PCAST Meeting – April 4, 2014

Welcome & Antimicrobial Resistance Report

>> John Holdren: Let me welcome, first of all, the members of PCAST to our bimonthly meeting, and public session. Let me welcome the members of the OSTP staff, the staff of the Science and Technology Policy Institute and the members of the wider science and technology community who are joining us. If any of our speakers from later in the morning are already in the room let me welcome them although I will welcome them again when that session begins and let me welcome those who are joining us on the webcast. We have a busy agenda today. We're going to be hearing progress reports on a couple of very important studies that are underway. I emphasize that these are progress reports. They're going to focus on the questions that PCAST is addressing at the request of President Obama. We will not be presenting recommendations and we will not be looking for PCAST approval of the reports at this juncture. That will come at later meetings, possibly in May and if not in May in July. The first of those reports is on antimicrobial resistance, antibiotic resistance, and I think the presenters on that are going to be some combination of Eric Lander and Chris Chyba who have been leading that study I'll turn it over to Eric first, since he is also my co-chair of PCAST and may want to offer his own welcome before turning to the antibiotic resistance report.

>> Eric Lander: I want to offer my welcome to all the members of PCAST who have been working very busily. It's been a very busy period and to everybody here in the room and on the Web. Our first order of business is to discuss a little bit about this antibiotic resistance report that we are beginning to prepare. As John has already said, we're not in a position where we have recommendations but we have a parsing of the important topics to consider and the questions on our mind, and I'd like to describe those a little bit, see if we have, I'll then turn to Chris Chyba the co-chair on this particular study, ask him to elaborate on this. If PCAST has questions they'd like to raise about it, engage in some discussion about it. It's really pretty hard for people today to imagine what the world was like at the beginning of the 20th century, when we didn't have antibiotics. Just the stunning rate of deaths, deaths of women in child birth, of children in their first year of life. It's a world that in the developed world we can't relate to anymore, and a world that changed very radically with the discovery and then commercial development of antibiotics culminating around the 1940s and then into the '50s with very expensive antibiotics. The problem, of course, is that antibiotics, agents that I mean by that agents that kill bacteria and fungi are agents that are in a constant battle with those organisms and the organisms themselves have a way to respond. They mutate and develop genetic changes that make them a lot more able to grow in the presence of antibiotics or a little more able to grow, and eventually overcome our antibiotics. There is probably no way around this. This has been an evolutionary battle going on for millions of years. The first antibiotics, of course, were invented by microorganisms fighting other microorganisms and the idea that this has been an arms race for billions of years is really very clear. We're now in on that arms race against microbes and we do not expect there's ever going to be permanent victory in that because there's an awful lot of microbes. And if one of them develops a mutation that resists, it can spread through a population. Nonetheless, we can have continuing victory, if not permanent victory, as long as we get a few things right. It has to be the case that the antibiotics that we develop, we don't lose them to resistance too quickly. We have to manage that rate of loss to resistance. It will probably be

inevitable over time, but if we proceed properly, if we steward the use of those antibiotics we can stretch out their lives longer and longer. And then we have to increase the rate at which we are producing new antibiotics. If we get the balance right -- we don't lose them too fast and we produce new ones fast enough, we do well. It means we always have a supply of antibiotics. We also need surveillance. We need to be able to look at individual patients and what they've got but more broadly across a population and understand how is antibiotic resistance developing and spreading. We have to be able to manage this fight back and forth with microbes. We haven't been doing very well lately. The rate of loss of antibiotics that is to say the development of resistance has been increasing. Right now it's the case that we have about 2 million people who are infected with antibiotic resistance organisms across 17 different types of antibiotic resistant bugs recognized by the CDC as important issues. There is an estimate of some 23,000 deaths per year due to antibiotic resistant organisms and that doesn't even include additional issues like deaths due to C-difficile an organism that outgrows other bacteria when people are treated often with antibiotics. It's a really serious problem. On the other side of the scale the rate of production and approval of new antibiotics has been falling rather precipitously. It went from perhaps 20 years ago, four new antibiotics being produced per year to over the period of 2008 to 2011 averaging less than a third of an antibiotic being produced per year. Not very good. There are reasons for all this. We don't really have great stewardship programs comprehensively in place across this country. Stewardship both in the healthcare setting and also stewardship in agriculture where 80 percent by bulk of antibiotics are used. We're also are losing the economic incentive to develop antibiotics to biopharmaceutical, a startup or established company it may be rather unattractive economically, obviously nationally important but economically not such a compelling proposition to develop a successful antibiotic to treat people with resistant bacteria when that might be a portion of the people with bacteria and you might succeed in treating them in two weeks meaning while it's a medically wonderful thing it might be an economically hard proposition to justify tens of billions or hundreds of millions dollars of research into if you can't figure out how to make the money back. So we have this economic issue. We don't really have the kinds of surveillance systems we would like technology is happily making new opportunities available to us, and there are things that can be done on the high end of technology. But we lack a lot of the basic infrastructure of the health system and we certainly lack the ability to bring in this high end technology. And then there's always a question when there’s a complex topic like this that cuts across so many domains from health to agriculture, to industry. Are we coordinating optimally? These are basically the questions we are asking ourselves as part of this study with regard to stewardship of existing antibiotics in the health care system. There are fabulous models of antibiotic stewardship programs, programs that have best practices for diagnoses and prescription and dosing and duration and data collection. The kinds of things that will minimize the rate, not abolish the rate but minimize the rate, at which bugs will become resistant. A good example is when someone comes in for a respiratory tract infection, they may want you to give them an antibiotic but the odds are pretty good they have a viral infection and that the antibiotic, the anti-bacterial is not going to do them any good and actually may cause harm, there may be harmful reactions to it. It would be very good to make sure you were treating appropriately in all cases and appropriateness has to do with when you prescribe, the dose you prescribe the duration, etcetera. We're discussing what are the ways that we can extend throughout the nation best practices of antibiotic stewardship in the healthcare setting. I think there are a number of possible ways to do it. A number of carrots and sticks

and things like that, and we're trying to weigh how best to do it. A second area is this question of stewardship and agriculture. I said we use 80% of the bulk of antibiotics on the farm. They're used primarily for growth promotion and disease prevention, and they're used in sub therapeutic doses, that is to say below the dose you would treat a sick animal. They are used because for reasons we don't fully understand you get more from meat from meat producing animals if you treat and you may reduce general infection by treating in this kind of prophylactic way. You also, the data are very clear, produce, as you must when you're treating continuously with low amounts of antibiotics, you produce resistance. And resistant bacteria are present in farm animals as a result. We do know that those resistant bacteria can get into humans on those farms. So there's no doubt there's exchange into the human population. In all honesty at this point given our reading of the literature I don't think we know quantitatively what fraction of our overall problem it contributes to. I wish I knew. I wish I could tell you 50% is due to overuse or misuse in the medical system, 50% is due to agriculture, but we don't know. We do know it's not zero. We do know it's not 100. And I really wish I could give you better numbers there, but I think regardless of it, it's important in any sensible plan to maintain our antibiotic stock to treat it as an area where we have to make sure that there's judicious use of antibiotics. And so we're very interested, and I think very supportive of the fact that the FDA issued a voluntary, two voluntary guidance, just in the past year, that voluntarily asked a couple of things. Manufacturers of antibiotics to consider withdrawing the claims on their label that the antibiotics can be used for growth promotion. Turns out the manufacturer withdraws the claim then you can't be using it on label in that way. Very interesting, when manufacturers who together produce something like 99% of the antibiotics in this country have indicated their willingness to engage in this discussion, I think signaled that at least many of them are very likely to do such things. In addition, the guidance calls on the agricultural industry to make sure that antibiotics are used under the supervision of veterinarians, rather than simply used by nonmedical personnel and just mixed with the food there, that they should be used for medical purposes under the supervision of the veterinarian. There has been a good deal of interest in the industry, voluntary guidance. They are voluntary. It's very interesting we're standing back and watching as we write our report some responses to these voluntary suggestions because if they prove workable, they still would allow use of antibiotics in agriculture, but if they prove workable they may be a very important contribution to the overall picture. The third topic that we're trying to think about -- before I turn to the third topic, there's one other thing that's come up in our conversations that's really worth thinking about is that we take very seriously the need to produce food on the farms. You can't say well growth promotion is not an important thing, and so it's important to ask whether science can find ways to replace antibiotics with other things. Pro biotic mixtures of bacteria that might produce the same beneficial effects. Vaccines you might give to the animals that fight off and protect against those diseases. It may be that the right combination is both to increasingly limit the use of these precious antibiotics we need for humans while serving the needs of agriculture by developing the science and technology needed to solve these problems through other means that don't create antibiotic resistance that could be relevant to humans. The third general area we're thinking about is the question of the development of improved antibiotics. Development of new drugs involves drug discovery, drug development, drug approval and finally drug marketing. You've got to ask if the pipeline has declined, in fact is nearly broken, why is it and where can you intervene? What can we do to increase drug discovery and are there roles that greater research can play at those stages in helping to identify new targets or

develop, the discovery of new drugs. Drug development. Turns out drug development means you've gotta go run clinical trials. Clinical trials for antibiotic resistant bacteria are tough. When do you actually know where you're going to get this patient. When you know you're going to get this patient you need to do something kind of right away. So having the right network set up are important. Do we need massive scale clinical trials. For many purposes we should. But for urgent agents like antibiotics against drug resistant bacteria, maybe we ought to have a different kind of a standard here to make sure that those can get on the market, in some limited way, for the patients who most need them. And it may be that narrower clinical trials, maybe you don't have to demonstrate what's called noninferiority, maybe you just have to have efficacy. We have no judgments on this but I can say that Congress passed the Gain Act an act that did streamline the clinical testing of antibiotics, and it looks like it had some pretty good effect, if not necessarily all there, but it may be that increasing the networks that are available for clinical trials can help for other things. Then there's the hard question of economic models. If in fact it is really uneconomic to produce these drugs or not sufficiently economic to attract in enough private innovators to do it, we may have to ask: Can you shift that model and how do you do it, and the problem it always costs money. Where should that money come from? These are many of the questions that we're trying to think about. And I don't think we're yet in a position to indicate even where we think we should come out. We may indicate merely what is the range of options and perhaps others should think hard about how best to deploy them. Finally, a couple of quicker topics here. New diagnostics. It would be great to be able to diagnose a respiratory illness is bacterial or viral in 30 minutes. That's a great challenge problem. We ought to be able to call on the innovation of the United States and others in the world to solve that challenge problem, because if you knew the answer during the same visit that you went to the doctor, it would really help with our stewardship. Similar questions, by the way, apply to a scary problem that affects the U.S. less but the world a great deal which is multi-drug resistant tuberculosis and ostensively drug resistant tuberculosis, forms of tuberculosis we essentially can't treat with any antibiotics. While they're not at high frequency in the United States thank goodness, that is a disease that's spread simply by walking around and coughing and they could become a problem in the United States and we can't ignore things like that. And the ability to, since TB is a very slow-growing bug, the ability to know in advance whether or not you're treating a patient with effective antibiotics or not is really important, because right now people start on treatment and six or eight weeks later, if they're not better, they say, oh maybe that was resistant and we should do something about it. Could we find out at least during the day that they show up at the doctor. These are the sort of questions we need to think about there. Improved surveillance. Well, I think it's become clear to us in the discussions we've had that the public health infrastructure in the states is not robust enough to really be able to provide the surveillance we need in many states. They're strapped very thin and there might not even be a full-time person in charge of surveillance of the antibiotic resistance problem. We've got to think about how can that be strengthened and improved. That's at the foundational level. And then at the high end scientific level the remarkable advances in genome sequencing makes it possible to sequence bacteria very cheaply. The inherent cost of sequencing a DNA from a bacteria, leave aside the sample preparation and sample acquisition, just getting that many bases of sequence is going to be about a dollar by middle of next year. There's no logistical issues around that. But at that price we ought to be thinking about seriously being able to acquire the information to know whether a patient in Washington who shows up with a resistant bacteria acquired it from New York who got it from somebody in Chicago who might

