PCAST Meeting – January 31, 2014

Welcome

John Holdren: Okay. Let me call this meeting of the President's Council of Advisors on Science and Technology to order. It's a pleasure as always to welcome the members of PCAST, the staff who are with us from the Office of Science and Technology Policy, the members of the wider science and technology community who have joined us in the room and, of course, people who are joining us over the webcast. Welcome to all. Let me start by saying that I think I can speak for all of the PCAST members in saying how delighted the group was with the remarks of President Obama in his State of the Union address on Tuesday night. His remarks on the importance of continuing federal support for basic research, his remarks on climate change and clean energy, his remarks about education, and workforce training in particular. PCAST remains enmeshed in a very demanding agenda. In subcommittee and preparatory meetings over the last day we have addressed a variety of topics, including PCAST engagement in the Quadrennial Energy Review, which the President has announced and was originally recommended by PCAST in one of our earlier reports on energy technology innovation. We have engaged with issues of systems engineering to improve outcomes in the healthcare system, talked about our assignment from the President to look at technological aspects of the nexus of big data and privacy. We have looked at continuing challenges in R&D policy in an era of flat or declining budgets, technological change and the jobs of the future, another topic that PCAST is thinking about. The use of educational information technology for workforce training is yet another. But we have a rich agenda for our public meeting this morning, starting with a couple of sessions on different aspects of improving scientific reproducibility in an age of international competition and big data. And I'd like to call on my co-chair, Dr. Eric Lander, to moderate these sessions on scientific reproducibility. Eric.

Improving Scientific Reproducibility in an Age of International Competition and Big Data I: Researchers

Eric Lander: I'd like to welcome all the members of PCAST and all of the people who have joined us here in the room today. And everyone who is currently or will at some point be joining us on the Web to view today's session. We have an interesting topic first up. We have two sessions on it on scientific reproducibility in an age of international competition. The topic is actually put on our agenda by Bill Press one of the vice chairs, who noticed a number of different articles and discussions on this topic. So I'm going to start before introducing the panel to just to turn to Bill Press to make some framing remarks then while he's doing that I'll ask our panel to come up situate themselves and I'll introduce themselves after Bill gives a bit of background.

Bill Press: Thanks Eric. It's unusual that PCAST schedules two one-hour sessions on a single topic and I think that indicates the importance we give to this issue. All science is based on the premise that there's an objective reality that is exposed by scientific experiment. And so it strikes at the very heart of what science does if experiments are not reproducible. And there have been a number of recent studies that may indicate that there is an increase in the amount of unreproducible science that's finding its way into the scientific literature. So in brief, the purpose of this session is simply to explore with the help of three distinguished panelists in this first hour and two distinguished and I might add powerful editors in the next hour, is this the case and why might it be true and, of course, what should we responsibly be doing about it. My view is that irreproducibility is not one thing. It's a kind of a catch-all for many things. It's a little bit like a nonspecific symptom of a disease, fever or something that could have many causes. And without doubt the cause that gets the most media attention is those quite rare cases of intentional scientific fraud. Quite rare but there is some evidence increasing since, for example, the 1970s. Now, intentional fraud at least that which is later discovered but since later can mean many years, when I say very rare, I mean that when people study this, they measure it in numbers like less than 100th of one percent of published papers. So fraud is a problem in science. We have to root it out. We have to understand what are factors that may be increasing its intensity. But in my opinion an equally important problem, probably a more important problem, is when irreproducibility in science is an unintentional result of other factors. Whether it could be due to

publication biases where there are selective pressures to publish a certain kind of result and not a negative result, for example. And that, of course, will skew the statistics of what we believe and what we don't believe. Or possibly, if irreproducibility results from poor training of a generation of scientists coming up or it may be in a sense doubly unintentional. It may be that we're not capable of giving the right training yet because the methods of experimental science have been advancing so rapidly that they're outstripping the training that people have. If instrumentation turns over in a radical way every five years but people are trained in statistical methods only once-in-a-lifetime every 30 years, then there's clearly a disconnect there. These are the kind of disconnects we need to fix. Big data itself, I think, in this as in so many other areas introduces a whole new range of problems. In big data, statistical errors in the classical sense of errors that get smaller with the size of the data set, statistical errors may disappear and be irrelevant whereas systematic errors, biases in how the data is collected, may be everything. And yet I think it's probably true of most of the scientists in the room that are training and dealing with imperfection in data, is largely in understanding statistical errors and much less in systematic errors. So these are just my thoughts, my biases coming in. And the purpose of this session is to explore thoughts of some experts and to have a discussion in PCAST to try to bring to the front what are the issues, the causes and possibly the remedies for this. With that, Eric, I hope you would introduce our panel.

Eric Lander: I look forward to that. Thank you for that taxonomy bill. One could add to the taxonomy. There is also, I couldn't reproduce your work and it's because somehow I didn't actually set it up the same way. And that's a really interesting question as well. So there's a whole range of issues. We invited three speakers here for this first panel, who all bring expertise and have written and talked about the wide range of different types of irreproducibility there might be and where there might be very specific statistical cures for it and in one case I think we'll hear about that. But we're going to start off with Glenn Begley who is the Chief Scientific Officer and Senior Vice President of R&D in TetraLogic Pharmaceuticals and who has written very interestingly about this topic. Glenn, all yours.

