But rather a constant pushing along the way and getting new friends on the Hill and so on during the process.

There was some unfinished business in CIPSEA. And I think one piece of it is very well-known and has to do with the ability to change information. Inherent in that was some tax data that are inextricably intertwined with some of the economic statistics that the Census Bureau. And as I live and breathe still to this day, we are still trying to get that companion piece settled.

And as early - as recently as this morning I received another version of our most recent attempt. I haven’t even had a chance to read it yet. Thank you. Because I don’t have it with me is my - I call my Blackberry because this is my first iPhone and I’ve only had it for a week. So you’ll forgive me if I refer to my Blackberry.

The less well-known activity that took place in the efforts with CIPSEA was that we originally in the administration, designed this to cover all data that are collected by statistical agencies. And as we came up to the final moments, it was actually within the administration that a preference -- strong preference -- was expressed for limiting it to economic statistics.

This was occurring in the late 90s when we were doing this run up to final presentation on the Hill and so on. And there was grave concern expressed about the potential effects on the taking of the census, in particular, if all those guys in Washington were sharing data. So I just put that down as one of those little things that can happen on the way. May I have the next slide please? I don’t have too many.

There are different ways that one can be involved in using CIPSEA. The principle users and named agencies in CIPSEA are the principle federal statistical agencies -- 13 of them -- that were grandfathered if you will. They were the longstanding principle statistical agencies. They were called out in the Confidentiality Order and then taken by reference into CIPSEA itself.

We did however, put a provision in that would allow us to recognize additional units that could be empowered to be statistical units and therefore take advantage of some, but not all, of the privileges if you will, of CIPSEA.

We have, since the passage of CIPSEA, authorized -- blessed; I’m not sure what word we used -- three additional statistical units in the government as statistical agencies for purposes of CIPSEA. And other agencies can also achieve this recognition if they wish to.

This is not a limited set of opportunities. It does not require a huge amount of work on the part of the agency to do this. So if that’s something that other agencies want to consider, that’s still up for consideration.

I would note, just in case it hasn’t been clear or that we’re all living on the past to some extent, that CIPSEA does apply whenever an agency is collecting data for statistical purposes under a pledge of confidentiality. So, it’s not just surveys.

The word in the law is acquire, and that was on purpose. That if we acquire data from administrative records; if we acquire data from a state government and bring it into the federal government as part of what we’re saying is being protected by confidentiality for exclusively statistical use, then CIPSEA applies to it.

Was that a ding? Oh, that was that ding; sorry. I got used to testifying and having the dings and lights thrown at me.

So I think that I just want to echo another thing that Marc said, that when the -- may I have the next slide please -- when agencies are using CIPSEA, we are talking about minimizing the risk of disclosure.

There are severe penalties in CIPSEA if you haven't read it. A quarter of a million dollars and five years in jail for willful disclosure of information. Those are the most serious disclosure penalties that I'm aware of in federal law for this - it's a Class E felony. I'm not a lawyer in case you can't tell, but it was a pretty serious business when these penalties were put in.

This kind of activity that we've listed down here under procedures, these are normal operating procedures for our statistical agencies and programs. They didn't begin with CIPSEA and they won't end with CIPSEA.

It's not a check-the-box thing. That's what I was going to echo Marc. This effort is not check-the-box either. This is a lifestyle or whatever the proper way is to call it.

I just would close these comments by, if you'll flip to my last slide, I'm just going to mention a couple of ways that CIPSEA has been particularly effective, we think.

As far as we know, that we have, under CIPSEA's success, we've protected information from requests from law enforcement or others that would have some sort of useful information.

By the way, I meant to say Marc, when we talk about confidentiality of information on the recipient end if you will, we talk about the broad array of data. So we're not just talking about individual or even…

Marc Groman: That's why I made the distinction.

Katherine Wallman: Indeed, indeed, on household data, but all of our economic statistics and so on, are equally protected. So I just want to underscore that point, which I meant to do sooner.

We do -- and it maybe is tied to my second point here -- that we do have some evidence that response rates from industry; from businesses actually are improved when we give a promise - pledge of confidentiality.

I'm hesitant to say this now that I know you're a lawyer, but we do know that lawyers in businesses in corporate offices often look and find that CIPSEA pledge and make that an important part of their decision to respond to most of our information collections which are after all, voluntary not mandatory.

I did mention that we've had three agencies who have added themselves to the core that are able to use CIPSEA. I don't have a good answer for you or clear answer for you at this point, of why more haven't been lined up outside my door.

But I think that invitation remains open. And we do have examples of what it takes to qualify. And we pass them on graciously to anyone who's interested.

So I think I would like to stop there and like Marc, entertain questions.

Katharine Abraham: Thank you both very much. We've got some time now for questions. (Christa), could I ask you to come up and join the panel up at the front. You've, I gather, been working with Marc on some projects.

Marc Groman: And if I may point out that, the Federal Privacy Council has brought in the statistical community because of the synergy of issues because we, in many respects care about and are working on very similar issues.

And that (Christa) has been a leader on that front and is somebody that I have the opportunity to work with often. In fact, the pleasure to work with. And I use (Christa) as a resource quite often.

Katharine Abraham: Right. So, please accept that you want to jump in to answer questions, please do. So let me open the floor for questions.

Paul Ohm: So, thank you both for that presentation. I'm Paul Ohm on the Commission. I think this is more for fellow lawyer brethren. So we had a pretty interesting, wide-ranging conversation this morning. And two questions I thought I'd love to hear your perspective from any of the hats you've worn in your career.

So number one would be a pretty open-ended question about the role of notice and consent in a changing time of data use and data science and data techniques. And in fact Katherine and (Christa), you probably have opinions on that as well.

And whether or not, you know, we need a rethink on that or what? That's plenty to get started.

The second, and pardon me for being a little open-ended here. I wish I had time to formulate this better. You know you invoke privacy on the ground and the idea that in some ways privacy now emerges bottom up, from responsible privacy professionals.

One thing that I think has spurred this Commission into being is, a history of running into people who are, you know, have great integrity and are trying to do their jobs but maybe are conservative or maybe have vague rules dictating what they can share and not share.

I wonder if you could just give us opinions generally on that phenomenon. Whether that's something that is prevalent or not. If there are ways to craft rules to kind of deal with that dynamic, if it exists or not.

So, tons in there. Please just pick up on whatever sparks your interest.

Marc Groman: So, this won't shock you that I have strong views on all of those points. And in this case for the most part, offering my own views but, they're consistent with things I've already said was administration views.

So, let's start with the last point first. Because I believe I've addressed a little bit in my opening remarks and it's fundamental to my core belief that privacy isn't a roadblock. And it is not and should not and doesn't stand in the way of innovation or technology or data sharing.

What - it's the way it's been applied, in many respects. It is in some ways - the problems are attributable to the fact that it has not been resourced. That we don't have the right people in the right positions.

And so I think that those concerns all fed into the President's creation of the Federal Privacy Council to build that cadre of true privacy professionals who will have a baseline level of training so that we're all speaking the same language.

It goes to my point about having to train our current privacy talent and bring in more talent so that we have people who can tackle the challenges of today and tomorrow.

So I share the concern that historically it has been perceived as a roadblock. Or that well-meaning individuals have not been able to foster the kinds of innovation at the pace we want.

But I don't think that's because of privacy. I think it's because we haven't resourced privacy programs. We haven't - and I - you know, and we haven't kept pace.

And so we've gone ahead with other initiatives that require this kind of skill and ramped up here. Didn't ramp up over here. And that's what we're trying to address now with the Council.

So I'm not disagreeing with your proposition. I'm actually agreeing with you and saying that - and it's interesting, I talked to DJ Patil who's also the Chief Data Officer. We're in complete alignment on this.

It's not that its privacy or it's any of the laws, but I need someone - and in fact I will tell you that the reason why I love privacy and my job is that few professions like privacy, are the intersection of law and tech and policy.

You have to know all three. And that means understanding for example, de-identification. And we have to bust the myth of de-identification. We cannot pretend that removing a name from a dataset calls it de-identified. Or that taking the tape off of my (unintelligible) which will then be used for (unintelligible) on the sequence is now de-identified. But we need people to understand that level and to engage.

And so I would agree, but we're addressing that. And it's fundamental to your success.

On notice and consent, I think that whether it's A130 or other privacy documents, I am a firm believer they're still a role. And we can't dismiss it.

I'm also a firm believer that the Fair Information Practice Principles, all eight or nine; however you view them, have a role to play. And they're like dials that you know, when one is more difficult to implement, you have compensating controls.

There's no question that notice and choice; notice and consent have limitations. And the Internet of things is challenging all concepts around notice of choice or notice and consent. And that we will have to grapple with it.

But first of all we have legal requirements around it meaning, the Privacy Act which applies across the federal government. But I believe there's a place for it. And what we're going to have to do is adapt to technologies. And one of the best examples I can think of at the moment is the President's Precision Medicine Initiative.

And the privacy principles which were developed specifically for PMI where we talk about what does it mean to give consent for the Precision Medicine Initiative.

And if you're not familiar with this, it is one of the most groundbreaking initiatives I've ever seen on healthcare. And the dataset is also perhaps one of the most sensitive I've ever seen.

Meaning it includes your genome sequence, your DNA, all of your medical records, sociographic information, in some cases, data from electronic devices, all put into a database for research that allows kinds of research that are really unprecedented.

And so what does it mean to have consent when you agree to volunteer and place your data into this kind of cohort to be used over time by different entities. And what the principles contemplate is not one-time consent.

So it is not a matter of presenting somebody with a very long form that you sign off once. And in perpetuity your sensitive data can be used for any purpose.

But rather it contemplates what we're calling, individual participation in the process. And that an individual can't have access and see, my data is being used for experimentation research around diabetes and cancer and something else.

And maybe you can consent to a more granular kind of research. You obviously should have the ability to revoke consent. But it builds the trust.

I mean I like that Katherine mentioned trust. Trust is fundamental to everything we want to do. When we lose that trust as a government or a program, it undermines the long-term success.

And so I think PMI is a great example where we recognize you know, there's some choice for level limitations. It can't be about clicking a form or one page for perpetuity. And it embraces the concept I love which is, the individual participating in how data is used over time. And then they get to see the amazing benefits, too. Did that help?

(Kathleen Rice): How are you - hi, (Kathleen Rice). How are you enforcing compliance with consent principles in that context?

Marc Groman: In the specific context of…

(Kathleen Rice): Of PMI?

Marc Groman: So PMI is really in its infancy. NIH is just launching. And so we're baking in all these discussions on how we're doing consent is evolving.

But there's a contemplated - an interface that would allow someone who provides information to see how things are progressing over time.

(Kathleen Rice): I understand that. I'm thinking on the back end, if somebody provides consent for uses A and B but not C, how are you making sure that it's not used for C?

Marc Groman: So there are multiple different PMI initiatives. And I'm not - I don't know the details behind how each one is doing that. But for - and I mean it could get very detailed given how data comes in from different sources.

But you know researchers are vetted before they have access to the data. They agree to adhere to certain principles if they're going to have access to the data. And there are controls that are put into place.

For example, you can allow research within a sandbox and give people access. But they don't actually extract all the data and put it on their own PC.

So there are lots of ways, whether its access controls or logs or sandboxing that will allow for robust research and the incredible development, while still having information security and confidence in how the rules are being applied.

Just one example, but I have no doubt there are other ways to do that. But that's the concept I'm familiar with.

(Kathleen Rice): I have a somewhat different question that I thought was maybe partly where Paul was going which is related to notice and consent.

I understand how that would apply in a case where people are voluntarily providing information to be used for some purpose along the lines of what you just described.

But what if we're talking about administrative data that were collected for some other purpose - federal tax records, social security earnings records, unemployment insurance wage records. How do you think about notice and consent in that context or, do you?

Marc Groman: So one of the fundamental and core aspects of the Privacy Act of 1974 is that the United States government is transparent with the American people about what we collect and how we use it. And that we do not maintain a secret database.

And that we are - and explain how the data collect will be used. And that is a core principle of the Privacy Act or other laws and of privacy generally.

And so one of the ways we think about that is that - and there are varying degrees of effectiveness agency to agency, but agencies are required by law to draft and publish a Privacy Act System of Records Notice for data that is in a Privacy Act System.

That disclosure explains what the information is that is being collected and the purposes and how that data will be used, shared and transferred. I do not believe that Joe Consumer reads those and, they're not intended for Joe Consumer.

On the other hand, I can tell you that advocacy groups do read it; IGs do read it. And it provides actually a benefit, because it's about transparency. And there is an accountability feature to it, as well as Privacy Act statements on different kinds of documents.

And so I think that is incredibly important. And that we need to honor the representations in those policies that explain how the data will be used.

I also want to highlight that -- and I think this is obvious -- but that all data is not equal in terms of sensitivity. And that goes very much into the risk analysis and to the thought processes about risk.

But you know PMI is a data set that is extraordinarily sensitive. And the collection and research is so groundbreaking that we want to do that. And other data sets do not raise similar sensitivities or be considered far less sensitive.