have gotten it from a farm or might have infected the farm. We ought to be able to trace provenance and see where things are coming from. Finally what more could the federal government do to help coordinate. This is something where we need industry playing a role, both pharmaceutical industry, the healthcare industry, but also the agricultural industry. But the federal government has unique coordinating role and a unique responsibility for something that is such an important public health matter. If we let antibiotic resistance get out of hand to the point where we have large numbers of bacteria infecting people that we can barely fight we risk slowly slipping back to that world that we have long forgotten of the early part of the 20th century, when people could die again of infectious diseases in large numbers. I think we've been aware of that possibility over the course of decades. But it's been on a slow boil and it's getting warmer and warmer and we're seeing now very disturbing trends, and I think for these reasons we've been tasked with thinking about what are the options and for these reasons we have been working very hard with a working group to see if we can come up with a coherent plan that will at least stabilize and I think reverse the situation. No conclusions to offer right now, no recommendations, but that's a quick summary of what we've been talking about. I wanted to turn to my co-chair on this work Chris Chyba for additional comments.

>> Chris Chyba: Thank you Eric. That was a very comprehensive, I think. Summary of the questions we're asking and I really have little to add. I might add in the realm of surveillance that a comprehensive understanding, were it possible to achieve one, would understand the reservoirs of antibiotic resistance that extend from agriculture, that include agriculture and food, human institutional settings and also the natural world and the flow of antibiotic resistance through those different reservoirs. Having said that as Eric emphasized, the first line of defense domestically with respect to surveillance is state health departments. Eric has already walked through the extent to which those health departments are strapped. The current CDC budget for direct assistance to states is $10 million a year, within the entire country. But this is an unglamorous unexciting topic on the low technology end of that problem. It has to do with personnel, with laboratory capacity and communications. And the question -- the question we're facing is what are the ways to try to address public health infrastructure at that level. In addition, of course, there's a fundamental international aspect to this problem because antibiotic resistance doesn't only arise in the United States, of course, so what can the United States do to better improve international disease surveillance, excuse me, antibiotic disease surveillance. As I said, Eric's summary was quite comprehensive with the questions we're asking and now our job is to come up with answers.

>>Eric Lander: The easy part was getting the questions perhaps. I don't know if there's some questions around PCAST. I see Bill Press put his flag up.

>> Bill Press: I have a simple question. Could you lay out for us what are the, what's the pattern of jurisdiction in the United States here in the sense of what part is FDA, what part is agriculture, if you take something off label, does that -- do both agencies have to agree with that? Just what's the mechanics of that?

>>Eric Lander: With regard to the labeling of a drug, it's an FDA consideration. With regard to --

>> Including for animals?

>> Eric Lander: Yes. Yes. The FDA labels the use of a drug, that approves use of labels of the drug, don't label but if they approve the manufacturer's proposed label for it. I think the USDA has a crucial role to play because the USDA is a trusted partner of our farmers across the country through the cooperative extension service. If this is going to engender changes in the use of antibiotics, I think the cooperative extension service has an important role to play in helping farmers that wish to adopt this voluntary guidance, and I think many will, to be able to make those changes.

>> Chad Mirkin.

>> Chad Mirkin: Wonderful presentation. I think you laid out the problem beautifully. Seems to me the solution is twofold. One is the development of new technology, which you outlined the challenges there and that's something that the country can overcome with the right amount of attention applied to it. But the other is the implementation of existing technology. So although we don't have 30 minutes diagnostics we have 90 minute diagnostics and many of these types of tools are not being used currently, extensively by the medical community and they should be. So is the group looking at ways of facilitating that type of implementation of new technology across the board?

>>Eric Lander: Yes. So the question of antibiotic stewardship and stewardship programs, a good stewardship program should include the use of the right diagnostic. Those should be technologically sophisticated diagnostics to get us the answer in the right time. The problem is we don't really have broadly adopted antibiotic stewardship programs that would include as an important component the use of the powerful diagnostics that we've got. Why -- well, probably because we don't have the incentives for it. Our medical system is strapped in all sorts of ways. And many of the best practices have spread throughout medicine through the creation of incentives, CMS, the Center for Medicare and Medicaid Services has a set of tools it uses to help spread best practices, by rewarding and penalizing the use of best practices and that's an area we're thinking about, is that a valid approach for doing this? Basically, I think many of the best hospitals are increasingly doing this. I'm detecting changes in the way they are starting to approach this. But to make it be a movement with real momentum behind it I think we have to couple it to the incentives that our medical system will pay for. We'll get the behaviors that we reward properly.

>> Jim gates.

>> Jim Gates: Thank you Eric. My fellow PCAST members it's a great summary of the problem. What are we thinking about in terms of the role of the public because in your recitation, you mentioned the fact that a parent might come to a doctor and say my child is suffering from this array of symptoms and can't I get something to fix this. And so there's a -- I would think -- a large role for involving the public in understanding the dangers of overprescription and are we thinking at all in that regime.

>>Eric Lander: Yes, and I relate to this as a parent of one of my three children a lot of ear infections, if you have a child with a lot of ear infections you know how painful that is for the child and also for the parent. All over-prescription of drugs is a dangerous thing to do. We need to be able to tell parents that

there are risks both ways, and that we can't be overprescribing medicines because it could harm children. But to support that, we really needed to Chad's point the best diagnostics, I think our public communication is going to be best the better our practices are and I think the antibiotic stewardship programs involve communication to patients as well because sometimes, honestly, doctors will prescribe antibiotics because the patient just wants something prescribed. I don't want to go home. It's not a nice thing to be told it's a viral infection and we don't actually have anything for you. It might be comforting to get a drug. But it might actually be the wrong thing. It might be harmful to get such a drug. There's both the doctors making sure they're adhering to best practices and the communication to patient that none of these things are risk-less. So it's a very important part of it. Maybe we should think more deeply about what aspects of public education might be involved there. Although it may be best that it does come through the doctor who will be the most trusted communicator.

>> No more flags up. Chris and I will be continuing this work with the working group we've assembled and hope to get back to you at a coming PCAST meeting not so long from now with a report and with real recommendations attached to it. But I think we've wanted to share that framework and we'll now turn on to our second topic.

Big Data and Privacy Report

>>Eric Lander: Our second topic is another study that's been going on. The President tasked John Podesta with bringing together a group to consider many policy issues with regard to big data and privacy. As part of that large effort, PCAST was given an assignment within it to understand from the point of view of technology what were the issues that will arise in the future, in fact, have arisen and will arise with regard to big data and privacy. What will potentially the technological levers available to the policymakers who might wish to fashion policy, and so because this has been a topic of great discussion in the country and of great importance, and frankly also great intellectual interest because it's a complicated topic, PCAST has, with relish, taken on this problem and I can report we do not actually have the recommendations right now because it is a hard topic, but as I tried to describe with antibiotic resistance, we've been trying to frame the question and would like today to share our initial thoughts about it. Susan Graham and Bill Press have been leading this work on behalf of PCAST. A number of us have been part of the working group I'll turn it over to Susan Graham to let us know where the thinking has gotten.

>> Susan Graham: Thank you, Eric. Eric gave you some of the context. Of course, what he didn't tell you is that we've been asked to respond in 90 days. Starting with January 17th. I was looking back to see when our e-mail correspondence started and it started on January 18th. According to John Podesta's estimate in a talk he gave earlier this week we're about two-thirds through this process. We have a working group within PCAST in addition to Bill and me. Jim Gates is participating, Craig Mundie, Mark Gorenberg, John Holdren, Eric Lander, Maxine Savitz, Eric Schmidt. I think that's everybody. And we've had conversations with many other PCAST members as well. One of the things we were asked in the President's -- or that was in the request that was in the January 17th speech was to consult widely with various constituencies and experts from outside PCAST stakeholders outside of PCAST, a we had a lot of conversations with a lot of folks and a lot of people have been very generous in talking with us about

this issue. So let me try to frame some of the topics we're looking at and some of the questions that we're exploring. The first issue that came up in our conversations is what's the problem, what do we mean by privacy? We all feel that we know it when we see it. And that's not good enough for the purposes of this study. So we had a conversation. We drafted some preliminary text and we thought we had nailed this issue and we find it keeps coming up. So even now two-thirds of the way through the study we're talking about what is privacy. And let me remind everybody that this is of course not a new question. We've had as a country and as a society continuing questions about what is privacy and our views have always been influenced by changes in technology. So if you think about the introduction of the telephone, when telephones were first available to people in their homes, there was a human telephone operator who operated a switchboard. And the operator, of course, could listen to any conversation. Furthermore, people didn't have private lines. That was a big breakthrough. And so in fact your neighbors could pick up their telephone and listen to your conversation. Now we're in a situation in which -- let me give you a second example. When e-mail was first introduced, it was introduced in research settings and primarily in the universities. And e-mail was public. Everybody could look at everybody else's e-mail. It was part of the openness of the research environment. And I can remember being surprised when somebody would say to me, oh, you and your group are having an interesting conversation about such and such. Now we're in that situation on steroids. In other words, we have an electronic world. We have a lot of information in digital form and we have a lot of ways in which privacy questions are raised and in which privacy might potentially be compromised. And so one of the challenges in doing this study is actually scoping. Figuring out what it is we should talk about and what topics are interesting but outside our scope. Big data is an important part of this question, because it's out there. Because we have a government program in big data and it's because it's a direction that technology is taking. The benefit of big data is we can have an enormous corpus of information we can do sophisticated analysis on that and we can learn a lot of new things. We can learn facts about the origins of the universe. That's a big data topic. It doesn't really deal with privacy. So in defining our scope, we wouldn't worry about the use of big data for the origins of the universe. On the other hand, a lot of the very important questions in society have to do with people. And so, for example, one of the promises of big data use and analysis is that we might discover rare diseases and we might learn enough about those rare diseases that we could find treatment for them. Part of the research in doing that is that we're going to look at a lot of information, health records, for example, genomic information. And basic science. And that gets back to information about people, which potentially compromises the privacy of individuals. So that's germane. There's a lot of hope that by studying online learning and looking at very large populations you can begin to see learning patterns that you can't see in the classroom about what kinds of questions students have difficulty in answering -- why they have difficulty, what happens if you rephrase the questions, what happens if you use different examples. That, again, is large scale aggregate information but now maybe it tells you why your child is having difficulty with certain reading problems or certain math problems. And so it maps down to the individual again and becomes a privacy problem. And worse yet, it's a privacy problem concerning children, which are a special case. So we've structured our project in the following way. We have this question about privacy. And then we've been trying to understand what is the technology that's being applied to big data. In other words, what are the analytics, what are their strengths, what are their weaknesses. The analytics aren't perfect. The answers we get from them aren't always correct answers. How does it

affect privacy if we're getting false information, which statistically can happen and can happen by bad experimental design. And what are the ways in which society is using and aspiring to use the outcomes of all of this analysis? What are the questions that people want answers to that could be informed by big data? So that's the first part of the problem. The second part of the problem is given that we're not going to turn this off, we're getting enormous benefits from this technology, what are the other technological approaches that can help to protect our privacy and where is the potential for harm, where is the great risk and how can we reduce the risk, decrease the harm without throwing out the baby with the bath water, without saying we'll just become so closed that we won't be able to get the beneficial information. To take an extreme if we decided you can't use any information about individuals we wouldn't be able to do any of this beneficial analysis. So we have to be able to use information about people without disclosing information that shouldn't be disclosed. But there's disclosure and there's disclosure. In other words, for example, many people feel that information about their health should be shared in the aggregate for health research. We'd like to find out how to attack disease. But we might not want our potential employers to know that we have a certain health profile. So the issue of context is very important. And that's one of the problems we're grappling with. How does one have the data there for its beneficial uses and define the context in which the data must be absolutely protected in which our goal is not to disclose it. So one of the things we're looking at very carefully is the nature of how data arises in the electronic world. Some of the new technologies, the disruptive technologies are things like cameras. If you walk around almost anywhere and you decide you're going to count the number of cameras that you're walking past, it becomes a staggering number. Those are only the cameras you can see. Cameras are getting smaller and smaller. And so there's a capture of information that's invisible. If you think about what a camera does it's got a purpose but it's scanning a scene and capturing a lot of information besides its intended purpose. If you have a camera in a traffic intersection and the idea is to look at the cars that are running the red lights or whatever, the camera will capture the occupants of the car. So part of the task of looking at technologies to protect privacy is to say, well, how are we going to grapple with that? How are we going to make sure that because you can see who the people are in the car, you don't use that information in inappropriate ways. We're looking at the relationship between privacy and security. Security certainly helps privacy, but it doesn't solve the problem. There are privacy issues that arise from the analysis of collections of data that are public data. And the power of the technology is that it can put together disparate kinds of information from disparate sources, and all of a sudden you know much more than you did than from one public database. So we're grappling with that. We're looking at the role of technologies like encryption, which is very powerful and surprisingly isn't used in a lot of contexts in which one would think, well, of course it should be used. Encryption has its strengths. It also has its weaknesses. There are times when in order to use the data you have to decrypt it. Once you decrypt it your level of protection goes down again. We're looking at technologies like that. We're looking at the research that's going in to trying to address those problems. And our responsibility in this report is in part to say that when are technologies that might strengthen encryption, for example, not having to decrypt, when do we think they might actually be mature enough to be in use in practice and how does that affect what we do about privacy in the meantime. We've also been asked to look at the international arena and we're doing some of that. I have to say that it's hard to really understand that in depth in a 90-day study. And then finally, although our task is on the technology side, there's an interaction between technology and policy. The group that

John Podesta is heading within the White House is focusing on policy. It's our view that technology informs policy. Doesn't make a lot of sense to have policies that we can't implement. That we can't actually carry out. On the other hand, if we consider the questions we need to ask they're informed by what the policy people see as the kinds of things they're trying to do. So it's a back and forth we're talking a lot with the policy group at the same time needing to focus so we get the report done in a 90-day period. I'd like to turn to my co-chair Bill Press to pick up on things I haven't yet told you.

>> Bill Press: Thanks Susan. That might be hard because I thought you gave a great summary of where we're going but maybe I can add a little texture for the PCAST members but also the public noting that since the session is broadcast on the Web it becomes a part of big data. (laughter) So really we are big data talking to itself here. And I want to tell it what we think. So what I've found very interesting is to try to trace back and

>>John Holdren: And everybody's picture is now part of big data. Bill Press: That's right. And what I found interesting is to try to trace back and understand historical roots of what we think of as our right to privacy because there are some things today that are really different. Scalability of effects and penetration into the life of everybody. But yet there are some things that are very similar, except that they need updating or may well need updating because of new technologies that have come along. It turns out when you look into this, you don't have to go much beyond Wikipedia to discover that an essential foundation at the beginning of the 20th century was a very famous paper by Warren and Brandeis. And Brandeis, of course, went on to become Supreme Court justice. And they attempted in this paper, which did not have any power of law, it was just a scholarly paper, but they attempted to lay out what they thought was a public consensus on what were rights of privacy and how they were suggesting that courts, mainly courts in those days, today we would say regulators of all kinds or even nongovernmental bodies, how they should view these rights. So it's interesting to look at the list of rights that Warren and Brandeis came up with which they summarized as four. The first was intrusion upon seclusion. That's the idea that basically you the individual have some perimeter of privacy that it's a harm against you if it's invaded by other people. Now, in 1890, for, say, people of a socioeconomic middle class and so forth, that was probably their home. It was a pretty clear boundary. It said what you do at home should remain private. But how should we interpret that today when many people and generationally many more younger people than older people feel that their home is out there in cyber space. Their home is the network of people that they interact with; it's their Facebook page and so on and so forth. So what is intrusion upon seclusion supposed to mean today? I don't think we're going to have a crisp answer to that, but I think we're going to try in our report to understand that. The second one of Brandeis's ones is very interesting is public disclosure of private facts. And this is supposed to be a right that goes beyond something like libelous or slander which apply which we're more familiar with, I think, and which apply in varying degrees to private individuals versus public figures and so forth. But this was that same idea that if you as an individual not a public figure had parts of your life you wanted to keep private, you should have a right to do so and you should have some means of redress if someone violates that. Now, today that's not only complicated because our private lives are out there in cyber space, it's complicated that it's no longer so clear what is the public setting and what is a private setting. It's clear at the extremes, but when someone posts cell phone pictures they took of you at a

party, was that a public session or a private session? Do you have any rights there? Or were you just out there and they had a right to be there and took the picture and post it? When we all start having wearable mobile devices that are constantly sensing our environment and recording it. Google Glass is the first one to have caused quite a stir. Does this mean that every conversation that I ever have with someone wearing Google Glass is a public conversation, even if it occurs, if it's a one-on-one in a private room, I'm just supposed to recognize that they're wearing that? So I won't go through the other two of the homes ones, but on those to point out although they're principles we all agree with we need to understand how to apply them today. Then, and I think we've just started thinking about this is the question, we shouldn't be bound by a list that was made in 1890. What would we write down today, with a clean slate? What are the generally perceived public values about privacy as they apply today about things that can cause harm and what do we think should be a set of options for the policymakers of redresses for those harms. In some cases it may be as simple as disclosure, disclosure of the fact so you know what you're getting into. I don't remember, does Google Glass have a little light on it that shows when it's recording. If it does or doesn't, should it or shouldn't it. But Susan raised a very important point, and that actually does get back to this private fact thing. In Holmes's day a fact was private because you knew it and you wanted to keep it private. Now we have private facts about an individual that he or she, himself or herself may not even know. We have private facts that are derived from public records. So you can see that you could imagine going anywhere on a wide spectrum on that. You could say no public records are public records and anything derived from them is public. Get used to it everybody. Or you could say no, the classic definition of a private fact needs extension. The individual should have rights in privacy in those facts, even if they are derived from the public. Then the technology question is how do you do that? Technology and policy question is how do you do that. Because in many cases this private information is only latent in the data that's collected and latent in a way that you couldn't even know at the time of collection if you wanted to, specifically big data involves not just big data sets but it involves powerful algorithms and algorithms that haven't even been invented yet. There may be data sets that seem to us today not to have privacy implications but tomorrow, because someone clever has invented a new algorithm, suddenly have privacy implications. We need a framework that's robust enough to deal with that kind of situation and that framework is not going to be something as simple as a checklist you may or may not do this because that checklist will never be complete. Maybe that's all I want to say in terms of what we're wrestling with. These are the same issues that others in the public space are wrestling with in the workshop sponsored by the Podesta group, who have been very helpful to us and in writings every day. You can find writings every day on the Web on this subject that we're trying to assimilate and we'll eventually try to put our oar into the water on this.

>> Eric Lander: Thank you both Susan Bill The floor is open for questions. I see Chris Cassel and Barbara Schaal.

>>Chris Cassel: Thank you and thanks to both Susan and Bill for the very nice and thorough explanations I felt like I was in a class with both of you guys. I know both of you are professors so it was very helpful. I just want to add that there's overlap just even as we speak literally with other things that are happening. There's a story in today's Modern Healthcare that the FDA is now calling for classifying

health IT products by their level of patient safety risk. So this has been an interesting debate in the medical and healthcare world about, because the language, the authorizing language for FDA includes medical devices, and it has long been thought at least legally that that includes health information devices. And of course that can take all kinds of forms with mobile devices and biometric things as well as what we think of as electronic medical records. So there's been a big debate about how exactly would they exercise that if at all. And I think they've taken a really interesting tack here which is to say that it's not the device or the platform, it's how it's used. And could patients be harmed by this use of this medical data. It's just a brief article I'm sure we'll all be learning more about it but I wanted to in some way add it to the framework of what you're thinking about.

>> Barbara

>> Barbara Schaal: I also want to add my thanks to a really well presented discussion of a pretty challenging topic. I think that clearly there's a framework you've produced a framework for addressing some of these issues. I just have a few questions about extending that framework a bit more to the international which can be daunting because there's different concepts of privacy in different countries and there's also different concepts about how to deal with privacy, whether you control the input of data, whether you control the data or you control the use of the data. And my question is: As we move forward addressing those issues in an international arena, do you see any particular parameters or challenges that are vexing related specifically to the international? Is there a framework for dealing with this, and is there an opportunity to produce some kind of a harmonization or how do we account for the variance in those concepts?

>>Susan Graham: So let me come back to the fact that our charge is to talk about or to investigate the technology in this space. And so we've looked primarily at what are the technology developments in other countries that we don't know about or that we're not doing. When we talk about, when we think about technological solutions or partial solutions, technology to help to protect privacy, a lot of them are not the -- the values are parameters in many cases. In other words, what you choose to forbid can be adjusted, if you get the technological solutions right. So while it's very important to recognize these different values in different countries, it has only an indirect effect on our part of the problem. Now the policy group has to worry about it in a very different way than we do.