And so all of these principles, they're certainly based on legal requirements. But I think it's important that you calibrate based on the risk. And that's fundamental to things like FISMA, where the sensitivity is a kind of information that potential risk of harm influence and inform both privacy and security requirements.

(Kathleen Rice): So is it just risk or is it risk and potential benefits?

Marc Groman: So I think that when the Privacy Act - I mean you have to disclose how the data is going to be used. But in any situation the program, right; the Agency Head for example, has to at the end of the day, make a policy decision about information.

And in making that policy decision, within the confines of legal requirements, that policy official will weigh costs and benefits.

And the important thing is, for the privacy person to present information about risks to individuals, risks to the agency, whether it's reputational risk for misuse or its financial risk from other factors. Risk to national security and the like.

And so consistent with the law it is the Agency Head who is accountable for the agency, who will have to weigh that. But has to take into account risks to individuals that come from the kinds of data.

Man: So if I put the two talks together, one of the jobs of the Commission is discern as wisely as possible, about whether the country would be benefitted by a clearinghouse. The entity doesn't exist now.

And I infer from what you’re saying that if the use scope inside such an entity were fully and only for statistical purposes, there’s the regulatory infrastructure to use CIPSEA to regulate the use if that entity were so deigned to be acceptable for CIPSEA use.

And the original descriptions for the systems of record for the agencies wouldn’t necessarily have to be altered. Did that make any sense and, is that true?

Marc Groman: I’m not able to answer that question right now. I think there’s too much there. And I also…

Man: Well if I understand what Kathy said, that CIPSEA allows protections that the law is applicable for statistical uses. That the agency that would be related to such uses has to be approved under the CIPSEA statute, right. As an OMB you deign acceptable an agency.

Man: Right.

Man: So, if there were systems of records for purposes of administering federal programs; had no statistical purposes in mind, yet they were transferred or made accessible to this new entity, and if the new entity were only permitted statistical uses, CIPSEA could be used.

(Christa): So, I can’t speak to what any type of new entity would be but, a System of Records Notice, if data are transferred to the Census Bureau from another federal agency, even though we are listed among the Privacy Act exceptions, you still have to report that in the System of Records Notice for that agency.

So the Internal Revenue Service, Social Security Administration; all of them report a transfer of data to the Census Bureau.

Man: I see. So they have to revise that? Okay. I didn’t know that. Thank you.

Marc Groman: So one of the things I think we’re going to - obviously one of our mandates is to discuss the potential creation of this clearinghouse.

I guess one of the things we’re probably going to struggle with is whether that inherently increases the risk of making - of violating privacy, or whether there are ways to envision a clearinghouse that might actually lower the risk.

Guess I - can you give me your thoughts on that - how you feel about those issues - those questions?

Man: Well I have to say that the concept of the clearinghouse is new to me. So I have not - that hasn’t been presented to me before. And so I’m not really familiar enough to…

Marc Groman: We haven’t defined what the clearing - what we mean by clearinghouse. I mean so, you can…

((Crosstalk))

Marc Groman: Yes, you can throw any definition you want, but it is part of the statute that created this, to think about that. And it uses the term, clearinghouse without telling us what it is.

Well I’m going to answer it perhaps this way and, by referring back to the Precision Medicine Initiative. Because that is an initiative that does bring together incredibly sensitive information about individuals that, if improperly used - released will cause real harm.

And so the mandate there was not - was to enable this to occur and to take incredible efforts at developing and tailoring privacy principles and security principles to the initiative.

And so you know, I'm a big believer in the question that we answer - that I want to is, how do we do it? Not, don't do it but, how do we do it, to your point.

And I'm generally - I'm a general optimist in that. And the more complicated the challenge the more fun I think it is.

And so in PMI we had to look at things exactly like I mentioned earlier, which is that we understand how sensitive this data set is. And so it's not going to leave the sandbox, right.

So researchers will have access to it and they may even be able to, with permission, add data for research. But they cannot extract it so that we control it.

And we have logging controls and access controls. And you have multiple levels of de-identification so, it's not just one. But you have multiple layers where different data elements have different kinds of de-identifiers. And that only a small number of individuals might have access to the keys that would allow for, you know, re-identification. It also comes down to understanding de-identification at a very, very granular, scientific level and not making it up.

But there are, you know, protection that goes to the consent I talked about which is, understanding and having individual participation in that kind of data set through transparency and disclosure.

And so you know PMI is one of the most fascinating, groundbreaking projects I've ever seen that involves PII because it comes from blood samples come from here. Your medical records are transferred from one system to another; from a VA hospital to another one.

But your survey data comes through an entirely different source and has to be scanned in and then added. And then a researcher can contribute data. And so I find, you know, it's not maybe a fully satisfying answer but, that was a very complicated issue that we were able to address.

The flip side is right, you have all the data centralized. It is a very ripe target for not just security but for a potential misuse. But you know I feel that a lot of these things can often be addressed.

Katherine Wallman: I'd like to take a slightly different tact in answering the question. Not to contradict but, a slightly different…

Marc Groman: No, it's…

Katherine Wallman: …tact. (Latonya), you don't remember, but we were in the same room. Do you remember? We were in the same room a long time ago and you're the lady who taught me about how easy it is to identify people from a single piece of information.

I mean just - I brain still, I can't even remember. I know you were in Massachusetts at the time. I'm using Massachusetts state records. I can't remember any more details than that.

But that really sensitized me just how easy it is. Now I'm a victim of tax ID, whatever you call it, theft. And so I really know. And I didn't give consent to anybody -- thank you very much -- to get those records or to have them mixed with any other records or anything else but, it doesn't matter.

My first reaction to your question was, to echo what Marc was saying before about we have the mosaic effect and all those kinds of things that are before us know. And I've just said everything I know about the mosaic effect so, don't push me too far.

But I think we also have a growing stable of experience in learning how to deal with these kinds of things. And sitting right behind you, in my angle, is (Ron Jarmon) who I'm calling out because through the Research Data Centers and the Federal Statistical Research Data Centers as they're now growing and so on, we're gaining a stable of experience in how to address these questions and how to get to -- now, I will agree with Marc -- how to get to the, can do, orientations towards this question.

And that's where I think we need to be. You know we may need to define what we mean by - what you mean by clearinghouse and what the scope and range and so on is, so that we can address these questions more specifically.

But once that's done I think we do have a growing body of knowledge and expertise to try to do some of these things more effectively without…

Marc Groman: I actually think what we said was consistent. I was also interpreting clearinghouse to mean, internal to the government, for government researchers. That's what I was interpreting. But if not, you know, again you left it to my interpretation.

So that raises - you know, there's a huge distinction between the access by - the use of information by another agency or reusing it for the government to make evidenced-based decision-making. And releasing it or making it available to the public which is data and it can be used for good and bad.

If we put it out, it's not going to just researchers here, but it's going to China and Russia. It's going to whomever needs it or wants to use it.

I also wanted to point out again, sensitivity - first of all, re-identification, I mean I'm echoing what you said, right. And that goes to the point that I've been hitting so hard which is that when we talk about it, it has to be based on technology and science.

And we have to acknowledge how easy it can be to re-identify and then take steps to mitigate those risks. Not eliminate them but mitigate them, truly based on science.

And I also want to urge the commission, as you evaluate your job and look forward is to make sure you consider different cultural perspectives on privacy because that is very important.

It is one thing for certain individuals to say, what is the privacy harm if we collect data about this, sitting here in Washington, DC? And it is very important to understand that we are a diverse country and we're a democracy.

And that cultural groups may view that very differently. And that came up in PMI for example.

And so that you as a group here may think something is a great idea. But a Latino community in Florida or New Mexico may really disagree, as may a certain population of minorities in an urban center, or a Muslim community somewhere else.

And you must take that into account as you think through policies, and not the implications on specific groups and how they will perceive. Again, because we are the government. We are not a company, we are the government.

And it's important that we think about all of our citizens and all of their perspectives on data use and how they perceive it by the government. And I want to just make sure that - hope that you will consider that as you look at this issue.

Man: So, let me just probe that - oh, I'm sorry.

((Crosstalk))

Bruce Meyer: Bruce Meyer, Commissioner. Can I ask both Katherine and Marc, keeping in mind the value of research and evaluation uses of data, how could we usefully strengthen privacy and confidentiality rules? Are there gaps in existing rules?

Marc Groman: I mean I - are there gaps; yes. I mean you know, I'm not set on an administration position, but I mean I think that - I think anyone would identify that there are gaps.

And that's why there is so much ongoing work. That's why NITRD published a privacy and research strategy for the federal government. And that's why there's ongoing work on things like, looking at the Electronic Communications Privacy Act and acknowledgement that that needs to be updated consistent with the current state of technology. And so, I think that's right.

I think that updating A130 for the first time in 15 years has gone a long way to bringing federal government policy and guidance consistent with today because, just a few things have changed since 2000.

And so that was a recognition that we need to fill gaps. And so there are, and I think that as technology evolves, keeping pace with the technology and the advancements, whether it's Internet of things or vehicle to vehicle communication or drones, there are most definitely gaps. And we need to continue to address those.

Katherine Wallman: Having mentioned how difficult it is to get legislation passed, but not to mention formulated, is it within the scope of the Commission to look at the Privacy Act itself?

Bruce you mentioned, are the rules that - I think in blocks of things from legislation to rules to best practices and so on down the line. And I thought you framed your question as regulations, but maybe I was keying in on a word that I'm used to.

Bruce Meyer: I was trying to be (unintelligible).

Katherine Wallman: No, that's okay. I just wondered if - you know I mean, somebody may shoot me before I leave the room.

I don’t even know if there are provisions in the Privacy Act that are detrimental if you will, or at least casting shadows on the kinds of things we’re talking about here. I just don’t know the answer to that.

But I think I would echo what Marc said earlier, that a lot of times it’s not what’s in the law itself but, how it’s being interpreted.

Marc Groman: And I would say that…the narrow blinders, whatever they used to call them, on horses, that you know, the easiest thing to do is interpret it as narrowly as possible and have the what I call, the can’t do approach rather than saying, is there room for us? Is there some constructive we can deal with the situation. That was philosophy.

Katharine Abraham: Did you have a last question (Bob)?

(Bob): Maybe I could just probe Marc on one thing. So this morning we talked about the variation inside the society and how perspectives on individual privacy varies across cultural groups, racial groups, and so on.

It isn’t clear however, after you made the comment you’ve made. So, what then? If you want to formulate regulations or statutes, what do you do with the knowledge of that diversity when you’re talking about privacy issues and privacy protection?

I get the point you’re making. But then, what? Have you thought that through?

Marc Groman: No, in many cases I think that it can be context specific. And for example you know, the emphasis on things like individual participation and trust in PMI. And understanding how different groups will perceive the ability of law enforcement to access that database, which they can, will change perceptions and influence the extent to which individuals want to voluntarily provide data, which that relies on.

Similarly, things like you know, the Census for example. You know I think that - and I believe I've seen studies that will tell you that when trust goes down in the government, individual's willingness or eagerness to respond to the Census also go down. And that actually might vary by group.

And so I think the important thing is to factor it into any particular discussion. But I don't - like everything else, I don't view it as a barrier to success. I don't - I just think that it's as you do a risk assessment, perhaps it doesn't allow for 100% of something that we think could or should be done.

But I go back to my, this is all about how we do and not whether. And so it's difficult, without a specific context. But I just think really it's very important that that just be thought through. And it comes up in many contexts.

Ron Haskins: Okay, we live in a federal system and everybody knows that states are very greatly, and this is no exception. So we're very pleased to hear from several good witnesses from the states who will give us their advice, presumably, and talk about what they've learned about privacy in the various states.

So first we have Aimee Guidera, who's the President and CEO and I believe the Founder of the Data Quality Campaign. Welcome Aimee.

Aimee Guidera: Thank you.

Ron Haskins: Justin Erlich who's the Special Assistant to the Attorney General for the California Attorney General's Office. What would any state panel be without someone from California on the panel? People from New York are probably saying the same thing.

And then Michael Basil, who's the General Counsel for the Illinois Department of Innovation and Technology. So we'll begin with Aimee. Welcome.

Aimee Guidera: Thank you so much Chairman Haskins, Chairman Abraham, and members of the Commission. Thank you so much. It's an honor and a pleasure to be with you this afternoon.

Katharine Abraham: Is your mic on?

Aimee Guidera: I think it is. Can you hear me now? It's a pleasure to be…

Ron Haskins: Just move closer.

Aimee Guidera: …pleasure to be here. And I can't think of a more important mission that what you are doing.

I'm a staunch, passionate believer that until we change the role of data, change the role of research, change the role of evidence in government, and in policymaking, and in practice, we will never get the results that our citizens deserve.

So thank you so much for taking on this challenge and this charge. It's a really important one.

And as Chairman Haskins said, I run an organization that is pretty passionate about this. We're called the Data Quality Campaign. We're a national, nonprofit organization that is focused on changing the conversation about the role of data in education to improve (unintelligible) achievement.