>>Bill Press: I think specifically on the international scene there's been a lot of public discussion on, is the U.S. going in a different direction from western Europe, from the EU. So as Susan says, we're largely the technologists in this, but one of the knottiest issues is what if there's a value that we, that the U.S. and the EU may share completely but we have differing views on whether that value can be enforced technologically. What should the balance be between trying to enforce it technologically, versus trying to enforce it by laws and penalties, because then it's not a question of not sharing the same values, it's the question that we might end up in quite different places practically, and of course that's bad for the progress of technology. That's bad for the realm of commerce, when the playing field is uneven, when there are different rules in different places. And so I think as several people have mentioned, it's not just the values and it's not just the technology, but it's really how they interact, which, as you perceived it, is very complicated on an international, in an international setting. Eric Lander: This is one of the most

complex topics we've taken on, as is clear from this discussion because it involves technology that spans essentially all possible human uses, not to mention lots of nonhuman uses, all the devices that we have that are going to be doing things in our homes and our economy without us. And at the same time fundamental question of values. PCAST is not a group that can resolve all of these questions. These are fundamentally policy choices that have to be made, value choices, but I think we have a unique responsibility to try to supply the answers to the questions where is technology going, how can technology help. To do that we do have to connect with these great larger questions here. I think it's an ambitious 90-day assignment. I'm grateful to Susan and Bill for leading us into this assignment. My guess is that it may not be the last word ever on the subject. But I'm hopeful that it will be a helpful starting framing of these questions and perhaps helpful technological input both to policymakers and more broadly to the public of the United States, and I suspect around the world, we're all trying to think about how do we ensure that we get the advantages of technology while still preserving the important aspects of privacy that we have always held dear that are important to the development and fulfillment of us as individuals. With that I'd like to suggest we move to our break. We are virtually on time. And we shall return, I believe, in about 10 minutes or so, according to the schedule. Where we have some guest speakers. Thank you.

Analytical Techniques to Improve Public Policy Decision-Making

>> John Holdren: Okay. We are now moving to a session on analytical techniques to improve public policy decision-making. I'm going to turn the chair over to one of our two vice chairs, Bill Press, who as I explained to a couple of our speakers he knows more about this subject area than I do and I'm very happy to have Bill to moderate this session. Bill

>> Bill Press: I'm no less interested in learning about this than you are, I think that's all I would say. So let me welcome the four speakers. And I think maybe the way to do this is I'll introduce all of you now a bit and then we'll just be able to go from one presentation to the other from my left to your right. Your right to my left. We'll start with Chris. I wanted to mention that the origin of this session is in a suggestion by Dick Garwin here today down there. Dick was a member of I guess the original PSAC, which was the predecessor to PCAST in the era of Presidents Kennedy and Johnson, and has been a trusted advisor subsequently I think to every PCAST as well as to every science advisor there's ever been. So it's great that we're able to implement Dick's great suggestion for this session. The idea of this session is to expose how computational and analytic techniques, including very much big data techniques, so there's some relation to the previous session, how those techniques have been moving from the realms where they were often invented, for example, in the physical sciences in simulation of physical systems, whether they be spacecraft or rockets or other such systems, into domains that affect policymakers more directly, can inform policymakers more directly, but also in some ways just have more relevance to the lives of individuals, of everybody. In some sense this is a look at new directions in the social sciences, but I think we didn't want to cast it in that way because we weren't looking for a disciplinary silo approach. We were really looking to bring in, as we have four practitioners who do very different things. So I hope they're not too uncomfortable while being under this same umbrella. And

they'll have a chance to tell us what they do. But things we think are, through policy and policymaker decisions, are going to affect the lives of people and that therefore PCAST should know a little bit more about this and so should the public. So PCAST members you have detailed biographies of all four speakers in your packets. And I won't read them in any detail. But for people, members of the public and people watching on the Web, let me just say very briefly a little bit about each person. Chris Barrett is executive director of the Virginia Bioinformatics Institute which is a part of Virginia Tech and also a professor at Virginia Tech. Before he joined VBI, he was for 17 years at Los Alamos National Laboratory where I knew Chris and had the pleasure of interacting with him where he was a leader in the group of basic and applied simulation sciences. Dr. Barrett has received a number of awards in his career that I won't read you the whole list, but he's been Jubilee professor in computer science and engineering at Chalmers University in Gothenberg Sweden. He won distinguished service awards at Los Alamos National Laboratory and I could say the list goes on. Claudia Perlich is chief scientist at a company in the private sector, Dstillery it's spelled without the first I, so I hope I've pronounced it correctly. And that was a company previously known as M 6 D which is easier to pronounce but no less transparent. I'm sure that's what I mean. But let me tell you what she actually does. She designs, develops analyzes and optimizes machine learning that drives digital advertising. Here's a very practical application in the modern world that has touched us all. Claudia has published more than 50 scientific articles and she holds multiple patents in machine learning. Her Ph.D. is from NYU and she spent many years with the predictive modeling group at IBM's Thomas J. Watson Research Center where she also concentrated on data analytics and machine learning for complex real world domains and applications. Charlie Catlett is senior computer scientist at Argon National Laboratory, and he's also a senior fellow at the Computational Institute of the University of Chicago and at Argon National Laboratory. He's the founding director of the Computational Institute's Urban Center for Computation and Data. Urban CCD. Which I think he'll tell us about so I won't give you the summary here of all of that. In some of his prior positions, he was chief information officer at Argon National Lab. He was director from 2004 to 2007 of NSF's Taragrid initiative. And going back earlier in his career he was chief technology officer for the super computing application at University of Illinois at Urbana Champagne. His degree is in computer engineering from UIIC. And finally Duncan Watts is principal researcher at Microsoft research and a founding member of the MSRNYC, if you can decode that, Microsoft research New York City laboratory. At the same time he holds or has held several other distinguished positions. He is the Andrew G. White Professor At Large at Cornell University's, Cornell's highest honor for external scholars, and he has been an external professor at the Santa Fe Institute as well as a visiting fellow at Columbia University and at Nufeld College Oxford. Duncan is the author of six popular books. And in the spirit of -- I know one should always plug the most recent book of anyone one introduces, that's my principle. And his most recent book is Everything is Obvious Once You Know the Answer. We can look at that. Duncan's PhD is in theoretical applied mechanics from Cornell University. So with that, let me go to Chris Barrett and tell us what we should know. Chris Barrett: Thank you. So to be brief is the art of knowing what to leave out. Not something I'm particularly good at.

>> Speak up a little bit.

>> Chris Barrett: Leaving things out is not my strong point. So I'm going to struggle with this. But so this is measuring modeling and managing interacting systems. I guess it begs the question -- how to make it change begs that question. There we go. So the relevance to a bioinformatics, bioinformatics originated obviously as a term worrying about the genome. To understand how things are expressed is moving off the genome in a big way. You can't understand pathogens if you don't understand hosts and you can't understand hosts if you don't understand host environments and to understand host environments you have to understand interactions among them. And with people, their environments are technological and social. And you wind up being really blurring the distinction between social and environmental. Socio technical and biological domains in the direction that modern bioinformatics is moving in. I'll talk mostly today about population scale things. But they pertain and connect immediately at the molecular level as well. The big thing here, relevant to this topic, is of incredibly diverse unstructured information. That was information that was never intended to have anything whatsoever to do with understanding anything about what I just said, how pathogens interact with the hosts or how hosts interact with one another or anything else. And it qualifies as big data in the sense that if you look at curves, I didn't bring one with me, but if you look at these exponentially growing curves that are growing in many documents, about big data, the vast majority of it is what's called unstructured, which means it wasn't collected for the purpose that you're using it for, basically. It can be more than that. But basically that's what it means for purposes here at least. It also involves insilico generation of really very large scale interaction dynamics. So by that, if we're looking at something like social epidemiology, which is a popular topic now, hardly the only important topic but I think it's one that probably everybody has some familiarity with, rather than using aggregated models, social epidemiology, you actually can get down to where you have Avatars for every single person, every single activity location. There's detailed interaction and reaction to diffusion, the spread of the disease over the people and then the internal dynamics of it. So if you have 300 million people in America and 160 million activity locations that can generate a pretty big system. And if you have 7 billion people on earth it can generate a really big one. And sometimes we actually know that that level of detail of that individual or individual characteristics, it matters to policy, because you may be talking about influencing individuals either behaviors or biology or sometimes it matters to how you can collect information and distribute it in these systems and the relevance to policy is different than the relevance to science in many respects in the sense as scientists we may be interested in certain kinds of things we keep trying to tell policymakers they need but it's not at all clear that's exactly what they need. For example, prediction may not be nearly as important as explanation or keeping up, I guess is a simple way of saying something about that. So with respect to how you can possibly scale, So I'm going to focus on the technology aspects and the methodology aspects of how you would approach such a problem. What are the things that are interacting, like who are you? What are you? Can I make an agent-based model of you that's a software object that's one of them is for each of us, and the answer to that really is on the kind of scaling we're talking about from a purely technical point of view that really isn't very feasible and leads to all kinds of other kinds of problems. Here's an example of some of the things I'm calling the un-encapsulated self. Vitamin B 12 is produced by cyanobacteria in your gut. Those bacteria don't have your DNA so technically speaking they're not you. On the other hand they behave like organs because you don't make any B 12. You need it otherwise you won't live. Moreover that particular bacterium is tuned to your matrilineal heritage. So the definition of what you are it's just not clear when I'm like on you, in you, or not. When I'm outside of you. This also

applies to thinking, to understand what a Bee does, doesn't necessarily make you understand obviously what a beehive does. Neurons don't know what they're thinking and sometimes we collect together in situations where we don't know what we're thinking either in exactly the same way that neurons and molecules don't understand exactly what they're doing. It's a concept that I'm going to put a name on here I'm going to call it a role, a simple word, we play roles and those roles interact in ways that produce things. They can be prices. They can be all kinds of things. But they exist outside any one of us and yet affect us and are associated with us. An example, to keep some sort of the spirit of the population biology is that you may have read about that there's -- there are a couple of different ways this happens. But there's a genome associated with a hopper. I'll call it a hopper. Depending on its environment, which it has some control over, because it interacts with it, it winds up in two very, very different morphological phenotypic forms, in phenotypes that are different are both behavioral and morphological. One is a locus form and one is a grass hopper form. They have an identical genome. To understand the expression of these things, you have to understand not only whether it's dry or something, you have to understand how many grass hoppers there are, when they're in these different forms their mating behavior is different, their swarming behavior is different. Their metabolomics are different. They are different creatures and they have always thought to be different species and from the genetic point of view they are the same thing, they can go back and forth depending what they do to their environment. This is simply to say in this case, if you want to view DNA as software, or as something people talk about, the program, it's middleware. To understand what this animal is requires something that lives outside any particular manifestation of the animal. So this is the inside/outside problem, how do we know who you are, and how can we exploit that from the point of view of technology to compute interactions on all these different levels that define us. How can we possibly do that if we know we need to. So un-encapsulated agency, encapsulated agency refers to basically, it alludes to a software convention taking encapsulated software objects and assigning them agentic properties and computing interaction among them and this argues against doing such things as a general technological issue for purposes largely for scalability and analytical access and flexibility. Interactions in the media if you look at the grass hopper example they concretely shape it. If I ask you where's your money it doesn't have a simple answer. In fact it doesn't really have an answer. In physical space your money isn't anywhere. And neither is your debt or the actions you take relative to it or your responsibilities, and it gets to be a complicated thing when you say where is your identity. We'll have to ask the question how do we get populations of all these you things that are not even local and interact them in all the ways they can interact for purposes of informing policy, because you're going to do things like have economic effects guide something else which cascades to, I don't know, juvenile obesity. There's a complicated chain of things going on there. And so how else can we possibly take care of these things. So what we've been working on many years and it started a long time ago, although it didn't have this form over the last few years we call them synthetic information platforms, where basically in the context of a query which can be a pretty complicated thing and can also be millions of queries I should point out not by one individual but by many, it organization the passing of this unstructured information into in context synthetic basically databases and a large data library where the individuals are IDs that connect across all these databases. So that you're un-encapsulated, and that you interact in the context of these abstractions in ways that are very informative, very microscopic, very granular, and they scale to hundreds of millions, billions of individuals and homogenous ones. I'm going to try to speed

up here. Okay. I'm not going to talk about this, except to say that since prediction in these kind of environments where you have mediating behaviors, let me just cut straight to the middle of this thing. There's all kinds of reasons why this is a problem. But I'm going to talk about this thing that says wrongness in temporary confirmation. This doesn't just happen with people, you know, gaming each other or lying to each other or something. It's deception. It's a basis of deception. And viruses do it, too, post cells. It starts happening very early in self-replicating biology but it's certainly true that you cannot have a unique mapping between the behaviors and the mediating states. If what you're trying to do is moderating the mediating states and this title of Duncan's book is really quite to the point. If you already knew what happened if you already knew why they do it it's obvious it could come out that way but it's not necessarily obvious that's what would happen otherwise deception would never work and that's an example of practical real world example. There's obviously places for prediction but there's also places where certain kinds of explanation and other things. So in a synthetic information environment, this has to do with this contagious disease stuff. These are all apps that have very different things. ISIS is an application for public health, people to do very detailed analysis, Bluecaster is a forecasting model My Foresight is what I'm going to tell you about in a minute. Granite is a very large graph analytics tool, there's visualizers and mobility models and training tools and stuff. These are all designed as apps that ride on top of the synthetic information platforms that have hundreds of millions of individuals and it allows you to do things in tractable ways that if you tried to build models for each of these things separately, then you would sort of run out of gas trying to get it all in hand, I guess, is the point. My Foresight is what I want to bring up because it handles certain things to do with guiding predictions and other things. It has to do with actually capturing and blurring the distinction between the users of information, decision makers and data sources, which isn't fairly new for lots of marketing things. It's new for science. The issue is that when, if you have counter factually many different potential futures, there's a lot of reasons you could have many potentials. You could have options or you just don't know. There's lots of reasons. You can ask people what they think, to rank them, for example, and vote on how certain they think they are of this ranking and include that in information about the forecast. It changes how the forecasting works, changes how the modeling direction goes, and it includes what is viewed normally as something that's outside which you would be able to collect, which is scientifically human intuition and expertise and build it directly into these systems. So what are some of the things that tools like this can apply to policy? Well, the epigenetics of juvenile exposure to violence is a topic that would be amenable to such an approach. Influence networks of all different kinds -- incarceration rates, personal success and college, what college you go to, smoking, stopping smoking. All these kinds of things are examples of policy relevant things. I'm just going to -- so what's the point of these things? Information and computing technology is layered into society. Guiding state assessment decisions actions at all levels. Not just at some policymaker's level. Johnny's parents have decided Johnny gets a vaccine. Or goes to school or something. They're decision makers, they're part of this thing. There are millions of them. You can include them. It scales to the natural size and richness of society as opposed to struggling with trying to get all these things that are abstract over them and leave them out in a more contained kind of an approach. It evolved as ICT and products and things evolve. It doesn't, it just naturally evolves. It changes how to think about computationally enabled social decision-making. So increasingly this ICT world, which is much more than the Internet is a part -- it's a part of both the laboratory and our computer models, rather than trying to cut it off from the world, we're

talking about insinuating middleware like the Locust and grasshopper into this world and realizing those things we want to -- for whatever our purposes are. People and technology are embedded actively rather than simply observed and managed in these methods. These methods are really the only way to deal with this level of complexity of these systems. One of the questions was how do these methods have something to do with these problems. There's no other way to address these problems. They're scalable. Unstructured data itself, upends a whole lot of scientific prejudice really about what data is non-demonstrative methods, hypothesis testing, counterfactual analysis. Things that lawyers are much more familiar with than scientists maybe have to become part of this practice. And prediction per se is diminished in both purpose and plausibility for many questions. And it really almost isn't the point oftentimes. And so it's the end of the contained model and the end of the great man theory of decision-making. It's the enacted ICT enabled sociotechnical world.

>> Thanks Chris. I think we'll hold all the questions until the end. So let's just move directly to Claudia.

>> Claudia Perlich: True to myself, I'm a problem solver, usually solve problems with data. This time the problem was we were given some kind of guidelines as what might be of interest to you what I do in my world as to policy making. I'm no expert on policy making have an idea about it I grabbed the questions and outlines and went with it. And I guess on the first topic here all the different methods that people talk about in the context of big data, analytics and so on. I think it's important to not get lost in the soup of terminology, because many different disciplines refer to very similar techniques with very different terms. And I think it's important to focus on what these different methods can be used for and what they do, not as much as what exactly they were based on what discipline they came from. So when I looked at the long list of things that I was given, I think there are really three separate groups identified here. One is tools to manipulate data. My somewhat personal reflection on the rise of big data is that it really has come from a technology shift in tools to manipulate data. That's where it originated and a lot of the analytics has been around for much longer, but the jump I have seen over the last two, three years is really on what I would call the plumbing, marvelous plumbing I love it. We have on top of that tools that describe the past and the present. So you can ask questions, for instance, obesity rate by neighborhood in areas. So you can now get much closer more detailed view on what the world really looks like or looked like in the past. And it relies on these manipulation tools and collection tools we have to bring it all into the same place. Now, the area that's my bread and butter really over on the tools to predict and my personal frame of this predictive modeling but I think simulation falls into that as well. So I would like to spend some more time on that. And the first reflection here is that while people understand prediction as about the future. Prediction most tools that can predict don't have to predict the future. They can predict the present, or they can even tell you what would have happened in completely different circumstances. And I think Chris already hinted at this at counterfactual. So examples here that I pulled up some work that I'm familiar with is, for instance, great work collaboration between UNICEF and IBM that helps categorizing and assigning importance to text messages sent by youth to in Uganda in medical emergencies versus case of violence and then they get routed to the right place. It's not about the future. It's just helping making the system more effective. And it's one of these application domains of predictive modeling. I've worked on medical diagnosis, and you all probably have forgotten about the fact there's suspenses on your e-mail, you only remember when it does something

really bad. I think the other point I want to get into is predictive past because I think that has very interesting implications in policy and understanding and evaluating. So going beyond the prediction as Chris said towards more interesting uses of being able to drive knowledge and change action, really what you ultimately want to be about. So my somewhat naive view on policy is really you want to change things. You want to interact with the world and hopefully it will be better outcome. The question really is what would happen if I did this? I would love to know but I don't. So all of the conversations we are having about really what are the implications of our actions and to what extent predictive modeling can be part of that. So simulation is a technique to predict. It starts basically with a set of equations and understandings that is already in human mind that you put down on paper or computer systems and then roll it forward and see what happens and you can tinker with the parameters and see what else might have happened under slightly different circumstances. The results are as good as your assumptions, if you know a lot about the world then they're great. And data is really playing a slightly less important role here and can be used for validation of these assumptions you had. Now, the world that I live in is completely on the other spectrum. It's predictive modeling, and my very dry view on it, function approximation, what does it mean? You give me a bunch of data and I'm trying to find what is that equation that underlies the data. So I start with the data and now the problem is my results are only as good as the data you gave me. I'm not talking about noise here. I'm talking about how comprehensive it is. Does it really provide all different aspects that should be in there? Is it biased, are there things that I don't see for whatever reason in the data I'm having. On the upside, I don't really need to have a perfect understanding of what drives things when it comes to human behavior, that's actually pretty much true. I don't know what makes me do stuff, let alone other people. So I would like to focus on a little bit on what I see in terms of the potential limits of predictive modeling specifically. I have seen an incredible shift in terms of efficiency and it's primarily with big data. We have so much more detailed level of information now I'll go into numbers a little bit. I worked for a start-up at four years old. We collect 10 billion events a day where people do things. So I have 10 billion actions that people take. From hundred million actions all online. And I can base on that information make pretty good predictions of who is interested in what product right now. I can predict a lot more if I wanted to. That's not what I'm currently getting paid for. But human behavior, once you see this very granular information is really exciting playground for this technology. What on the other hand where it doesn't work very well, the idea of predictive modeling. If we talk about rare events like Katrina, as I said it's only as good as the data. I need to see a lot of instances of that thing that you want to predict for. I only get a hurricane once in a while; I will not be able to make good predictions about its impact, who would be affected and so on. The other is if I have very few instances, slightly different, I was talking in the good case about hundreds of millions of people that I make predictions for and that I observed. If we're talking about predicting something on a country level, I don't have that many countries. There are only maybe 100 best case scenarios, 2,000. It's not enough for the technique to really play out at its best. And finally, and I think it's important for the policy context, I can't predict when the circumstances are completely different. If I only have seen data in a world where there was no healthcare, I can't tell you what people will do if there is healthcare. I simply have no data for that scenario and that technique wouldn't be able to really answer questions. And this is the part that technically known as extrapolation. If I have data in a certain area, this is extrapolation in time it's the stock market, I know what happened. But to tell what will happen in the stock market in a year from now it's completely impossible. Predictive

modeling will not be able to do that. Because it's out of time. And that also applies to circumstances. If I have information about how people behave in the U.S., doesn't mean I can predict how people behave in Germany. It's a different culture. Behavior might be very different for certain aspects like privacy considerations. So one of the questions posed here is will you have better predictive tools? My reflection is the analytic tools, the methodologies, the algorithms, while changing, actually have been around for a really long time. We know how to do this. We have been doing this 20 years ago. That's when I got started on predictive modeling. What has changed, my earlier point, is really the data that is fueling it. So the tooling although having evolved has become more easily accessible to people isn't that much of a driver. It's really the data that goes into it. Now, my somewhat word of caution around predictive modeling goes back to is the problem you're asking even answerable by predictive models. In particular, do you have the data that could build the model to answer your question? And most often the answer is no. Quality control is extremely difficult. At least in simulation, if somebody gives me a set of equations, there should be a couple of people in this world who understand what they mean, and can explain it to me. Since the model is only as good as the data, you need to understand the data. If I drop a terabyte of data on you there's no understanding, there's no way to really grasp and fully explain what is going on. So ultimately it becomes a scale of the modeler, it's very hard to control and I would even -- and I think that's what you run into also in the privacy conversation, is that it really is hidden in this vast stack of complexity of data that's very hard to get to. Now, one upside I wanted to mention here that I see a lot of potential, are there are better techniques. Predictive modeling, learning from data is currently used in a very exciting application called causal analysis. Now, we all really want to know why things happen. That is the objective. So typically it's done by a beta thing but you can't really do that with human subjects. You can't just apply policy to half the U.S. population. It's not going to work. So what can you do? And that's my closing point here is that you can use predictive modeling to estimate causal effects on just the data you observe anyway. And this has been done very successfully in medicine, and we do it also in advertising, and I'd be very happy to share any information you're interested in. Thank you.

>> Bill Press: Thank you.