And I took the liberty in true transparency since that’s the theme today, of changing the title of what I was going to be talking about because, I am not a lawyer. And being on with general counsels and a counsel from states, I wanted to make sure that you had the expectation I was not going to be talking about state legal perspectives.

But instead, really talking about the core work that we’re doing at DQC which is, trying to change this culture to value-evidenced; to value data in education.

And I was asked to talk about education because one, that’s what we work on. But two, the lessons learned and what we’re seeing happening in the education section I think, are very applicable to all the work that you all are focusing on in education. So hope that you find it valuable and, thank you again for having me here.

Today this conversation about starting with privacy couldn’t be more important. Because what we found in the last 12 years that the data quality campaign is, people won’t use the data unless they find value in it, and they can trust it. And trust that it’s not going to hurt them or their children.

And this is a really important part, because if we don’t get this right then all the work; all the other work that you all are talking about, all the work that we’re doing day in and day out, and that people across this country are doing, won’t amount to anything changing because, people just won’t use the data. You know, teachers will close the classroom doors and nothing will change.

But to gain this issue of privacy right matters, but I also want to concur and reinforce what the first panel said. That this cannot be a compliance exercise.

You know privacy is not a project. It's not something we're ever going to be done with. As long as we're talking about evidence and research and data, we have to be talking about privacy.

And more importantly, this word trust really, really matters. Because trust is not a compliance exercise. And for too long in education and I think in probably lots of other sectors, what we talked about is saying yes, well I'm compliant with FERPA, trust me.

Well, guess what, that does not breed a sense of warmth in a parent's heart when that's the response that they get when someone says, what are you doing to protect my child's information?

Being compliant with the federal law that no one understands and people don't get with within it, doesn't breed trust. And as a result, we won't have people using data. So, it's a critical, critical piece.

And I think it's really in education what I also want to concur with because we have to remember this value piece. Because the trust piece comes out to value.

And so it is so important when we're talking, in any sector, about data or researcher evidence. It's reminding us, especially those of us that get excited about research and data, why we're doing this. Why is this data actually being collected? Why are we doing a research study on this? Why does this matter?

And if we can’t answer the question about why we’re collecting something; why we’re studying it and how it has an impact on a stated goal, or it means something to one of our citizens or to a stakeholder, we probably shouldn’t be collecting it or using it.

And that’s the core piece of trust and privacy is starting with, what is it we’re collecting? Why are we doing this? And this is a major challenge for us right now in education.

It’s helping people to understand the, what’s in it for me. Why does it matter that somebody has access to my kid’s - to my child’s information? And how is it going to help them be more successful in life?

So the focus on helping people understand that value-add matters so much, and it’s a critical piece of this trust. And it’s what drove us at the Data Quality Campaign to start a year-long process of working with a broad-based group of stakeholders in this field to define a vision. A big idea of, so what.

What does it look like when data is actually used to service students and to serve student learning?

And for ten years we’ve been talking about longitudinal data systems. We’ve been talking about all kinds of data. What we realize is until we start putting the students in the middle of the conversation and remind people that this is about -- we actually call her (Grace) -- this is about (Grace), and making sure that all those that are working with (Grace), that’s closest to her, have the information necessary to help (Grace) succeed. Nothing else matters.

And every data conversation needs to come out of that vision of how do we make sure that information, which is just data that you know, has evidence behind it that turns into evidence, is helping those that are making decisions for that child to move forward.

And this vision piece is important because people see the value add. They see themselves in that vision. They say, I need that. I deserve that. I haven’t had that information and I need it to be successful and help my child be successful. It changes the entire conversation. And so looking at that people side of the equation really matters.

And I want to pause for right now. So one of the tasks I was given was also to think about, so what does it look like, and what is the state of play right now in education? And how are we getting closer to making that vision a reality for students across this country?

And I’m going to have two other parts to my conversation. One, giving an idea of what’s happening right now in the education sector, what’s happening not only in the growing infrastructure. To be able to collect and use information to inform student achievement and to make that vision come alive, you’ll have what’s happening to protect it.

But second of all, to also get some recommendations for you all to be thinking about, about what are the appropriate - things to think about for appropriate federal action to ensure that this information is kept safe, secure, and that we’re building trust.

So one of the challenges that we have in education, and I think it would be in every public sector is that the nature of data has changed so much.

If we ask people in the education sector, so what is data, the number one thing people say is, well it’s a test score.

Well, that’s so different. And you know this complex side shows that data is everything from teacher observations to parent input to attendance information, to course taking, to remediation rates.

All of this information provides a richer picture of when these data points can be connected and linked up together rather than just having a single data point in time.

That’s a really important idea that we’re struggling to get out into the hands of people to understand that when you’re able to connect the data points and connect it up to an individual, you have a much richer picture of what’s happening in that student, with student achievement in that school, in that district, and it changes conversations.

This conversation is a big one. But also the conversation of who’s using this data for what purpose, and who needs what at what level?

You know one of the big myths that we’ve had to do some myth-busting about is that people believe that they hear about this richer information. And that what’s being collected is going directly to Washington, DC on an express train.

So what’s happening, all this collection of data and all these new ways we’re collecting data right, that we’re not just collecting it through tests any more. Information and data is being generated constantly now in the classroom with the use of new technology.

So all of a sudden that raises all kinds of issues about privacy and security. When we don’t know how that’s covered right now, there are questions.

So one of the other questions is, if it’s electronically done it makes it so much easier, the perception is, that it can go straight to President Obama’s desk you know, that your kid’s fraction scores are going straight to the Department of Education, to the White House.

And so part of what we’ve been doing is also helping to build that case of helping people understand that not only is that illegal, that are four federal statutes that say that it is illegal for the federal government to have any student level information identifiable in the federal government.

But also that the vast majority of that information is being kept in the classroom. Is being used with families, with teachers, with students. And that at every point there are roadblocks to keeping that information from flowing and, for good reason.

And also helping people understand that there are different uses for this information so that when a teacher is using the data for continuous improvement, for thinking about tailoring the lesson plans to students in the classroom, it’s very different than how a district is going to use it to make sure that they’re allocating resources to get the biggest bang for their buck. Or that a state or the feds are looking at accountability provisions.

These are all complex issues that all data is not equal. All data doesn’t do the same thing. All data doesn’t have to go to the same people.

And so there’s so much to unpack that we - you know, there’s a good reason that data is a four-letter word in a lot of people’s conversations. It’s been used as a hammer; not as a flashlight. People see it only as accountability. They don’t see it as a tool for improvement.

So there's a whole lot of just understanding and helping people understanding the, what's in it for me, that I think is critical to have as we're also having these conversations as trust, this value piece.

And what I also want to pause on one minute in these next two maps, is showing you all the incredible progress that has been made in the infrastructure of this nation in terms of the ability to make those prior to slides possible.

This is a map of the country in 2005 when the campaign was launched. The map is showing the number of states having what we considered, a complete statewide longitudinal data system as defined by having ten essential elements of a data system.

What you can see there is that the country has turned remarkably orange. And now every single state now has the statewide longitudinal data system in place.

Likewise, this is a map that we -- sorry, it goes back and forth -- that we had created ten actions to - that we launched in 2009 to also look at the change in the state focus on policies that made sure that data was not just collected, but that it was also, the focus was getting data into the hands of people who needed it.

And there was a focus on using the information coming out of the data systems, not just collecting it. And what you'll see here, over the three years is that increasing numbers of states have put into place the tools, the policies to ensure that the focus is not on collection, but on the use of this information to improve student achievement.

And we argued, since we stopped surveying on this, every state now is in position, and has an infrastructure that allows the conversation not to be about collection but, about use.

But what comes of this -- sorry, this map, let me see. Sorry, that's a better map. And now it's stuck -- is that more of this data becomes accessible as people keep hearing about, we have more data; we have richer data. They're seeing it. They're seeing their classrooms. There's more data points.

The testing piece comes into this, as well as education. There are more and more questions arising about privacy. And this is part of a natural evolution that as we have more data, more people are using it and more questions are being raised.

And this is also in large part due to the emergence of just a whole lot of different kinds of data that people aren't clear of.

You know, one of the big words being used in our data world right now is, biometric data. And it sounds really scary and like Star Trek type of stuff. And people are having questions about how it's being used and how it's being impacted.

And as a result.

Ron Haskins: Aimee, we have two witnesses and we only have about 45 minutes.

Aimee Guidera: Yes.

Ron Haskins: Okay.

Aimee Guidera: I'm going to go really fast. This is - so when we go through this and we see that this data is being used more and more, it prompts lots and lots of conversations. Because questions are being asked, concerns were coming up. And as a result, Americans do what they do.

They call their members of Congress. They call their state legislators. And what you see here, that you all talked about this morning, that we've always had these laws since 1974. And yet really in 2013 we saw this explosion of media focus on education on privacy and concerns. And we saw an explosion of federal legislation.

And then I'll run through three slides just to show you maps. I won't read the numbers to you, but let your eyeballs go across them.

The year before this in 2013, one state had one piece of legislation introduced. And one law was passed in Oklahoma. This was in 2014. This is in 2015. This is in 2016.

And you have summaries of these years in your packets that we'll hand out to you in a second. And also we'll be getting you the 2016 legislative one next week.

And to give you a summary, as a result in the last four years, every state in the nation except for Vermont has now introduced and discussed a piece of legislation around collecting and protecting student data. Of these, four to ten bills were introduced, 40 in 49 states, 72 laws have been passed in 36 states.

So the conversation is on. And the discussions that you all are having about federal actions are not happening in a vacuum because states are having these discussions. Governors are involved. State Boards of Education are involved. Legislatures are involved. This is a conversation that is happening.

And when you look over the last four years, some of these trends I think, are really important for you all to be thinking about as you're thinking about the federal action.

One is when the conversation started, many more of these bills were being introduced and laws passed, were more of the restrictive type. They were much more about regulating governments, collection of student data, and putting restrictions on this.

But over the time what we've seen is more of a focus on regulating vendors and online service providers. And then this year we saw people going back and refining laws that were on the books and being much more constructive in what we call a governance model.

And what we're seeing, the vast majority of bills and laws right now being discussed and passed in this country are dealing with four pieces.

One is the focus on communicating. Not just the value of data, but also communicating the need to be transparent, calling for audits of data that's being collected at the state level. Why it's being collected, and who has access to that, and mandating that that is being put on state Web sites so that it takes the boogey man out of the closet and people have to see what's being collected for what purpose and why. And also, states are encouraging sunsets of data that no one needs.

The second piece we're seeing is this focus on really talking about roles and responsibilities; the governance piece. Who's responsible for what, at what level? Who’s making sure that someone is in charge of defining access, destruction, who has access to our data?

Third of all we’re seeing a mandate on talking about revising and refreshing policies and practices about privacy, security, and confidentiality. Whether states have them or whether they are outdated, making sure that they are reviewed on a regular basis.

And fourth of all, a focus on capacity. How do we make sure that at the local level these places are put - these pieces are put into place? But also that people have the training and the knowledge and the capacity and the conditions to be able to be on top of this at that ground level privacy piece.

So give all that and given where the states have been, and the fact that everybody has a role in ensuring that we have a culture of trust around data, what is it that you all should be thinking about as you’re making recommendations in this Commission report?

One thing I’ll say before I div into these four recommendations is one, that there’s so much, in addition to the states that have been working on this. We also have seen others playing a role.

The private sector, the vendor community out and many of you have probably seen the pledge that the parents that the private sector has made, that puts out there and says, this is what we commit to do and not do with your children’s data.

We helped organize with the Consortium on School Networking, and effort that you have in your packets called, The Student Principles, which education constituency organizations; both unions, the state Legislators, the Governor’s

Association came together and made a statement about exactly what they would do in making a statement of the value of data. And then you all, the federal.

So what I've put together here, and you have a summary on one page of your federal things that you should be - the three recommendations we make. I want to briefly put them into a format that we released of roadmap for states. A format in a framework for thinking about how do you ensure that that vision I showed you becomes a reality?

And that's a framework for thinking about, how do you make sure that data isn't just collected, but that it's servicing students at all times.

And so bottom line, measure what matters. The federal government needs to make sure that what it's collecting and making sure that what states are collecting and locals are collecting actually matters in achievement into the results that we want.

What that means, there should be a federal review on a regular basis of every piece of data being collected. And somebody needs to be able to say whether or not it leads to a conversation about student achievement and how it matters.

And if it no longer can have any impact on student achievement, it should be sunset. Because every collection of data, there's a cost. It's a burden cost. There's also the cost of risk with every single time you collect a piece of data, a risk.

Second of all, this whole capacity and conditions piece, how do we create the conditions of trust and people wanting to use it?

The states are working to get their data houses and their privacy houses in order. The federal government also needs to do this. You know the laws that you all were talking about this morning in the reviews, they’re there. And a lot of people have argued that they are serving their purposes and out there. But there is a lot of confusion.