>> Charlie Catlett: I'm going to just build right on top of these last two presentations and give you a tour of some of the things that we're doing and what are being done in other cities with these various technologies that we've been talking about. In the context here is we started an activity a couple of years ago where it was my vision I would say to try to bring together different groups of people with different expertise to solve problems that they wouldn't be able to solve on their own and to build capabilities that one set of people might be able to provide to another. Especially going across computational science and social sciences. So we've been focused on I'd say three different layers of policy and practice in cities that are defined in terms of time scales. When we look at time scales of years to decades some of the things that Claudia just talked about, we want to include in computation models to get a sense for what are possible outcomes of urban design or policy. In terms of closer policy, what might happen over the next few months or hours or maybe the next few weeks. There's a lot of new data that's available and the two communities I'm talking about, particularly here, there are a lot of people with passion for analyzing data that don't have the training or background to know what are the

right questions to ask of that data when it comes to policy or in our case the cities. Then there's a vast community of people who ask the right questions but they're not aware of the data being available. And if they were they wouldn't have the tools or the methods to be able to analyze that data. So we're trying to bridge between those two communities. And then in the very short scale, short time scale of even seconds to hours, we are exploring ways to collect data about the environment and even to build capabilities that would change the way people interact with the environment. I won't spend a lot of time on the bottom part here in the optimization interaction, but I'll try to spend most of my time in the top two. Then I'll close with a little bit about what we're learning with interacting with students who want to work in this area. So the first problem we started to look at is the 600-acre site south of Chicago called the Chicago Lakeside Development, 600-acre peninsula. It is steel slag, artificial land built over the decades by U.S. steel. It's brownfield. It's surrounded or adjacent to the west to some of the areas of Chicago hardest hit over the last hundred years or so as economies have changed and as demographics have changed in the city. Skidmore Owings & Merrill along the with the developer McCaffrey have an interest, have elaborate and nice plans for the site that talks about how you won't need a car to live there. All the storm water will not go down to the Mississippi by way of the river that we reversed, but it will go back into the lake and they have the vision to use renewable energy to power the site. When we sat down with them and asked how do you know these things will be true, their answer was: Well, we're hoping it's true and it's our intuition that it's true but we don't really have a basis for saying that this is actually going to be the case. Now, it was more clearly illustrated to me when I first started talking to Skidmore Owings & Merrill about a real estate development in central Beijing, Beijing central business district. They said we're going to plant all these trees beautiful conceptual picture. This will help clear the air. I asked how do you know it will do that? They said we don't. We said we actually have some computational models that will help you answer that question. So this project here is just about bringing the tools and practice of urban design at very large scales. These are unprecedented scales, 500 acres, 700 acres, Hunters Point in San Francisco, not very many in the U.S. Almost every city in China is building out at 500 acres or more scale. To think the designers and policymakers don't have what-if scenarios over the period of a few decades is a little unnerving. So we took a set of software we had was to model energy in every single building. We put it into an architecture design software from ESRI, and then we allowed the designer or the policymaker to turn two knobs. One knob is what's the zoning of each of these districts within the 600 acres. The other is when is that going to be built over the next 10, 20, 30 years and based on that we would run energy models for every building and come out with the curves off to the right, with the real purpose of giving them an ability to design the energy demand or the energy supply based on the demand that they were building. And then recently we've begun to say, well, to try to answer two other questions. And the first one energy is also dependent on what the climate does. We have decades of historical data about weather and we want to, by mixing up the wet and the dry years and cold and hot winters that's happened in the past and changing their frequency, we can get a sense for what the energy demand might be if the climate goes this way or that way. This is not to say that here's the predictive answer but to be able to compare multiple scenarios and try to zero in on ones that make the most sense. Similarly, for transportation, this site is on the other side of a choke point, where five lanes in each direction go down to two. And then back up to five in each direction. And there's a question from a policy point of view, when to expand that road and build more capacity. And what might be the impact of that decision if there were a rapid bus transit that was put in

place or a rail, light rail line. These things come together, and with transportation it's the same two knobs, zoning and phasing, these things come together if we want to ask a more precise question about energy, which is not can we meet the monthly capacity needed but can we meet the capacity over a 24-hour period of time where we have renewables and storage that come into play and if it's not the same capacity over that 24-hour, the bottom right graph here, just made up numbers in there to give you a sense for the change over 24 hours. And there human occupancy and buildings which is the inverse of taking a trip that you would learn from actually Chris's software that we use TrendSim, might give us an ability to bring the temporal resolution of these models down to the hourly and then to be able to ultimately feedback to the designers the design parameters that say if these are your energy supplies over the next 20 years this is how you should phase and zone the build out of this property. I don't know exactly where to point this. Now, to do the transportation designs, you have to know something about who people are that live there, where they might go. There's some very good survey data, this is some screen shots from a beautiful visualization that Marta Gonzales at MIT and others did from two days of detailed surveys of 10,000 households in the Chicago area. If you have this in your notes, if you go to the website you can see how the dots change color and move around. That data tells us for something like lakeside development what are the catch basins for different shopping and entertainment districts and do they overlap with the lakeside development. Can we imagine people going there for entertainment. And then we take data from the Census Bureau that gives us home and work destinations and income levels and types of jobs and all those things need to go into these models in order for them to do more than just give us some entertaining visualization. I'll just say one word about collecting more data. There are a lot of scientific problems that we think we have a good handle on to one degree or another but we don't really have data to validate the models. One basic question is there any impact on the regional climate that you can have by different policies with respect to green or white roofs or spatial distribution of buildings or any sort of architecture urban planning, we don't really know the answer to that question. And part of the reason we don't know is we don't have good data about the boundary layer between cities and the atmosphere. So we've got a project that we're prototyping that started out as a way to blanket the city with sensor nodes in order to get some of this data. And it turned out every scientific group that we talked to said, gee, if you had this sensor or this capability of bluetooth device we can do this other kind of research as well. So we think there's a lot that we can think about cities with data that doesn't exist right now if we can begin to collect it.

>> Charlie, you have about five more minutes.

>> Charlie Catlett: I have my timer to go off a little bit early, but I think I can do this in five. So we've seen in many cities, in the upper left of the screen here, city of Chicago data portal, city of New York, Boston, Seattle, Portland, San Francisco, all of these cities and counties and states and some governments, country governments, are using one of two capabilities or one of two technologies for their data portal. And there's this community I mentioned before that can take the data out of that and do all sorts of applications like you see in the bottom left there, all of which give you mapping or trending or other things, and some of the tools we've been trying to develop have been trying to put these kind of tools in the hands of social scientists where they can sit in front of a map, draw an area and say what's happened here over the last 10 years, five years, ten months. In the right-hand side you see where we'd

like to go, this is based on a system the city of Chicago built for the NATO supplement a couple years ago where the mayor wanted to ask the question if I point to a map what's happening here right now whether social media 311 calls, 911 calls, where the police cars are right now, stuff that's not published out in the public but that the city knows. Now they want to ask a different kind of question, which has to do with predictive analytics and that is can I look at the map and can you show me areas of the city where something unusual is happening or something unusual might start to happen soon based on some leading indicators. So a lot of the work we've been trying to do is based on that. I'll say a lot of this data as Claudia said, it's not clean, it's not complete. It's not collected necessarily for the purposes we're using. There's some very good work that Rob Samson is doing at Harvard to look at and adjust for bias in the 311 data. If you just went by 311 frequency on potholes you would conclude that places were young well to do people lived were the worst roads and that's actually not the case. Because they happen to use 311 more than others. Working with San Francisco on other ways to use data to measure things that are not directly measurable, such as the list you see in the upper right, and that means bringing in econometric models and pedantic models and things like that to figure out what is this data telling us. And for that you need, not just machine learning people but sociologists and economists. Similarly with both Chicago and Memphis, where we're trying to inform policy with respect to Chicago it's vacant properties. Where do you invest the small amount of money to buy vacant properties, where you can have the largest improvement in the health and safety of the neighborhood and the people that live in it. With Memphis it's where should we put our infrastructure investment over the next 10 years based on where we're seeing growth in productivity happening. All this, a lot of this is based on some very good work by Dan Neil at Carnegie Mellon University. And then I'll give you three examples and then close with the city of Chicago that gives you a sense of where policymakers are in predictive analytics. In the top one, rodent infestations essentially looking at 31 different 311 calls saying where are we going to see rats in seven days or less, rather than sending the abatement people out based on calls now they're sending them out based on predictions. That's some things that you might actually predict yourself and that's my timer. And then things that aren't probably actually laws of nature like the common name and store name is probably not the real reason that it has a higher probability of black market cigarette sales. I'll just close here. Eric is not here, but he has funded this wonderful work in Chicago that Rod Ganny is running, called data science for social good doing projects that range from looking at Mesa public schools data and saying where are our kids under matching at schools where they could go to better schools and they're not and why is that to alleviate crowding on buses by giving the Chicago Transit Authority tools for predicting the impact of policy change and equipment change and schedule change on crowding. And the real point here is that these 40 students that came last summer were picked from 560 applications, all of which came in with a two-week period where nobody knew anything about the program, we announced it through social media. Two weeks later we had 560 students to choose from. This year we'll have 50 students. And the point here is that there's a generation of people here that are very interested in using data to make the world better. So we want to capitalize on that and train them so that they can use these tools appropriately and so this is a program that's not just about learning to use the tools, trying to figure out how to take a business problem or opportunity and some data and put them together and learn something useful to a real organization. Thank you.

>> Duncan, please.

>> Duncan Watts: Great I'd like to start by making a broader point about social science and then narrowing down to some of those and reinforcing some of the specific points that have already been made. So there's the broad point is this: If you look back over the last sort of century of social science, we see a tremendous proliferation of theories. You can pick any major topic in social science, there's hundreds of theories by economists, psychologists, sociologists, political scientists. So there's plenty of ideas about how the world works. But what you don't see is a cohesive cumulative and empirically tested body of theoretical knowledge. In some sense we're generating as enormous number of ideas about how the world might work but we're not coalescing on any cohesive ideas about how the world does work. Of course there's exceptions to this rule. But I think if we think about mechanism design applied to actions, this is a pretty successful body of theory but I say the reason why it's successful is because it's a very artificial setting. It's something that we create and it works the way we design it. I think several years ago macro economists might have said that they had this kind of success with monetary policy but they don't say that anymore after the financial crisis. And there are some sort of I think smaller exceptions like if you think about nudging or the poverty action lab at MIT they can make some claims for success but they're very limited in their applicability. And the overall result is when we think about social practice how people in business or government or policy go about what they do, it is strikingly uninformed by social science. They are essentially all doing things that social scientists study and they do not pay attention to what social scientists have to say. Nor should they in my view. This is true in sort of designing and conducting marketing campaigns, optimizing organizational performance, this is something that we're trying to do at Microsoft right now. We're going through a massive reorganization. I'm pretty sure maybe Craig can correct me that the senior executive team at Microsoft did not go and consult the organization science literature before deciding how to reorganize the company. Enhancing collective action, resolving conflict, predicting market demand public opinion managing systemic risk and financial systems and designing communities and cities as Charlie was just saying. These are efforts that are based on sort of hope and intuition and these stakes here are enormous. There's clearly sort of an important salient gap between what social science ought to be trying to do and what it's actually doing. This is true for development, economics. So why is this? I think there's lots of reasons that you can point to. But the sort of core one I think is scientific. First of all, when you think about individual human behavior it's very complicated. We have sort of for centuries tried to figure out why people do what they do and how to make them to do it differently. When you talk about social phenomena you have a complexity that's different you have individuals interacting with each other in complicated ways to produce collective entities the things that social scientists study are collective entities: firms, markets, cultures, political parties social movements, audiences -- we talk about them as if they're single things as if they act as if they have intentions but we also know that at some rock bottom level they're all aggregations of individual people with their own desires motivations and plans. This is what the sociologists call the micro macro problem we go from the micro level of individuals to macro of social fact. Elsewhere in science we call it emergence. This is something that Chris was alluding to earlier. It's certainly true that emergent phenomenon arise everywhere we look in science and everywhere we look in science where we run into emergent phenomenon they're difficult to solve. I think one way to express why social scientists had trouble keeping up with the natural and the

natural and engineering sciences, is that every problem in social science is in some sense a multi scale problem that involves emergence. You can't understand how an organization functions without thinking about individuals, without thinking about collectives without thinking about hierarchy and chains of command, without thinking about markets and competitiveness and regulation and government, there's a multi scale problem. If you throw any part of it out you in some sense lose the essence of the problem. These are extremely hard problems to study empirically, first they are hard just because of collecting data at an observational level of thousands or millions or hundreds of millions of individuals interacting with each other and behaving over time is already an extraordinarily difficult task. Now if you want to understand causality and run experiments on this same sort of macro scale, you have all sorts of other difficulties both practical and also ethical. And going back to the scientific method, it's very hard to do science when you can't measure what your theories are about and you can't do experiments. So hence we have sort of a lot of the first part of the scientific method which is hypothesis generation and not too much of the second part of actually testing and verification. So why is today exciting, why is now an exciting time to be doing social science? I think you've already heard from this panel that there is this technological revolution we've been living through in the past 15 years which is starting to lift some of these historical barriers, we're seeing an enormous proliferation of new types of data on a scale and on a scope and on a granularity that was unimaginable just 15 years ago, and this has been generated almost by accident by all these digital platforms that we use to communicate with each other and to shop and to express our opinions and to generate content and do all the kinds of things we do an everyday basis. A second thing that has attracted less attention which I personally find very exciting is the ability to use the Web to create in a virtual environment what you're all familiar with from universities, the typical sort of psychological experiment done in a lab using university undergraduates usually. This sort of lab experiment can now be transported into a virtual environment and we can run these experiments on a much larger scale and faster time scale than we have in the past. And as Claudia mentioned earlier, we also have the ability to conduct experiments increasingly in naturalistic environments which is what we call field experiments. So these two capabilities together, I think, have the ability to revolutionize not just society itself but how we study society and those of us who are sort of prone to metaphors point to the telescope as a potential sort of analogy, but also I think the collider is another analogy of being able to do experiments on a large scale. And I think all of this is very much complementary to the notion of doing modeling and simulation either with traditional mathematical theory or with computers. So with all of that said, and it's very exciting, it's a great time to be doing research in this field, and we have a lot of people piling into from physics and computer science, entering and working on problems that were the traditional domain of social science, we have new conferences and research centers and journals and phrases that have ended the vernacular like computational social science or network science or big data. There's a lot going on. But I would say there's a sort of frustrating and stubborn fact that in spite of all of this progress and the tens of thousands of papers that have been written, you know the problems that I throw up on the first slide we really are no closer to understanding these big questions of social policy than we were 10 or 15 years ago. We still have no idea what to do about managing systemic risk in financial systems and the academic literature on that is kind of a mess. And emerging epidemics, cultural change, organizational performance, we really haven't even begun to scratch the surface of the sort of big serious substantive questions. And even simpler questions about why do people change their minds when they do is sort of

notoriously difficult to answer. So how are we going to go from where we are now, where we're starting to see what can be done to where we need to be actually solving some of the questions that people in the policy world and in the general public really care about. And I would point to a few possibilities for or directions for future progress. So I have four points specifically to make. The first is that in spite of the enormous volume of data that we are generating from all these digital platforms, what is sometimes called digital breadcrumbs, it's sort of badly compromised by the fact that these different modes of behavior are all collected on different platforms. They're all essentially contained in different silos. So Facebook has wonderful data about self-reported social networks. The search engines have great data about what people are thinking about through the search queries. Amazon and eBay have a lot of information about what people are buying. Ratings companies like Nielsen have a lot of information about what people are watching on TV or using their shoppers club card, et cetera. E-mail providers have a lot of social data. But all of these different modes of behavior are owned by different companies and they're essentially inaccessible to each other and many of the questions that are of interest to social science really require us being able to join these different modes of data and to see who are your friends what are they thinking what does that mean that you end up doing you cannot answer this question in any but the most limited way with the data that's currently assembled. And now this is I think a very sensitive point, because to some extent this is what the NSA has been reputedly doing, joining together different sorts of data and you can understand how sensitive people are about the precisely the reason why this is scientifically interesting is also the reason why it's so sensitive from a privacy perspective. Figuring out how to do is enormously tricky you had a session dealing with privacy issues. They get much more complicated when you start joining different sorts of data together. Nevertheless, I think it's something that we need to think about. Another big problem that is less appreciated is what you might call the algorithmic confounds of digital data. That when you go to Facebook, you're not sort of seeing some kind of unfiltered representation of what your friends are interested in. What you're seeing is what Facebook's news ranking algorithm thinks that you'll find interesting. So when you click on something and the social scientist sees you do that makes some inference about what you're interested in and why you're sharing what you're sharing, this observation is, in some sense, hopelessly confounded by the fact that Facebook also has a theory about what they think you'll be interested in and to some extent you're just performing out that theory. So this is a huge problem, I think. In some sense it's always been a problem with evaluating the effectiveness of social policies. But it's something that's deeply embedded into all Web-based platforms. So one possibility as some of my colleagues urged we should not actually use Twitter and Facebook and other sort of wonderful sources of social data precisely for this reason and that we actually need to go out and collect our own data and design the data collection the way we wanted it to be designed. Building on that, I think the experiment is another answer. And if we look we already have a lengthy history of experiments in experimental psychology and behavioral economics. But for these going back 80 years now, these experiments are really defined by the unit of analysis being an individual. You come into a lab, you're manipulated in some way and you give your results and then you leave. At best you might study a small group of a small handful of people. It takes enormous amount of time is to design and to execute the experiments so if you look at the behavioral economics literature you'll see one sort of manipulation on a public good game published in the paper and a year later the next sort of minor variation on the same theme will come out. So you're sort of thinking about a time scale of a year between when I test my hypothesis and

then when I get to test my revised hypothesis. So one thing we can think about doing with the virtual labs idea is to first scale up the number of people that we can get in a lab at the same time to potentially hundreds of thousands. This completely changes the kinds of questions we can ask. The other way we can scale things up is by shrinking the hypothesis testing cycle. So going from something that's on the order of a year to something that's on the order of a day or hour will dramatically increase our ability to essentially search the parameter space for some of these problems. There's all sorts of new research designs that you can think about that run for longer, that allow us to keep track of individuals over long periods of time that give us novel sampling insight. As many, many directions that we could go if we could solve this problem of scaling up lab science on the Web. And the third point is that both the observational work as well has the laboratory work could be aided by more traditional simulation modeling approaches. So as Claudia mentioned, the weakness of all modeling essentially is that it rests on assumptions which might be questionable and in the case of agent-based modeling in the social sciences, rarely gets tested against data in any sort of detail. And so again we have this sort of proliferation of theoretical models that exist more or less in isolation of each other and often are incompatible with each other and that these incompatibilities never get resolved. So I think one thing that we can do with some of these new sources of data is not just calibrate the models, test the models that we have and possibly reject some of them, but also use data from human subjects in experimental settings to calibrate agent-based model. This idea that agent-based models could be empirically informed. And then the benefit of doing that is that the modeling, the simulation environment allows us to much more rapidly explore different parameters and discover new areas of parameter space where we think interesting things might be happening essentially do hypothesis generation with the simulation models, and take those results back to the lab and test them so you can see how this might operate in something like the scientific method. Finally, and I'll try to be quick here, I think I'll return to a point that Charlie made that we really need to think about changing the institutions that do this kind of interdisciplinary social science. That when the people who are good at dealing with data typically don't know the right questions to ask and the people who have the right questions to ask typically are not good at dealing with data. So we need to have ways of bringing them together. We also need like I guess that was really my second point. The first point is that we need to think carefully about how to manage all of these new types of data that we have. Even just sort of standards for collecting it and storing it and sharing it is sort of really new for us in the social science world. We tend to sort of take our data sets and sit on them and not share them with our colleagues. A third point that I think we do very poorly in social science we need to do better at is really coordinating our research designs across different methodologies. You take any big problem in social science there are people who do mathematical modeling people who do ethnography, people who analyze quantitative data, people who do experiments, none of these people talk to each other. They don't publish in the same journals. And there's certainly not coordination across research designs, I think that's a really important step that we need to take. And maybe we need to take it in entirely new kinds of research institutions and maybe something like Charlie's organization or the cusp which exists at NYU a model we can follow. There are possibilities some public/private partnerships that might be useful to pursue, especially around this issue of data sharing because so much of it is generated in the private sphere. Maybe we need something grand like a Janelia Farm (research campus) for social science. It's a nice place, I'm sure. Wouldn't mind spending some time there.

>> We're going to have to wind up pretty soon.

>> Duncan Watts: I think I'm done. So thank you.

>> Bill Press: I'm sure my PCAST colleagues are bursting with questions but we've fallen 20 minutes schedule. We'll interact with you after the session. As individuals we'll do that. I'd like to ask my colleagues to thank you for your presentation. [Applause].

Public Comments

>> John Holdren: Well, we are now at the point in our meeting where we entertain comments from the public and our other Vice Chair Maxine Savitz will be presiding in the public comment session. Maxine.

>> Maxine Savitz: Thank you John and we have five speakers today, most of them related to the antimicrobial bacteria resistance so while people are getting settled we have a larger number than we've had in some of the past. First speaker is Amanda Jezek, Vice President of public policy and government relations, Infectious Disease Society. The way we work this is come to the table and, while you're getting settled it doesn't count against your timing. And essentially you have two minutes and I'll give you a warning after 90 seconds, please proceed.

>>Amanda Jezek: Thank you very much for inviting us to give public comment today. The Infectious Disease Society of America, representing over a 10,000 infectious diseases, physicians and antibiotics is very happy to share our views on the issue of antibiotic resistance. The interagency task force on antimicrobial resistance was established to help coordinate the federal response on resistance. But the ITFAR needs a high level centralized leader and dedicated funding. IDSA calls for a an ITFAR director to be appointed in the Department of Health and Human Services and an ITVAR advisory board of non government experts to be established. The European Union is far ahead of the United States in encouraging antibiotic innovation. The EU has launched an impressive public/private partnership "new drugs for bad bugs" which brings together government, academia and industry for an unprecedented collaboration. IDSA urges the US government to establish a complementary effort. IDSA also recommends a new tax credit to support antibiotic R&D. Analysis by Ernst & Young found our proposal would yield an additional five to seven new antibiotics in the pipeline every year. IDSA also encourages the establishment of a limited population antibacterial drug or LPAD pathway similar to the special medical use pathway that PCAST previously recommended. LPAD would allow new antibiotics to treat serious or life threatening infections for which there is an unmet medical need to be studied in smaller, more rapid clinical trials. LPAD drugs would be approved for use in limited populations

>> 30 seconds.