People know that these laws are over 40 years old. That the world has changed. They are confusing about how they apply to online digital work. And as a result we have confusion and ambiguity which is not leading to trust, it’s leading to issues.

So one of the things is to update and revise laws such as FERPA and COPPA. And to make sure that these are not only reflecting the realities of today, but also being responsive to the confusion out there.

And we need to break down the federal silos. You know PTAC has been doing a world of work of making sure that they are coordinating more with FTC, as COPPA is doing that.

But the important thing to think about is, if you’re a Principal, thinking about doing the right thing at an elementary school, you’re getting guidance from three different federal agencies thinking about, how does this all add up?

And it needs to instead be thinking about, what does that federal - the Principal need to get from the federal government to make it easy for him or her to do the right thing when it comes to protecting data.

Third of all -- I’m almost done -- this issue of transparency cannot be talked about enough and reinforced. It is absolutely critical that one, data that governments have has to be turned into public indicators so that people get value out of that data.

And two examples of this that the federal government has done and it cannot do is one, the college scorecard of taking data that the feds were collecting and now turning it into indicators that are incomplete, but it's a step forward, in terms of showing people and giving people information they need to make better decisions.

And the second piece to the clearinghouse conversation, the prior panel, this issue on the band of the student unit record system, we're on record of saying, should be overturned.

There needs to be a conversation about solutions of making sure that data that is sitting in different silos can be linked up responsibly. There are lots of ways to do that by linking state data systems. By having a federal clearinghouse.

I'll leave it to you all to figure out the parts of that. But by having a band means you can't talk about it and you can't talk about solving those issues. And as a result, people do not have the information…

Ron Haskins: And you are going to have to move on so…

Aimee Guidera: And I'm done. And the last one is just the issue of - I'm sorry, I thought I had ten minutes.

And that the last thing about privacy is that access and privacy, those closest to students in this country still are not getting a guarantee of having access to the information that they need and deserve. And as a result, the federal government is silent on that in its latest reauthorization of the Elementary and Secondary Education Act.

And that's something that the federal government can do to incent states to ensure that those closest to students themselves, parents and teachers have it. But also there's a whole lot more that needs to be done about capacity building, both within the federal government of making sure that offices like PTAC, the Privacy Technical Assistance Center, has the ability to provide the guidance, the support, the technical assistance to those on the ground.

But also to provide direct assistance through funding to those on the ground that need it. They need the training, they need the capacity-building because this is happening on the ground level. And how do we make sure people can be responsible stewards of the data that they need to use?

So, I'm sorry for that and thanks so much for the time to talk. And I look forward to questions and answers.

Ron Haskins: Mr. Erlich?

Justin Erlich: Great, thanks. Sir, thanks very much for having me to this meeting. I am Attorney General of California's Advisor on Technology, Privacy, and Data.

And so in that capacity I both spend a lot of time thinking about how consumer privacy and California privacy can be…

Ron Haskins: Can you bring the microphone a little bit closer?

Justin Erlich: Yes.

Ron Haskins: Pull it over to you.

Justin Erlich: Oh yes. That's great.

Ron Haskins: We actually would like to hear you.

Justin Erlich: Great. So to start briefly from the beginning, I'm Attorney General Advisor on Privacy, Data, and Technology. And so in that capacity I spend a lot of time thinking about how California's privacy can be protected, but also how we can use our administrative data to better inform criminal justice policy.

And so in that regard we sort of wear multiple, competing hats. And so very must respect and appreciate the work that you all are doing today.

I think you've heard at the federal level from sort of a wide variety of how to tackle this issue. I'll give you the criminal justice area to sort of complement the educational one that you've just heard.

I brief, we launched an open data effort caked, Open Justice which is really about trying to be transparent around our criminal justice data to respond to a real need in the community for really wanting to understand what is going on in the criminal justice system. Use of force, all these questions, the FBI Director making comments about how there's a real (unintelligible) of data.

And so we wanted to enter into that space. But as we did we became very aware that as we sought to strengthen trust and focus on metrics, we used a dashboard to show information, but we also launched an open data portal which provided the raw data, or started to.

And we started to think quite a bit about all these privacy risks. And so the sort of tension of trying to be transparent to meet that call from the community while also making sure we weren't exposing privacy risks, became very real to us.

So I'll just give you some of the things that we have learned in the field. I'll be brief so you all can - so we can have more time for questions for any of us later.

In brief, the goal of all this was, you have data, you analyze that data, you share that data, you get awareness, and you might actually get some real policy change.

And we released six data sets - crime rates, clearance, deaths and custody, arrest rates, law enforcement, and also the agency and county level. So if your privacy radar is one, you might see lots of different opportunities for where this might cause some eyebrows raised.

Just to give you a quick sense of how it looks, we did a lot at the statewide level, and there's not a risk in that. It also doesn't always hide some of the - show some of the trends you might want to see.

So we've also started to break down at a country level, and we were also down to the agency level. And you will notice, in California some counties are as small as 1500 people. And so the risk of re-identification become very real and this is something that we're still struggling with the best ways to do.

And you'll see some missing colors there because it's such a small set of numbers where were not quite sure what to do with them right now so, we're punting.

Why we’re doing this - we think data is a public good. We think transparency builds trust, open by default is a key signaling function from the government that we’re trying to reduce transaction costs and help researchers.

It increases the potential simply from a capacity perspective. We have a very small research team internally. We have some research partners but, there are only so many that we can work with directly.

And there are whole sets of things with data that we can’t even think about the uses of ourselves, but someone very smart somewhere would. And we want to make sure that we’re unlocking that.

At the same time we’re quite concerned about identity disclosure and the risk of re-identification among others. In the criminal justice space there are sites like Mugshots.com which many of you may be familiar with, which is exactly the sort of thing we don’t want to see, people posting booking photos and then charging to take them down. And this is actually a growing industry across a variety of sets.

And so this is sort of our greatest fear is, as we try and help criminal justice policy that we also are creating a potential for those who want to manipulate and exploit the data.

There’s also risk which is a little bit less about privacy per se. But as you put out open data and the data may not be perfectly reported, you always run the risk of making some inferences on bad data. So we do some work to make sure the data is clean.

I think a few things I'll just highlight on a criminal justice side that sort of may be different than the educational side or the healthcare side, there's actually a fair amount of data that becomes available in local jurisdictions through court records and PRAs. But our criminal offender record information is confidential.

And so your arrest records, your dispositions, all of the demographics about it. So there's a little bit of a tension in that. And we know that balancing privacy is so important in this area, these criminal records can often be stumbling blocks to applying for jobs or housing. So we're quite cognizant that getting this balance right is very important.

Our task is made more difficult because there are many other access points that just get written up. Like press releases from district attorneys or law enforcement. News coverage of trials. And so if we think the linkage attacks, we know that, particularly as we go to smaller communities, if you find a 70-year-old White male arrested on a certain date in a small county, it would not be that hard to find that from a data set later.

And we also find ourselves without the same guardrails that might be offered from HIPAA or FERPA. HIPAA as one example I think, has rules for counties or jurisdictions of under 20,000 and provides some other ways. And the law enforcement space doesn't have that.

And we're actually seeing our open data effort as part of a broader effort in the community. The practice around law enforcement, the White House Police Data Initiative is something you may know, is pushing for more transparency from a lot of local enforcement agencies.

And people have been following on many different sides, some publishing information that then later created issues.

And I'll show you some of the information that we have - gender, race, age, offense type, date of offense, all sort of quasi-identifiers which is, we won't publish any names or unique identifiers. But these are all types of things that can help inform. But they also can help identify.

And we have problems, not just the small cities and counties, but small law enforcement. So we will be collecting a lot of use of force data due to a bill that was just passed. And we're supposed to release incident level data on law enforcement demographics. But of course if your law enforcement is small, that is also easily identifiable.

So we're currently exploring with (unintelligible), combining age buckets, month of offense, and trying to aggregate jurisdictions. There are some experiments of how to strike this balance.

Just a couple of other things that we've started applying is each data set is a unique snowflake. And so there are many different stakeholders from each of our groups.

So victims of crime in some data sets, that has been a lot in the news lately around some data sets were released that you could actually I think in some states or cities, someone was able to figure out victims of domestic violence. Who to report it. That creates a large issue.

We have the Decedent Rights. If someone has committed suicide we want to make sure that their name is public. That creates a set of issues.

We particularly want to be sensitive around juvenile data. So, it's sort of complex and the balancing changes from data set to data set.

So what we're doing for now, we're looking to do tiered data access. So anonymize some data on an open data portal with the Data Use Agreement. I put anonymize in quotes because we've all heard today, that is difficult to do completely.

Making some data available just through external researchers. And some data is actually so confidential that it can only be used to internal researchers.

In terms of how we're trying to anonymize the data right, this is an ongoing process. And we'll also look for guidance on what you all come up with. But we have been having a lot of internal conversations on our risk appetite on how to best balance that. We've been speaking to experts.

And then also looking to collaborate with data scientists or statisticians on how to model risk so we know, for these smaller jurisdictions, so what is our true risk exposure.

And finally, as just was alluded to in the last talk, we're trying to be transparent on transparency, which is how we choose to create our policies. We'll be creating some sort of white paper policy to make clear on the journey we've been through so people know how we have landed on our efforts. And also it might hopefully be a guide for local enforcement agencies who are being asked to be transparent as well. Thank you very much.

Ron Haskins: Thank you very much. Mr. (Base), can you pass - yes.

Man: Well, thank you for the opportunity to be here. I now have a little sharper focus on what's expected here. I'm here on a little bit of short notice but I think there's a fit for what information I'm bringing to you. I want to start by pointing out, I'm the general counsel for the newly formed Department of Innovation and Technology for the State of Illinois.

We did not have well centralized IT function before the current administration and this is a step we've taken to better consolidated both the hardware, the people, the software and the thinking and hopefully we're on the right track.

And I think you're going to see today a couple opportunities where we can demonstrate to you that we are on the right track and we might have a couple of answers to some of the questions that were raised.

We've heard a lot about anonymous data, de-identifying data and you at the federal lever here in Washington, D.C., need to think about macro ideas and big picture trends.

Don’t forget that a key collector of data is the state government and we are an important social service agency across the United States and we have direct one-on-one contact with people who need social services from the health and human services area, public health, to a lesser extent education because that's' locally driven but criminal justice is largely influenced by state policy.

And we actually need you to consider the concept that (Mark Growman) brought up and that is calibrating your approach to data and data privacy because there are many instances where we need the data to be identified and we are looking at a lot of different initiatives in the State of Illinois to aggregate data from different agencies on the individual level to take a 360 degree view of a child in the educational system, of a person in the criminal justice system, of a person in the health and human services realm, of a family unit in that system or in those systems or in the workforce datasets that we collect.

Being able to and not prevented from drilling into individual data and compare them across agencies is fundamental to what we do and really important for us improving the way state government interacts with citizens.

With that said, I have two things that I brought to you today and hopefully they build on what was spoken about earlier. The first one is public versus private universities; accesses. I was asked to consider one particular dimension on unemployment insurance data and there's a current federal regulation that I'll talk about for a moment briefly.

And then the bigger picture idea, Number 2, data sharing agreements and the Illinois Enterprise Memorandum of Understanding. I want to talk about those two ideas with you and I think they'll integrate well with what we've heard today and maybe what you're trying to accomplish through the commission.

Public versus private universities, access to unemployment insurance data. Right now there are federal regulations that requires understandably to keep unemployment insurance data confidential and those same regulations define what it means to be a public official and a public official can include a public university.

We have several fine, many fine, private universities in the State of Illinois and federal regulation is prohibiting us from sharing important data with those researchers. Those researchers are willing to give us their support and analysis, and for free, because it's a great educational opportunity for them and we're more than willing to take their analysis and information and thinking but we are prohibited to often from sharing the information with them.

So, in terms of areas where our data privacy and federal policymaking and where there's room for improvement here's a specific example of an opportunity for the federal government and the regulatory environment to help improve the ability of states to make use of data.

Number 2, and I think this is really going to help, I hope, advance your discussion and consideration of the problem you're being asked to solve. We have recently accomplished in Illinois an agreement amongst state agencies. So this is within the State of Illinois. It's an Enterprise Memorandum of Understanding which no good bureaucrat would do without an acronym here so we're going to call it the EMU.

And our EMU is, I think, a really helpful construct for you to consider. And let me start by asking, is everybody in here a member of something like Avis Preferred or a similar rental car arrangement? And think about that for a minute, you've filled out a whole bunch of information about the type of car you like, whether or not you want GPS, whether or not you want a car seat, whether or not you're going to waive insurance or if you take the insurance, which of the three confusing levels they offer you.

And it facilitates your experience when you get to the airport. You, in some cases, can go right to your car, you get a notification on your phone, you go right to the car and you get in the car, you drive out, you show your ID to make sure you're not a thief and you're on your way. It streamlines the process.

Well, we've done that for data sharing within state agencies within the State of Illinois. It doesn’t take care of every item on the checklist you need to but when DCFS, our Department of Children and Family Services shares data with the Department of Human Services but if that list of things they need to agree on is 25 elements long, this knocks out 21 or 22 of those elements and it allows us to turn the data request or the data sharing arrangement around in a matter of usually five days or less.

The last couple of items that need to be ticked off are what are the specific use, what's the specific duration and a few other internal control items that allow us to move nimbly, quickly and stay in compliance with all of the relevant state and federal laws.

And we're trying to rinse and repeat this for a lot of different approaches. Our dealings with local government. You know, at the federal government you have the balance of federalism between states’ rights and the federal governments rights. At the - at least at the State of Illinois our relationship with local units of government is a lot more authoritarian. We have a lot more control and power over a county, a school district, library district or a park district.

Nonetheless, both of us can gain and benefit from the sharing of data so we are currently drafting an EMU for that type of inter-governmental exchange outside of not just state government but state with other units of government.

And as you see from the slide, right now our EMU is internal but we are contemplating an external data standardization sharing agreement. Our EMU establishes common agreed terms so there's - you don’t have to reinvent the wheel every time two agencies who have never dealt with each other want to share their data and try and create a synergy there.

We have an active operational committee chaired by our state CIO and the various agencies have representation on that and they can be nimble and respond to changes (in siti). This is undoubtedly a really interesting and important time (in siti). We are going through a digital transformation.

These offices that I drive by in Washington D.C. are filled with desks that used to be filled with paper and cabinets and people couldn't walk in off the street and access the information. Now anyone in the world can get access to this information quickly, efficiently, sometimes surreptitiously and that creates the risks and the opportunities that we've heard a lot about today.

We've not dictated a technological platform in Illinois on which data is shared because technology changes so quickly. Part of the process of sharing data involves the two agencies agreeing on what platform they're going to use and making sure that they understand it.

The goal of course, and the result is that we think we have a rapid compliance and flexible data sharing arrangement in Illinois and it's a new and exciting and great opportunity for us to improve the way we deliver services to Illinois state residents.

I guess I jumped the gun a bit - we are looking at this at many different levels and trying to standardize the process and create the Avis Preferred service. I'm not showing for them, I'm just trying to create a metaphor that you can all appreciate.

The Avis Preferred service for data sharing at every opportunity that we interface with our sister agencies or within state government, our local use of government, the public, we're trying to figure out a way to make this safe, quick and transparent and so when the question was asked earlier about whether or not there should be a clearing house in a new federal agency that can be constructed that will house all data, I don't think there's a right or wrong answer to it.

I want to suggest to you that a data sharing agreement and standardizing the way each agency in the federal government thinks about protecting their data and sharing it, an EMU type of arrangement might be a good fit for you.

You can come up with the list of 20 things that everyone agrees on and use the same words and use the same standards and use the same reference points and requirements so that when anybody within the federal government or interact - external interacting with the federal government try and engage the federal government to get that data, it's a much faster, clearer and well, better, understood process.

I collected as part of this admitted short turnaround assignment three different data sharing agreements from three different federal agencies and each has a different form, look, feel, but the Venn diagram of them, if I had better PowerPoint skills and time, and more time, I would have tried to create that visual for you.

The requirements are very similar and so what your, you know, ABC agency requires versus X, Y, Z agency requires, they really have the same core concerns at stake and if we can get the federal government to standardize their approach to data sharing this EMU approach might be a good fit for the federal government.

Man: Thank you very much, thanks to all three of you. Questions from the members of the Commission?

(Kim Lawland): I have a question. Hi, I'm (Kim Lawland), Commissioner. So, you've got this data sharing so if I was working on something in an agency but I needed data from education, data from agriculture so I have to go to each of those individual agencies to get the data, is that how it works…

Man: Well, yes, you will but it's not a in a physical sense, it's in an electronic sense.

(Kim Lawland): Right.

Man: We have an online portal that allows you to fill out a data sharing request that you can send to each of those agencies and explain to each of them what you're doing and we approach it like a FOIA request. If you're at the department of education and you want data from your sister agency if I was an external member of the public serving you with a FOIA request, you've got five days to turn that around.

This process and this set of agreed terms and conditions, agreed terms and requirements allows - and we've been using it successfully, it allows each agency to be nimble and quick in responding to you and saying, yes, you can have the data - you've filled out the purpose, I understand the purpose, we've already agreed on what the data protection standards are and we understand the duration and the other key components that are defining your request and we're able to turn those around quickly.

(Kim Lawland): Okay.

Man: So it's not in the bureaucratic or literal sense that you have to go to the agency, it's an electronic push right from your desktop and it could land back at you electronically.

(Kim Lawland): Okay, thank you.

Woman: But there's still human involvement in the sense that somebody is looking at the request and ruling on the request?

Man: Exactly. Yes, and they can decline the request and explain why and our organization has a mechanism for resolving disputes on why it should or should be resolved and we have the reliance on which statute you're basing your request is built into that system and so it should and so far with the couple of months we have under our belt, it has worked reasonably well.

Woman: Does the MOU parameters on when something would be appropriate for declination or is it just left to those types of considerations?

Man: We have a dispute resolution mechanism. I don’t - the level of the tail on what's reasonable or what's not - the dispute resolution mechanism determines when something is - how that process will work. There's a specific sort of hard and fast bright line rule. It's - I mean the data's world is so fast and nimble, so fast and ever-changing that we didn’t feel having articulating, you know, very clear bright line rules on when something is or isn't out of bounds would be helpful.

Woman: And is it the same person, same title at each agency who passes on each request or does that vary within an agency?

Man: The agencies get their - they choose who it is that has that responsibility within their…

Man: So I guess follow-up a little bit and then I'm going to turn to Mr. (Urlich), it sounds like every time you want - so (Kim) is interested in doing something, some project and so she needs data from here and here and so she puts in a request, so these are all of these one-off, I need it now, so then I get permission. Well a year later (Kim) wants to follow-up, extend it, do something - a slight variation - another request?

Man: We're very early in our process and so what I think the answer to your question would be is we would say the first time (Kim) makes the request she would say for the next year I need this data. If it's just turning around another quick five-day turnaround or less the following year to follow it up, it wouldn’t be that big of a burden for her to make that request again.

Duration of use in the ask is one of the parameters used in the initial request. So it is in theory possible for her to say for the next five years I'm going to be pinging you every January 1 for this dataset and then that might trigger at the receiving agency a light bulb or an idea that says, oh, that's - I understand your request, but for whatever federal reason or other reason I can only do it one year at a time. So they reach an agreement that says just come back to me each year and just put a reminder in Outlook to re-request it.

Man: And the second is do you - would you envision coming up with a similar relationship with the federal government?

Man: Why not?

Man: Well, yes okay…

Man: Why not?

Man: Because that is - you know, part of what we're talking about is comingling federal and state data.

Man: Yes. Why not? I mean, we're - I would imagine you could come up with 15 to 20 of the commonly accepted standards using the same terms, the same words, the same common form electronically or on paper that says if we're going to share data these - you have to agree to this up front and not reinvent the wheel for every single federal agency with which we interact.

Man: So Mr. (unintelligible) I wanted to follow-up on your sort of one-off comment about whether you - and it's something that we have talked about a little in the commission, you made the statement that it's hard to protect the identity of a 70-year-old, somebody who committed a crime in some local community.

I live in a reasonably - I live in Lexington Kentucky, not a huge community, but I get plenty of articles in my local paper about 70-year-old guys who did unusual things in small communities. Well, and you laugh a little bit but of course changes are good that that was written up in the local paper when that crime was committed.

You take that into account - and so you're worried about protecting the identity of these people for whom chances are good all of, everybody in that area, already knew about it, whereas the same guy could commit a similar crime in San Francisco and no one knows about it but you're not going to reveal that either, San Francisco or LA because that's not news in San Francisco or LA or it ranks far enough down the list that it doesn’t make the local newspaper.

Man: Yes, so I think what we're up against is, or what we're grappling with, is the idea of when you make a centralized repository available, widely accessible to anyone who wants to download it and then can run a set of scripts, for example, I couldn't because I don't get how these things work but I know there are very smart technical people.

((Crosstalk))

Man: Yes.

((Crosstalk))

Man: The scenario you might be worried about is we have this confidential data, we - but we are allowed to use it statistically so you anonymize it, you publish it and then someone very smart runs some script that crawls all local newspapers, matches the demographics of those newspapers to your centralized database, identifies, I don't know, 0.3% of a very large number which still ends up with a small number and then they sell that to Mugshots.com who does another nefarious things, then you're posting information right?

So like I think there are risks inherent when you have a very large dataset that we are just cognizant that as you - as the state, when you publish data, you have a sort of high degree of trust that the communities need in you and so we should just be very mindful of that. I don't know if that necessarily counsels that we could never do such a thing but rather that if we do so we should do so very thoughtfully and aware of the risks.

Man: I'm interested in the EMU that you describe but I missed on thing. So, what's the incentive structure? So if I'm an agency and I get one of these requests, what's in it for me to share my data?

Man: Well, the incentive is the reciprocal treatment you get when you're looking for data to cross reference.

Man: Let's say I'm never interested in anything other than my agency?

Man: That's why…

Man: I know that doesn’t happen in Illinois but…

Man: That's right, that's why we have a governor I guess, you know. The ultimate arbiter.

Man: So your point is that there could be consequences if an agency doesn’t cooperate, is that your point?

Man: Yes, yes. I mean, I - we're really, this is fresh ink, this is a 2016 initiative so to say I know the end game in every one of those disputes is probably a little optimistic but you know, you never want to go to the governor and say we can't agree. They'll figure something out. I can't yet conceive of an agency that's going to plant a flag so firmly on a data hill to say there is no way I'm giving you this without trying to reach some sort of compromise and working it out.

Woman: I guess I understand how that would work among the state agencies, you were talking as well about potentially using this kind of approach to say making data available to outside researchers. Do you have thoughts about what would make people within the state interested in sharing data for those kinds of purposes?

Man: Hopefully it will be just building on the success of the current internal EMU and some of it may require the moral suasion of the governor to say this is important, it's for transparency. I don’t - we're not there yet to know, and I'm not speaking on behalf of the administration in that regard but I can imagine there's a great benefit to society, to us being able to share that information externally in incremental steps and making sure there are proper safeguards and when appropriate, anonymity.

Man: Well and to the agency too in many cases, there - researchers can solve issues that the agencies can use to improve their services.

Man: Absolutely.

Man: In fact, I think Illinois involved in something like that right now in the child protection agency and their (unintelligible) foundations have national competition that's going to be based on those data.

Man: Okay.

Man: And I could imagine the direct use to the agency of - depending on what we find out.

Woman: That sort of echoes some of the things you were saying (Amy).

Man: (Amy), in the last couple of years we've had some, I think it would be fair to say tough even draconian legislation introduced in the house and the senate to restrict access to school records. Do you think - can you imagine provisions of privacy that would give you an argument as a person lobbying the congress and trying to talk congress out of doing this sort of thing to say, look, we've already got these protections and so forth. Is that going to be a useful thing if we have good provisions on privacy?

(Amy): Yes, and for the record, we don’t lobby. We do a lot of educating though but so…

Man: I meant to say educated.

(Amy): Good, thank you. My son just thanked you too. So, I think what's important is that why this commission and these conversations are so important is that it creates a safe space to have these discussions that aren't in the heat of legislative discussion and I think - so that's why I think this commission and the work you're doing and having the chance over 18 months to talk about it and provide recommendations to congress is unbelievably important.

The changes proposed to FERPA has actually been a fascinating process in the last three years and you've been very involved in it as well. That we actually think that the drafts that are on the table right now - that the draft that was released in September - goes a long way to addressing a lot of the concerns.

And actually one of the issues that's really at heart, it's one of my quick slides you saw, a lot of the technical issues and ambiguities about who can have access to what data, what could researchers have access to, how can data be shared across agencies was a big issue that FERPA, to be blunt, to tie this into the prior conversation.

A lot of agencies were using FERPA as the greatest example, the greatest reason why not to give anybody access to data, let's be blunt because it wasn't in anybody's interest, it's just more work and literally we had somebody say to us at one point, if you make FERPA clear it then I'll have to do more work and give people more information when they ask for it.

And people - it was just another thing to do by providing access to information even if it was legitimate. So when the regulations happened it clarified a lot of the legal pieces of it but it actually ironically caused almost more problems because it was seen as being gutting privacy, opening the doors to data in a way that because it was done by regulation and not in the public sector and not in the public sphere and the conversation.

So we actually think there's a lot to be said by having conversations in the people's (house), in state legislatures and in congress about this because it's a public conversation, it's not something done by FIAT and to regulation. And we actually think that because of the conversations we've been having and on congress listening to researchers coming up and talk like yourself, listening to stakeholders talk about this, that what's on the table actually improves and updates the law.