>> For whom they were proven safe and effective. As we stimulate more antibiotic R&D we must also combat resistance through research and public health activities. Real time publicly available data on antibiotic resistance and antibiotic use are critical and we support increased funding for these efforts. Antibiotic stewardship is a critical tool to protect antibiotics from misuse and we recommend that all healthcare facilities be required to implement stewardship as a condition of participation in Medicare. And I would hope that you'll all see our written comments for further detail on these issues. Thank you very much.

>> Before you go --

>> Eric Lander: You're done with your two minutes but we have questions.

>> Amanda Jezek: There's much more I would have said.

>> Eric Lander: But you supplied this to us in written form. And that's great. This Ernst & Young study, will you be able to supply us with that?

>> Amanda Jezek: Absolutely, I'd be happy to.

>> Eric Lander: That would be great. And are there any other materials that you think you should supply us with?

>> Amanda Jezek: So in 2011 IDSA put forward a combatting antimicrobial resistant report that provides detailed information on these policies on recommendations, it's available on our website and I'll be happy to provide that as well.

>> Eric Lander: Great. If you could try to provide that today that would be super. Thank you.

>> Amanda Jezek: Be happy to. Thank you.

>> Maxine Savitz: Thank you very much. Our next speaker is Richard Carnevale Vice President, regulatory scientific and international affairs. Infectious Disease Society.

>> Thank you. Hi. I'm Dr. Richard Carnevale with the Animal Health Institute. We're the trade association that represents the manufacturers of animal health products. I appreciate the opportunity to make a few comments with regard to antibiotics and food producing animals. As Dr. Lander mentioned this morning FDA in December published a guidance 2013 they called it to ask voluntary request sponsors of these applications, these approved applications to remove claims for weight gain increasing weight gain and improving feed efficiency which are commonly referred to as growth promotion from the labels of these products. They also asked these companies medically important antibiotics I would mention. They also asked these companies to restrict these marketing of these products to use on the order of veterinarian, through veterinary feed directive or prescription. I'm happy to report as Dr. Lander also confirmed that all the companies involved with these products including the AHI members, the generic Animal drug Association members and those independent companies have all agreed in writing to cooperate with the FDA. So we've taken a big step here. And there's a lot more to go. There are

challenges to getting there. But after we are completing these changes, which hopefully will take place in about three years, is that all the products that are considered medically important for use in feed and water will be labeled only for control and treatment of disease under the control of a veterinarian. So we hope that that will take place.

>> 30 seconds.

>> The reason I come to you today is my message to the council is that the FDA process should be allowed to work. We have heard calls for making this process mandatory. And I can tell you I've been around this business for quite a few years, I've spent 20 years in government service between the FDA and the USDA, and I know that the only way to get this done is through cooperative approach with the industry. If the FDA were forced to go to a mandatory approach, the administrative procedures that would be required for them to make their case and to carry out this effort would delay this many more years and would frankly have an uncertain outcome. So that's my message. That a voluntary process should be allowed to work. It will work. The companies are committed to doing it. And I think we're on our way to changing the way these products have been marketed for many years.

>> Thank you very much.

>> Our next speaker is Lance Price who is a professor at the department of environmental and occupational health at GW but he's speaking as a private citizen today.

>> Thank you for the opportunity to speak today. It's a real honor to be here. I'm Lance Price molecular epidemiologist at George Washington University where I use genomics and big data to study emerging antibiotic resistant bacteria so this feels like the right place to be today. I was pleased that Dr. Lander mentioned several times the importance of antibiotic use in food animal production as part of this formula to bring down resistance, because the research that we've been doing and several of our colleagues have been doing only underscores the interconnectedness between food animals or livestock and humans when it comes to antibiotic resistant bacteria. And not just the classic foodborne pathogens but colonizing pathogens like E Coli. and staph aureus, which I think are critical to this formula. So like you I was also encouraged by the FDA's new guidance, but I also have two, at least two major concerns about this new guidance. One is that we lack the granular data necessary to evaluate the success of this program. We don't know really all the drugs that are being used, in which animal species, what quantities, for what purposes, duration, and frequency, we need the data to evaluate the success of the program. And my second concern is also related to that, that without these data, I'm concerned that the industry may make nominal changes that is switching from antimicrobial growth promotion to routine disease prevention and the actual quantities of antibiotics won't be minimized or won't be reduced in any meaningful way and bacteria don't care what you call the antibiotics. In closing I know I have 30 seconds in closing I'd like to thank you again for the opportunity and to ask you to read the letter included in my written statement that's cosigned by several esteemed colleagues like Stewart Levy who dedicated their lives to studying antibiotic resistance and share my view that it's critical to address the antibiotic use in food animal production if we're going to address resistance in people in general.

>>Eric Lander: All these wonderful people I can ask questions to. Do you know of any research that really answers the question what proportion of human antibiotic resistance can be traced to agriculture? I certainly with much less study than you don't know the answer to that question do you think we have a definitive answer to that question.

>>Lance Price: I'm glad you asked that question that's what I've set out to do. I can set up studies to start to quantify how often certain things happen but I think it's actually an impossible and maybe even a logical question to ask. Because let me give you an example. What if we create a new strain or a new strain of bacteria evolves into food animal production due to antibiotic use, spreads to people through one transfer from animals to people and then gets established in the community and hospitals and kills a million people. Do we quantify that as one person associated with food animal production or is that a million people? And then when you think about the transfer of resistance genes, you know, gene from a benign bacteria from food animal production that gets transferred to a human pathogen adapted to humans and then goes on to cause millions of infections, drug resistant infections, how do we quantify that.

>>Eric Lander: There's multiple causes to a disaster, it might arise here due to an issue in agriculture, it might spread in the human population we can sort out Providence versus spread. It would be great if you could send us any material you have on what we do or don't know. I'm encouraged to hear that you're working on genomic approaches which should at least shed lots more light on it.

>>Lance Price: We're seeing lots of light.

>> Eric Lander: Send us all the light you can shed. That's great. We'll take all light, we endorse knowing more. So thank you very much for coming.

>> Lance Price: Thank you for your time.

>>Maxine Savitz: Next will be Elizabeth Youngman director of drug safety and innovation at the Pugh Charitable Trust.

>> Thank you for the opportunity to comment on the council's important work to address the urgent threat of antibiotic resistance. My name is Elizabeth Youngman and I direct drug safety and innovation at Pugh Charitable Trust, an independent non partisan research and policy organization. Pugh works to advance policies to reduce unnecessary use of antibiotics in food animal production and to spur the development of new anti-bacterial drugs for people. My focus today is drug development. As you already know we have a pressing need for new antibiotics. Doctors already face patients with untreatable infections and threats such as NDR 1 and NCRE potential for worse to come. Pugh's analysis of the antibiotics pipeline which is available on our website and will be updated regularly suggests some reason for optimism. About 45 new antibiotics were in clinical testing as of March 2014. However many drugs in early phase testing don't pan out. Perhaps only one out of five or even one in ten will reach the market. To keep up in the race against resistance newer and better antibiotics will be needed but scientific, economic and regulatory barriers all slow the pace of antibiotic drug development. Scientific questions like how to get drugs into gram negative bacteria are hard and greater investment and

coordination is needed to help solve them. We urge the council to consider this. The Gain Act passed in 2012 is a step towards addressing the economic barriers by providing additional exclusivity for some new antibiotics. A remaining challenge is regulatory. PCAST in its 2012 report recommended an approval pathway for drugs in a limited population of patients. This approach has been called special medical use or applied to antibiotics LPAD for limited population antibacterial drugs, intended to aid antibiotic development by making testing more feasible and less expensive. Pugh has extensively vetted this concept with drug companies, FDA, providers and payers. Along with the Infectious Disease Society of America we issued a core set of policies to guide the establish of such a pathway to including the need to foster for labeling effective labeling to foster judicious use of LPAD products. In addition to creating a regulatory pathway for drugs that might not otherwise have one the narrow indications contemplated by LPAD can allow for premium pricing for high need antibiotics which would address economic incentives as well. Renewed PCAST support for this type of pathway could help create an option for getting antibiotics to patients who have no other option.

>>Eric Lander: I make the same request for you with regard to your noted things here about core principles for LPAD and your interesting comments about premium pricing and how that might work, if you could try to today provide any materials on us. You actually cite some such materials.

>> Can I take the opportunity to thank, we had five speakers in the public session. All -- we're done.

>> One more.

>> I'd like to thank the people who took the trouble to talk about antibiotics at this very useful time and send us materials. We like reading things.

>> Maxine Savitz: Our next speaker and last one is Robert Levin, chief executive officer of Transclick incorporated.

>> Thank you for the opportunity to be here. Good morning. My name is Robert Levin. I'm CEO of Transclick an innovator in digital platforms for the developing world, Transclick is a tech laureate of the Tech Museum of Innovation and a tech pioneer of the World Economic Forum. Today my comments will focus on what PCAST may do to coordinate and support the federal government's role in advancing and incentivizing innovation and coordination with the next wave of entrepreneurs of digital health tools, apps, sensors and cognitive computing for medicine. So I recommend that PCAST consider extending the roadmap report issued in December 2010 on digital health with a new connected health initiative which would call for a mandate to increase collaboration with the nation's digital health clusters in New York, San Diego, Silicon Valley, Boston and elsewhere, to build, to help continue to build the connected layer of medical devices, medical sensors, medical apps, both regulated and non-regulated as well as cloud-based cognitive computing, like Watson and others. There is an enormous wave of innovation in the digital health start-up community including university incubators and commercial accelerators many of which started since the 2010 report and many of them are focused on connecting medical devices, medical apps, medical sensors and cognitive computing to electronic health records. So I believe that the federal government has the opportunity.

>> 30 seconds.

>> To build on the successful smart health research programs of the National Science Foundation and increased allocations to SBIR and STTR programs as well as increase incentives and outreach on behalf of military health systems like Tricare and the Veterans Administration health care system to incorporate some of these innovations. For example, the Veterans Administration is rolling out the mobile care provider program including a digital health app store for 18 vet centers there's a commission being discussed in Congress called the Healthcare Savings. Through Innovative Wireless Technology Commission which will examine scientific research regarding medical effectiveness of digital health technology. So in the next 20 years as our global population grows to over 9 billion people there will be at least 10, 20 million shortages of doctors and nurses, and there's an opportunity to leverage American innovative healthcare technology to supplement this shortage and healthcare capacity around the world.

>> Thank you.

>> Maxine Savitz: I'm going to thank all the speakers for their very timely talks. The first four as Eric mentioned on the antimicrobial but also PCAST has a study going on about applying systems engineering approaches to the healthcare industry in the same way it's done in many others, and part of that is the whole digital way electronic. So very timely. All of them. And really do appreciate the public speaking and getting a larger number of a group to talk to us, good. Thank you all very much.

>> John Holdren: Thank you Maxine. This brings the PCAST meeting to a close. We're grateful as always to the PCAST members for their participation and attention and observations. We thank the invited panelists and the public speakers, the OSTP staff and, of course, the wider community who have joined us both in person and through the webcast. Until next time. [Meeting adjourned]