There's still some things that need to be, I think, fine-tuned in that draft around researcher access; there's some concerns there, and about the capacity and training issues and about how it applies to local data use a little bit more. But other than that it's a process of talking about it and listening to each other, we're actually going to end up with something and I also think it will take away, it will build trust, because people have had a chance to talk about this in the public arena.

Man: Other members of the commission? Well thank you to the panel, we appreciate it and we are going to have a 15 minute break so we'll be back here at 3:15. Thank you again to the panel.

((Crosstalk))

Woman: We'll go ahead and get started so we'll leave times for questions and discussion. We have two speakers who I'm very pleased are able to join us in this final panel for the day, final panel of speakers. Our first speaker will be Mark Rotenberg from the Electronic Privacy Information Center and then we will have Cynthia Dwork from Microsoft Research who is joining us remotely. So Mark if you want to go ahead?

Mark Rotenberg: Great, well thank you very much I wanted to thank the commission for the opportunity to speak with you today about privacy issues. I also want to think all of the data scientists, invariably it's the collection of personal information that leads to the creation of new privacy laws.

So we are very grateful for the work that you do. I want to say a few words about my organization and also my background. I will be talking a bit about two key topics in your mandate, the scope of personally identifiable information and also the concept of data minimization but I do want to begin by pointing out that like many organizations in Washington D.C., we do like to think of ourselves as an evidence-based policy organization.

In fact, we rely very heavily on government data, government reports for the advocacy work that we do, I'm going to use, in fact, as a case study one of the key government statistical reports of…

Woman: Actually, would you mind pulling the mic closer?

Mark Rotenberg: Yes, certainly.

Woman: Great. Thank you.

Mark Rotenberg: I'll also point out that as lawyers often in federal courts we file a type of brief known as a Brandeis Brief which relies heavily on data and not so much on legal argument or doctrine so we have a genuine commitment to the mission of the commission but of course a big part of our focus is the development of new laws and new techniques that enable the use of data while safeguarding personal privacy.

I don’t believe that you should approach policy mandates with a goal of trying to balance competing interest. I think that's invariably the wrong road to take. I think the much better approach is to achieve competing interest which is to say you need to find policy solutions that enable the use of data while simultaneously protecting privacy.

I'll point out also I've served on many expert panels over the years. I'm currently involved in a panel for the academies. It has a really long title and I apologize to (Bob Groves) so I'm simply going to say Big Data and Privacy. If you want to find the actual sort of multidimensional, multi-source data source, multi-(unintelligible) title, it's available somewhere online.

And I'm also on a panel for the OECD on risk assessment but the views I'm sharing today are really my own and they're not intended to represent those of the organizations I'm affiliated with. I would say that the work of the national academies in this field I think is particularly important and I hope you will give a close look at the preliminary report which will be out soon as well as the final report which will be out a little bit later.

I think that will be a very useful background. But let me dive right in with a case study to try to make concrete both the value of data and also how privacy issues play out in the debate over access to government information.

I have a particular interest in the legal authority for electronic surveillance in the United States which is probably a good thing because I run an organization called the Electronic Privacy Information Center.

And our organization pays particularly attention to a report that was mandated in 1968 by the U.S. Congress - data about the use of electronic surveillance by law enforcement agents in this country. It was one of multiple approaches developed by congress to try to establish some oversight and some accountability and I think I can say fairly that regardless of your views as to the use of electronic surveillance authority, the data is enormously useful to understand the issue.

Whether you're proposing expanded authority or limited authority or predicate authority or the significance of encryption as to the effectiveness of electronic surveillance invariably people come back to the provisions in this 1968 law. Here's a screenshot of the Web site of the administrative office of the U.S. courts which has dutifully, for almost 50 years, been publishing this data.

Here's a screenshot from my own organization where we have taken the data regarding federal wiretap, you probably can't read the small print, the blue line at the top is the total number of wiretaps authorizing the United States, the line colors, the state number, the red is federal, that line which is essentially at zero is the number of orders denied but this is a graphic representation of federal wiretap authority and just a quick take we'll show you that after 9/11 in the United States there was a significant increase.

Okay, as I said, I'm not here today to make a case for or against privacy but I do want to share with you a few insights about this particular dataset that may help you understand how we view access to public data as well as the competing concerns about privacy interest.

The first thing that I'll say that is actually quite important for us is the fact that this data is stable over time. The fact that the categories remain the same, the descriptors remain the same and the participants have remained the same for almost 50 years makes it enormously useful when we're trying to assess trends.

Now, this is relevant in part because when thinking about sources of data in use by the federal government, oftentimes they're proposals for private sector sources or voluntary reporting, all of which can be useful, of course, in my field we have now what are called transparency reports issued by private companies about electronic surveillance but none of them compare to the value of the 1968 (add) the methodology as to the data's transparent and the data we can literally go back to the administrative office in each of the judicial districts and verify how it's produced.

Now, here's where it gets interesting. It gets interesting because a valuable as this data is and as sensitive as this data is keep in mind we're talking about data associated with the prosecution of criminal investigations which typically implicate dozens if not hundreds of people, we're talking about the investigators and their families, we're talking about witnesses, we're talking about targets, we're talking about people associated with targets, this is sensitive stuff.

Nonetheless, the federal government has found a way to make this data available to the public so that groups like ours and others can make use of it in a way that's enormously important. And this is in support of my point to urge you not to think about balancing privacy against public access but rather to think about how do you make the data available and simultaneously protect the privacy interest of the individuals. Congress achieved it in 1968 in it continues to work 50 years later.

The other point, which many of you are likely aware of, the data is still enormously relevant. We are - we continue to be in the midst of a big debate in this country about the scope of electronic surveillance. Again, I'm not asking you to take sides, I'm simply asking you to see the value of the data that's produced on an annual basis by a federal agency that protects privacy and is useful to all of the participants. I would propose to you that this is actually a fairly good model for an evidence-based policy determinations.

Now, second case study which I'm going to point to quickly, partly because as some people know, it's related to a hobby but partly also it helps make my point about the use of public data that minimizes privacy risk and that is that the data that's typically produced by NOAA. This is weather forecasting, this is hurricane information, this is climate data, this is satellite imagery; enormously valuable to people, industry, farms, shippers, fisheries, all across the country.

Lifesaving functions, emergency functions are all supported by the access to NOAA data, yet again the remarkable point. There's hardly any privacy issue present in any NOAA database. Now I'm sure there may be marine life that's kind of upset with me right now, I'm not defending dolphin privacy but that's not really my point today. My point is to say that you can gather large amounts of data, make it available to the public, support foreign policies without jeopardizing privacy interest at least in some circumstances but we have an existence proof in these two cases for how that can be done.

Let me talk now a little bit about some of the key concepts before you and the commission and I mean I'm here with some of the superstars, Professor (Sweeny) and Professor (unintelligible) at enormous impact on the work in this field. These are my, again, my own views. They may vary a little bit from some of the commission members but let me just set them out.

The concept of personally identifiable information at least from a legal perspective is one of the key, core, concepts of modern privacy law. You just can't escape that. You cannot write a privacy statute, and I've written several, without trying to understand what the relevant dataset is that you're seeking to regulate and the most familiar way that we do this in the privacy world is by talking about personally identifiable information.

What is personally identifiable information? Well to answer that question I could give you 30 different answers depending on the law, the regulation, the jurisdiction, the year, the dataset, the user, the application, it is endless but if you want a simple way to understand PII I would suggest to you that the most robust concept is data that identifies or could identify a particular person. If it makes you comfortable to insert the modifier, could reasonably identify a particular person, you can do that as well. But you understand now with that rough approximation for what PII is, what essentially the data is that raises privacy concerns.

Now, when we're talking about the collection of PII we're talking about the obligations associated with the collection and use of personal data. This is typically known as fair information practices. The rights and obligations are necessarily asymmetric. If you have the data in your possession you owe the data subject rights just as if you're a bank and you have someone's money, they can't control what you do with their money. The security obligations and everything else are in the control of the bank and the rights of course, go to the individuals who have given up their personal data.

This is rarely about secrecy although I know that's the intuitive understanding of privacy protection far more often it's about the fairness of the decision making when educational data is taken from this source and that source and a determination is made about someone; was that fair, accurate and was it timely?

Here's a key point I will make for one more members of the commission. I recognize that the boundaries of PII have been increasingly permeable because of new techniques that call into question the robustness of de-identification and anonymization and I will come back to that.

But I would suggest to you for a variety of reasons, fairly well grounded in economic analysis, that if an entity has the ability to recreate PII from what is believed to be non-PII then it necessarily has the obligations for protecting the data even if at that moment it is not PII and you understand the point that I'm driving toward is that if you want to make the claim that you're not subject to PII obligations the burden is on you to establish that in fact your techniques can achieve that goal.

And if you cannot achieve that goal then you carry the obligations. I, along with a few friends, many years ago, put forward this concept of privacy enhancing techniques, it's different from privacy by design and some of the other concepts but we had in mind very precisely that these would be techniques that would minimize or eliminate the collection of personally identifiable information, the most familiar of course is cash or a debit card.

The point is that the technique must be robust scalable and provable. Now, we believe in the development of these techniques and no doubt you're going to be looking at a number of them but we're also quite willing to challenge ineffective techniques for privacy protection.

We had one interesting case before the U.S. supreme court a number of years ago involving the privacy of prescription records and that issue in that case was the privacy of the prescriber who had made available prescription information to third parties and we looked a little bit more closely and we said, wow, the patient data is actually being hashed with MD5. MD5 is no longer a secure technique.

We wrote a brief and we got Ron Rivest who created MD5 and (Bruce Schneider) who busted MD5 both to sign onto the brief and said it was actually a greater privacy interest here than people are aware of because MD5 is no longer a reliable technique to safeguard data.

EPIC has brought complaints also regarding AskEraser, it said it had a search technique that would delete all searches, it didn’t. SnapChat said it had a technique that would make photos vanish, it didn’t. These are (paths) that don’t work which is not to say that there couldn’t be (paths) that could work.

I'll go quickly through this. I know there's a lot you'd like to cover. There's a genuine risk in the collection of personally identifiable information. Data breach, identity theft, financial fraud according to the Federal Trade Commission, the number one concern of U.S. consumers over the last 15 years has been identity theft and the risks now, of course, are expanding. Those are risks to institutions as well as to data subjects.

Quick word on data minimization. As I've said, I've written a bunch of privacy laws, this was one from the 1980's but back then we were actually struggling with the privacy of video cassette rentals. Does anyone here know what a video cassette tape is? A few hands went up.

There was one last night in an episode of Mr. Robot they did that as a privacy technique, so an analogue recording mechanism, it was very clever, I was taking notes. But you can understand the sensitivity, you know, around what people view and certainly the retention of that data and in protecting the privacy of the type of information that people had access to, there was a provision in this act that explicitly said if you're in possession of this data, it should be destroyed after it's no longer reasonably necessary.

Now, of course you'll find other provisions like that in other laws and regulations and it always gets the data scientists a little bit on edge like oh my goodness, bonfire of data. But there is a real concern here about the misuse of personal information.

All right, I'm going to wrap up here with a few just broad thoughts. I think one of the points that obvious today about the collection and use of data aside from its value, when you're talking about personal information it is truly increasingly dynamic.

Increasingly difficult to control, it's difficult to anticipate outcomes but the corollary, of course, is that it's difficult to assess risks. So the people who are driving the debate over big data and anticipating new outcomes that result from emerging datasets to be sure, I mean, those outcomes will occur but oftentimes I ask the question, to what extent has risk been assessed in trying to evaluate what the downside may be in some of those enormous data matching projects.

The other clear trend, I think, is that datasets are increasingly under attack from malicious act and this is important to keep in mind, I think it requires a certain amount of humility on the part of the agencies and the data scientists that may have the very best plan in mind and may, in fact, have a project with very little adverse outcome for the members, the participants, in the dataset to nonetheless have to recognize that there may be people out there who don’t share your view of the world.

And if they get access to the data you are now responsible for how they use the information about those people who's data you've collected and this is a new dynamic in the privacy world, I don’t think it's been adequately understood. As you know, certainly there's increasing focus on AI and data analytics, the Stanford Group just released their 100-year study on AI. The Whitehouse report on AI is expected soon to be sure all of those new models are going to be driven by the enormous data repositories that are available in both private sector and public sector organizations.

And you almost need to anticipate as you're thinking about the future use of federal data not only the present uses that look familiar but the new uses that are not yet well understood.

We increasingly are of the view that when this data is merged in the analytic world the transparency of the decision making becomes almost as important as the privacy protections associated with the collection use of data.

So just to close here, I mean, there's no dispute and certainly in the important work of your commission is to recognize that data really is the basis of research and innovation and growth and in foreign policy decisions but it's also the basis for profiling, tracking, segmentation and discrimination and I think ultimately the privacy issue is best understood as trying to maximize the benefits and minimize those risks. So I will stop there.

Woman 3: Thank you very much. Are we set up for Cynthia to join us?

Cynthia Dwork: Yes.

Woman 3: Oh, hello.

Cynthia Dwork: Hi.

Woman 3: Great. Thank you, I will turn the floor over to you. Can you see us?

Cynthia Dwork: Yes I can.

Woman 3: Okay, we can see you.

Cynthia Dwork: Great, okay. So good afternoon. I'm grateful for the opportunity to speak to the commission and I thank you for inviting me and making it possible to participate remotely. I will speak about differential privacy.

A definition of privacy and a collection of supporting algorithmic techniques tailored for privacy preserving statistical analysis of large datasets. Differential privacy is a mathematical guarantee that an individual data contributor will not be affected adversely or otherwise by allowing her data to be used in any study or analysis no matter what other studies, datasets or information sources are or will become available.

At their best, differentially private algorithms can make confidential data widely available for accurate data analysis without resorting to data clean rooms, data usage agreements, data protection plans or restricted views. Nonetheless, data utility will eventually be consumed. The fundamental law of information recovery states that overly accurate estimates of too many statistics can completely destroy privacy. The fundamental law can no more be circumvented (can the) laws of physics.

Every useful computation results in some loss of privacy. It yields some statistical hint of private values and these hints accumulate. Differential privacy measures and controls privacy loss accumulating over multiple analyses. This signal capability makes it possible to program in a differentially private fashion.

So in ordinary, non-private computation, anything computable can be computed using only addition and multiplication but that is not how programmers work. Algorithm design is the creative combining of appropriate computational primitive to carry out a statistical computational task while minimizing the consumption of key resources such as computation time.

Similarly, differentially private algorithm design is the creative designing of simple differentially private primitives to perform a sophisticated analytical task while also minimizing privacy loss and inaccuracy. As a rule, when the dataset is large, the signal dominates the noise injected for privacy. When the dataset is small this is not the case. This is correct, think of the case of the dataset of Size 1.

To ensure privacy, the noise must dominate the signal. Designed to preserve the privacy of everybody, even the needles in the haystack, the goal is to elicit participation in a study without fear of repercussions or a public good such as learning that smoking causes cancer and other facts of life. Indeed, it is often the outliers who most need protection.

Differential privacy also provably controls privacy loss accruing over computations on multiple possibly overlapping datasets making it especially relevant to the kinds of analyses that will be needed for evidence-based policy making. Now the fundamental law tells us that meaningful privacy guarantees come at a price. Other disciplines such as ethics and economics cannot be brought to bear nor can actionable policies be articulated without a measure of privacy loss. Differential privacy provides such a measure.

Finally, differential privacy strengthens the scientific method in an unexpected way even when privacy is not a concern. The rise of big data has been accompanied by an increased risk of (unintelligible) scientific discovery. A great deal of effort has been devoted to reducing this risk from the use of sophisticated validation techniques to deep statistical methods for controlling the false discovery rate in multiple hypothesis testing.

However, there's a fundamental disconnect between the theoretical results and the practice of (data) analysis. The theory of statistical (hand prints) assumes a fixed collection of hypotheses to be tested selected before the data are gathered whereas in practice data are shared and reused with hypotheses and new analyses being generated on the basis of data exploration and the results of previous studies on the same dataset.

This leads to over-fitting, that is learning about the dataset rather than about the population from which it is drawn. Differential privacy automatically protects against this source of false discovery. Differential privacy holds great promise but requires great effort. The fundamental law firms the imperative for high quality differentially private algorithms but the field is young and many of these will be the content of doctoral dissertations not yet written. The literature is silent and crucial data preprocessing sets such as imputation of missing fields and other aspects of data cleaning. Working with formal privacy guarantees, requires a new skillset foreign to most statistical agencies, social science researchers and data scientists?

Recent adoption of the approach by Google and Apple will draw talent away from the public and research sector. Nonetheless, I see no alternative to formal privacy guarantees when an individual can be placed as the scene of a crime or identified as a member of a case group in a medical study based on overly accurate fractional protein counts from a forensic DNA mix.

So I want to close with three policy recommendations; first, publisher Epsilon. So differentially private algorithms are equipped with a privacy parameter usually called Epsilon capping their privacy loss. In a non-private algorithm, Epsilon is infinite.

By maintaining a registry of privacy loss akin to a toxic release registry we can observe the accuracy/privacy tradeoffs actually made and stimulate competition to obtain better analyses at lower privacy costs. In those - engaging those who traffic in the data of individuals in the effort to protect their privacy.

Second, establish a list of approved private data analysis techniques and appropriate applications and keep it current. Third, consider restraint. In a data rich world the challenges revolve around the tradeoff between what can be done and acceptance of the fundamental truth that overly accurate estimates of too many statistics can destroy privacy. If we are interested in privacy, sometimes restraint might be the right approach.

I'm done.

Woman: I think there will be…

Cynthia Dwork: I can't hear you.

Woman: Sorry. I'm sure there will be questions for both speakers and I think Paul wants to jump in.

Paul Ohm: Yes and I'm not going to try and come up with one grand question for both of you, hi Cynthia, it's (Paul Ohm), how are you?

Cynthia Dwork: Hi Paul.

Paul Ohm: So instead I'm going to ask you each a different question. So Mark, and I already gave you a heads up about this at the break, I think an elephant in the room is the 1960's and 1970's proposal for a federal datacenter and that by historical accounts kind of backlash that occurred ultimately leading to the Privacy Act and other things. We haven't really had a discussion or an account about why that happened, how it happened and perhaps how it could be avoided, history repeating itself.

Cynthia for you, this is funny, it's a question I never thought to ask you before because your paper takes such pains to define things in such a kind of formal manner, would you go so far as to say like differential privacy is the broad umbrella definition that encompasses everything god in the world that's ever going to come?

So in other words, the work of the commission could be really quick if we write one sentence that says, use differential privacy techniques, nothing else will do, period. So as kind of one of the lead progenitors of that would we be making - we'll be making some kind of mistake if we do that but would it be an enormous mistake or is it kind of right but we should be a little more careful about it, if that question made any sense?

So I don’t know which of you want to take on this first?

Cynthia Dwork: Since Mark got the heads up, I'd like him to go first so I can think about your question.

Paul Ohm: Okay.

Mark Rotenberg: And thank you Paul, it's because of the heads up that I made the reference to thanking the data scientists for the advancement of privacy protections in the United States.

But the history is relevant. I mean, it was in 1965 that there was a proposal led by social scientists, by the way, for national datacenter and it came about in part because of changes in technology much like today and people anticipated that with the increased automated data processing of the personal information that was being collected by federal agencies, it would now be possible to bring together stores of information that could not previously be brought together to make government work more efficient and to identify new activities.

Does any of that sound familiar? Okay, so that was actually the early 1960's culminating in the proposal for the national datacenter 1965.

And it did, in fact, create an enormous backlash. Now, there's a lot of history about the period even just through the lens of privacy and there are other things taking place in the United States that contributed to public concern but you see throughout the 60's Alan Westin publishes Privacy and Freedom. Vance Packard publishes The Naked Society. By the way, a New York Times bestseller the year after the national datacenter proposal was made and congress begins holding hearings and by 1974 you get both the Privacy Act and FERPA.

What does the Privacy Act say at its core? It says keep the data in separate agencies. It says we do not want a centralized data repository in the United States and it creates a whole bunch of restrictions and procedures on what we now call the sharing of personal information across public agencies. Although, by the way, you'll never see that word in the privacy law.

Where are we today and what do we take away from the National Datacenter experience? It feels very familiar. I think the risks are still there, I think the passage of the Privacy Act in 1974 was probably a good thing for the United States.

It actually enabled trust and confidence in the adoption of new technologies by the federal agencies at a time when many people were simply concerned about the automation of personal data. So it was a legal regime that restricted the use of data but at the same time enabled the use of data and incorporation of new technology. So I can say more but that's the quick answer.

Cynthia Dwork: So I took all of that time to come up with an answer of yes and no. So, if we said yes, things would get very private but everything would come to a standstill. My answer would be you have to try.

So trying, trying to be differentially private will first of all eliminate a bunch of stupid mistakes that people make. If you try and you can't do it, you should explain why. Maybe it's because you need an algorithm, you need to do some kind of analysis and that Ph.D. thesis hasn't been written yet. Spell it out, somebody may work on it.

If you have some use of the data where it's fundamentally not statistical, okay, then explicitly say this is not statistical. If it is statistical explain why what you're doing has to depend so much on the data of a single individual because differential and privacy works well when the data of any single individual doesn’t make too much of a difference.

So I'll stop there.

Woman: Thank you. (Bob)?

(Bob): So I don’t fully understand this differential privacy concept yet and I'd like to read something about it. But I guess my question is, is your presumption in this differential privacy world or algorithm you're talking about that everyone plays by the rules?

Cynthia Dwork: There are many different parties involved. So when you say everyone do you mean the people holding the data, do you mean the people who are asking questions of the data? Who do you have in mind?

(Bob): I'm not sure who I have in mind but my - intuitively what I'm worried about is you create a perfect system but there may be folks out there for whatever reason who don’t want to play by those rules in which case you end up sacrificing privacy.

Cynthia Dwork: Right, so differential privacy has its roots in cryptography which means that we view the data analysts as an arbitrarily malicious questioner who has no goal in life other than to compromise privacy and who has access to arbitrary information from other sources than the dataset including often a lot of detailed information about the data themselves. So, the algorithms protect even against such a malicious and informed adversary. However, let's say that the algorithm is being run by the Census Bureau. If there is somebody in the Census Bureau who deliberately miscodes it or, I don’t know, the algorithm really has to be the algorithm. There has to be a good guy somewhere in the picture.

Woman: (Ken)?

(Ken): I guess I have two maybe questions or comments and building on what Paul said but the conversations we've heard. So there was this effort in the 60's to create a national data statistical agency and it failed here. But we've got lots of evidence from around the world particularly Northern Europe where those efforts succeeded and I think have succeeded fairly well.

We seem to be fixated on the U.S. attempt that didn’t work in the 60's without sort of looking at the attempts in other parts of the world that have succeeded fairly, I would say fairly spectacularly but that may be a bit strong but have succeeded in ways that have created national databases that have combined information.

You can tell me those people in Northern Europe are different than us but actually they're not to - I would discount that the fundamental difference in those human beings from the people in the United States.

And so I guess can we look there and draw conclusions? That's my first question. My question, I'm like (Bob) in I'm not familiar with the algorithms that you're suggesting we apply but it does seem as if part of what you're tasking us with is to prescribe to people that are using the data how they analyze the data. I would view that - and maybe we want to go there but I would view that to be beyond the call of what we're supposed to be thinking about but maybe that's appropriate that we sort of push and say, if you're going to use these data you have to do so in these manners. Am I missing something?

Mark Rotenberg: Let me make a couple of quick points. I think those are excellent questions. I've actually done a lot of comparative work in the privacy field and we also work closely with the OECD which itself is the provider of many useful statistical datasets that are comparative across country.

One of the different, well - multiple differences between the EU and the U.S. cultures on data use, one, is that Europe has a robust regime for data protection, right? So there's actually somewhat less of a risk that data that is in the statistical realm will migrate into the commercial realm or the law enforcement realm as it could in the U.S.

The systems, I mean, the lines are not bright and clear but there is clearly a difference in terms of legal protection in Europe that provides more assurance that statistical data will be used for statistical purposes. I think the experience in the U.S. is different.

And if I could make a quick point actually in support of Cynthia's work, I mean we do see differential privacy as one of the several important new techniques that enable the use of data while protecting privacy. Now, it does come with some tradeoffs as you say, to give researchers access to a dataset for limited queries over limited time period may actually be the tradeoff to ensure the privacy will be protected but we think increasingly at some point to think about data in that way because the alternative is to say, oh, here's all of the data and you can do with it whatever you want which from a privacy perspective probably won't work.

Cynthia Dwork: I'd like to ask you what you meant when you talked about constraining how the analyst asks questions? So let me just give you an example. Suppose a data analyst asks the questions on how many Microsoft employees have the sickle cell trait. That is one gene for the sickle cell disease and they get the answer and let's say the answer is 298 - I've just made that up.

So this quantity feels nondisclosure, it feels like it's not compromising anybody's privacy at all but if they then get an exact answer to a second question which is, say, how many Microsoft employees other than distinguished scientists with very curly hair have the sickle cell trait, then my sickle cell status could be deduced.

So, should we allow this kind of combination or not? It - do you think it's okay to control? Do you think it's okay to allow, to control and say no you can't ask for the sickle cell status of Cynthia? Some control clearly is warranted.

Now what happens with differential privacy is that it's handled automatically. You don’t have to think about what do these questions mean but in some sense the - yes. Okay, so we need some kind of control but it needn't be explicit. It can be captured by the definition of privacy.

Woman3: Does that answer?

(Ken): Yes, I mean it does. I was just trying - I wasn't saying right or wrong or yes or no, I was trying to figure out what was being suggested so that I could better understand that, so…

Cynthia Dwork: I mean, for the most part anything that people think about computing on a dataset that should be statistically useful we will hope has a relatively good high quality differentially private implementation that gives an answer that's accurate as long as the dataset at least is large enough while having minimal privacy loss.

Mark Rotenberg: Cynthia I think you also said that, correct me if I'm wrong, that at this moment in time the development of a near-universal suite of analytic tools employing differential privacy is not there, one can imagine it, so in a way to (Ken)'s point, there's a statistical computing cultural change that has to go on before people like you would be comfortable, I think. So you'd switch software, you do different things like that.

((Crosstalk))

Mark Rotenberg: Yes, it doesn’t mean that you change the statistics you'd be trying to estimate.

(Ken): Got you.

Cynthia Dwork: That's true. It may end up that you want to consider some new statistics for various reasons. By the way, I know that (John Aboud) submitted to the commissions the text of his Shiskin Memorial Lecture.

Woman: He's actually here.

Cynthia Dwork: Oh great. So, I think this is a good place to look for some of the - for the answers to some of these questions about getting from here to there.

Woman: We can circulate the text of (John)'s talk to the other members of the commission.

(Hilary Coins): I wanted to - hi I'm (Hilary Coins), commission member. I wanted to ask both of you to talk, to respond to the sort of thinking about sort of issues around tiered access and about differences in statistical analysis of existing government data by folks who work in agencies versus, I guess, what Mark talked about maybe in the first panel, the use of the term sandbox gated protected setting such that has been developed at census over the past decades and I wondered what your thoughts were about that in relationship to the issues that you're raising?

Mark Rotenberg: It's really not something I'm very familiar with so…

Cynthia Dwork: Nor is it something that I'm familiar with. However, one of the things that - one of the things that I would hope is that - I don't know really how you define what a researcher is. I understand how you define what - how you distinguish between somebody who was at an agency and who isn't but I don't know how you define a researcher and so I would - part of the motivation in differential privacy was to make things widely accessible without having to decide who is a researcher, who is a bonafide researcher.

The sandboxing clearly is intended for people who have credentials who feel can somehow or other be bound to good behavior, who maybe would suffer repercussions like losing their job if they misbehave with the data but even things reported out by people who have seen the data in a sandbox could accumulate and cause privacy loss.

So I'm not quite sure - I see the necessity for allowing people to see data in detail in certain circumstances but I don't have a general formula for it.

Paul Ohm: So I feel like I - that almost felt like a setup to me, my earlier question. I didn't mean it to be a setup Mark, so I asked what happened in the 60's, what was the response, what lessons could we learn? Is it avoidable now and I think you ended right before the last part.

And maybe that's your answer but rather than just ask the same question, you know, one thing I've been thinking about and talking to commissioners about is the way the privacy community tends to notice things, respond to things, interact with proposals so if you could just give us guidance on the care and feeding and privacy advocates, right? So what are the kind of best practices for how to interact, when to interact…

Mark Rotenberg: Probably less expensive than the care and feeding of academics.

Paul Ohm: But if you could - in a somewhat abstract way, tell us like here are the rules of engagement that will at least get you in conversation at the right time…

Mark Rotenberg: I mean, Paul, I mean look, in fairness I think that's kind of a disparaging comment. You know, privacy is a concern that's widely shared in the United States across all demographic groups, age groups, political groups, it is remarkably the one issue that seems to unite much of the country when so many others should seem to divide us.

So the first word of advice I would give to you is it would be a terrible mistake to think in terms of your work as trying to protect against criticisms from privacy groups. That’s almost exactly backward.

I think you have a fundamental responsibility to the American public who the mandate and the act of congress to come up with techniques that enabled the appropriate use of public data collected by the federal government in a way that protects private information. It's in the law, it's not something I just said and that really should guide your work.

Now, as I've tried to suggest to you, I really don’t believe there's a necessary tradeoff here and I think if you go down that road, that will be a mistake as well. I respect what Cynthia has said to the extent that I understand it with certain techniques you may need to adjust a dial for your Epsilon to decide how much privacy risk you're prepared to accept for a particular dataset but if viewed broadly, as a policy challenge, you really need to maximize both goals.

Now the other thing which I haven't said so far but since you push the point a bit, there is a great deal of sensitivity still in the United States today about mass surveillance and it's been given renewed interest because of what the American public learned about the activities of its government over the last few years so this is not the 1960's anymore, this is, you know, post-2000, June 2013. And again, I think the public and the members of congress would like to see solutions that don’t put them, again, in the position of trying to defense mass surveillance of the public by its own government.

Paul Ohm: Yes, thank you.

Woman: Let me thank you both very much for taking the time to come and speak with us. You've given us a lot to think about, I appreciate it and I'm sure others - I'm sure the other commissioners do as well.

Let me turn the floor over now to (Bob Groves) and (Bob Han). Okay.

Man: Thanks, can you hear me?

((Crosstalk))

Man: So I thought this was a fantastic set of panels. My head is spinning sort of just trying to digest what I heard in the last panel from both our speakers. I'm going to take a trick out of the Passover Haggadah and try to simply raise a few questions as academics want to do sometimes and I'm not necessarily going to provide answers for these questions.

But I think some of these questions at least are high level and worth us considering and worth the individuals assembled here considering and they're not all my questions, they've come up during the day and I'm liberally borrowing with that attribution from contributors.

So one question that was asked earlier was how do we consider addressing data access in a world of non-zero risk? And I really think that we need to come to grips with that. A couple of representatives of the administration suggested that privacy, and these are my words, privacy isn't a roadblock, maybe an opportunity.

We certainly can have innovation while protecting privacy. We need to be very sensitive as you pointed out for the risks that associated, that may be incurred by individuals. Others pointed out that we need to be sensitive to the risks that may be incurred by agencies.

Another, I think, important insight was that all data aren't equal in terms of their sensitivity. A central question, and this is perhaps because I come out of the economic cathedral is how should decision makers think about this risk benefit balancing and I think I may differ with you Mark a little bit. I think at some point tradeoffs inevitably have to be made, you can't necessarily maximize two things at one time unless there are inefficiencies in the system.

We heard about an example today about the potential misuse of criminal records. I think that's a real concern, sort of illustrative of the kinds of things that can happen. Mark you raised several good illustrations in your talk as well.

A second question that I think is important for us to grapple with and we raised at some other points in the conversation is how much better can we do or can we be doing than we're doing now and what needs to be done to actually get there from here? We heard in the earlier panel today about, I'm not going to pronounce this correctly, but EMU or E-M-O-U, what's the potential for vehicles for like that to address some concerns?

The third question is what might realistic solutions look like? And I would love Mark to get feedback from groups like yours in thinking through some of those issues and finally since we're all about measuring and evaluating things, how do we go about measuring and evaluating our own success in this area or the likely efficacy of whatever recommendations we come up with respect to either privacy or lower the transactions cost of obtaining data for socially useful reasons.

So let me stop there. I've avoided offering any constructive answers and I'll leave it to my distinguished colleague down the aisle to try to take a stab. Thank you.

Man: So he's not here and I'll give it a try. You know I think it's not by accident that the commission spent one of its first days on privacy and I think the work that we did today, the discussions we've had, the presentations we've heard really are sobering or sobering reminder that as someone said, we really need to shift from the central focus on (bringing) up data for more analysis to one of making the headline, how do we take advantage of new developments in privacy protective devices in order to achieve greater information flow from data.

And I think that shift is probably important and I think relative to the (interchange) you guys just had, the interesting question is can we not revisit 1965 by leading with, we want to take advantage of new developments in privacy protection and those new developments will permit a wonderful new data world.

We've experienced a lot today, we had a 100-year history from Paul of the meaning of the word privacy and it's clear that that simple word that we use in everyday speech is not - first of all, is dynamic in its meaning over time, heavily contextually dependent, complicated because of changes in technology, something that we can't rather flippantly apply as if we know what we're doing.

I thought what I learned today was how we need to think of risk of harm through privacy breeches as not a static phenomenon in time, space or subpopulation that what is risky today may or may not be risky tomorrow and vice versa so that the notion that I think we've had for some decades that you can sort of have these agreements and you can have data of a certain character and that solves the privacy problem forever more is probably not a wise one to take going forward.

And then I thought today also taught us that we're a pretty diverse society that for cultural and historical reasons and for variation in just understanding of data, we have to be sensitive in the work of the commission to regard those differences with some seriousness.

You know, the key question we're facing as a commission I think is can this country build a safe and sustainable environment that utilizes modern technical developments to protect the privacy of people who's data are kept and at the same time permit the evaluation of the governments performance through the use of already existing data.

That's a tough nut, it's not something that for which there's a simple model sitting out there where we can just pick it up but I think that's our task. It's complicated by the fact that risk - some of the techniques that we know approach risk as something that can be easily measured or that can be measured but we're so aware, I think, that it's not just the real risk that we're talking about but the perceived risk that we're talking about and indeed perceived risk may trump real risk when it comes down to it. So, there's work for us to do, I think, on that relationship between real risk and perceived risk but are we going to - once again goes back to 1965, I think.

If you read the - there's just a brilliant transcript of the House debate of the 1965 bill where it's clear that many of the risks were all perceived risks. It was actually quite difficult to produce the harmful event that was being feared if you wanted to but that trumped the day, that took the day and so we have to worry about that.

Let me just end by a note of optimism I guess. So, I can't believe a better time in our lives to take this on because we do have new things. The developments on the technical side didn’t exist several years ago. I think the sophistication of data users and the devotion for data ethics and their own behavior is new, wasn't there before. These are two good things that we ought to build on to see if we can crack the nut and it's a lot of work to do but I hope we can achieve it.

(Ken): I just wanted to make one - you said a comment on what, something (Bob) said because earlier today you talked about having a (parado) improving solution. Now I know it's been a long time since I sat in a theory class but I think (unintelligible) told me the only way I can get a (parado) improving outcome is if there were inefficiencies in the system.

So, in some sense you're - and when I… I view everything as an economist because I've been doing it for so long I don't know how to think anything else but that's what I view (Latonya)'s presentation that you weren't here for today is that there are a lot of inefficiencies in this system and in some sense one of the things that I think about is if we somehow bring these data together and in a common area we can achieve some of those efficiencies, improving the efficiencies out there.

I think in some sense, my view is the distributed world we live in produces a lot of those inefficiencies and we can improve efficiencies and reach a (parado) improved outcome. So your question seemed to be in conflict with something that you had said earlier, that's just - in my opinion.

Man: So I agree with the sentiment of your remark and maybe I misspoke in my question, clearly if there are opportunities of the kind that (Bob) suggested to take advantage of new technologies or new knowledge or to take advantage of ways that the wall was written 25 years ago that may not be relevant today, I think that this is a golden opportunity for us to go after it and try to be more sensitive to privacy needs, protect them better and to also to the development of new knowledge that would be socially beneficial. So I agree with you (Ken).

Woman: Mark, did you have a comment?

Mark Rotenberg: Well, I wanted to respond to (Bob)'s summary if I could. I'm not an economist but I did marry one and when you're - I think when you talk about tradeoffs as an indifference curve let's say between privacy protection and data use, one of the first things that you learn is that if you're trying to increase both what you're really trying to do is move the curve outward and that typically occurs through innovation. So, my comments were actually partly inspired by that view of what tradeoff is really about, how do we move the curve outward to improve privacy protection and data use?

Now, again, we have a good existence proof of that in my field which is in communications privacy which is to say before we had public key encryption it was difficult to assure communications and privacy protection. I mean, you could send stuff anywhere you wanted but anyone could open your envelop or listen in on your electronic channel.

Once we had private key encryption we had a technique to enable robust and widespread encryption that literally moved the curve outward in the tradeoff. And I look to people like Cynthia and others and (Latonya) for that kind of insight that makes whatever tradeoff might occur, you know, viable.

But, part of this is also about economic incentive and this is why the concept of fair information practices and personally identifiable information actually fit together because they create a liability regime that puts the responsibility on the data collector to ensure the adequacy of the privacy protection. So I'm really - I'm speaking to you partly as a privacy advocate but also someone kind of thought about these issues through an economic lens.

How do we create the structural incentives and the regulatory incentives that enabled the use of the data so that people's privacy is protected. That would seem to me to be a very worthwhile goal for the commission.

Man: So I thought (Latonya) - here's an example. You tell me whether I'm right. So, what's our world right now? It's a world where our data use is being limited because we haven't been able to crack the nut fully on a set of techniques that would allow conjoining of a lot of different datasets and indeed that same system has its own privacy weaknesses in it.

So, if there exists a set of new analytic techniques that preserve the privacy at equal or better levels and permits the same analysis, then the use goes way up. It seems like you're out there in that level.

So if you talk about a single dataset it's a little harder, I think, to make the case but Cynthia would make the case I think. But that seems to be a clear example to me. That would be a much better world both on privacy and on data use, yes.

Woman: Well, I think we have been given a lot to think about. I don’t think we're going to be able to fully synthesize all of this on the fly and walk away with final conclusions today but we made some progress so unless there are things that other Commission members want to say? I would - I will thank everyone for being here today and we will adjourn. So…

END