Glenn Begley: Thanks very much. It's a pleasure to be here. Thank you very much for this opportunity. You've already heard my current appointment the work that I'll describe was performed while I was head of the Oncology Group at Amgen. Before that I had more than 20 years -- I know it's hard to believe when I look so young -- more than 20 years as a medical oncologist and a researcher scientist before I joined Amgen. At the outset I wanted to clear my position. The results that I'll discuss do not, to my mind, challenge the validity or legitimacy of the scientific method. I'm not talking about fraud I'm talking about scientific sloppiness or desperation. The vast majority of investigators do want to do the right thing, they just don't know how to do that. And the fact that this debate is occurring in public, to me, confirms the strength of our scientific system. Cancer is evolution, and I like to tease my cardiology colleagues by saying cancer is evolution, cardiology is just plumbing. The other advantage that they have, of course, is that most Americans at least believe in plumbing. (laughter) When it comes to evolution, we don't get to set the bar. The bar is set by the biology. So there are two parts to this slide. The top part talks to the inherent problem we have with cancer. Our models are very poor. We don't understand the disease. And the disease is more heterogeneous than we like to imagine. There is however low-hanging fruit. That's the opportunity. And this is inherent to our system and it's much more easily addressed. It relates to issues like poor experimental design, selection of results, deliberate bias and so on. That's the subject of this presentation. And I should say at the outset I'm not talking about big data, I'm talking about little data. Often too little data. So we can set the bar in some areas when it comes to these areas, the low-hanging fruit that I talk about. But it's important to understand that we get what we incentivize. Industry relies very heavily on what happens in the academic arena and in the decade when I was working at Amgen we tried to reproduce 53 seminal publications and were able to do in only 47 publications. These were all publications that were describing something completely new. There was a spectrum of irreproducibility; the most shocking to me was when the investigators themselves could not reproduce their own data in their own lab with their own reagents. All we did simply was take their reagents, put it behind our back and give it back to them blinded. They were unable to reproduce their own data published in Nature Sciences. There were two occasions where we could reproduce the specific data but it was not a robust finding. For example, they said two SRNAs could inhibit the growth of two breast cancer cells.

We completely reproduced that. But when we looked at 20 breast cancer cells, the finding was not robust so that was not sufficient grounds to take forward a drug discovery program and I recognize that's a subject of assessment. And there were many examples of data selection where investigators chose the experiment that they wanted to illustrate the point that they wanted to make and ignored the other data that was actually to the contrary. These studies --

Eric Lander: How many there were of that first bullet where the investigators themselves was unable to reproduce with their own reagent?

Glenn Begley: At least a dozen.

Eric Lander: So a dozen out of those 53 you think fell into that category.

Glenn Begley: At least, I don't know the precise number.

Eric Lander: The majority, you would say, fell into the other category of it was reproducible with the reagents but not generalizable?

Glenn Begley: There were two examples.

Eric Lander: Only two examples so the others you don't know what category they fall in.

Glenn Begley: Correct. But in every situation we tried to get the reagents from the investigators and as often they would comply. And then when we tried to reproduce the experiments in-house we were unable or we generated our own reagents. For example, if there was a small molecule inhibitor they had used we might have used a more selective small molecule inhibitor that we had generated in-house. So, as I said, the data that was most shocking to me were the dozen or so examples where the investigators themselves could not reproduce their own data. So this has had substantial impact, principally in terms of the opportunity cost because we spend so much time working on things that really have little value. The papers have, some of them have spawned secondary papers. Some of these papers have been cited thousands of times and had hundreds of secondary papers. And many clinical studies have been initiated based on this work. Unfortunately, Amgen's experience is not unique but a health indicator about 80% of the time they were unable to reproduce published literature and since then both AstraZeneca and Novartis have come out publicly, and others, saying this is also their experience. So in my view this is driven principally by our current incentives. It's a systemic problem. It's not localized to one or two laboratories, it's across the whole field and it's principally because careers are built on publications in top tier journals. This is what drives grants, fame, promotions. The top tier journals want simple compelling stories and positive results are rewarded. There is little recognition of the value of negative studies and there is a positive reward for finding the result that a reviewer wanted because that will guarantee publication in the top tier journal. The investigator and the host institution, to my mind, are ultimately responsible for this. However, the greatest likelihood of change is going to come from the journals and the granting agencies raising the standards for publication, raising the standards for grant applications. So as I reviewed these papers, there were typically six areas where publications failed. Typically they were almost never blinded. The results typically a seldom complete, shown in their completion. I'll give some examples. Experiments were typically not repeated, positive and negative controls are typically excluded. Reagents are seldom validated and statistical tests are inappropriate. I'll give you some examples. So we published the first paper and I received a desperate e-mail from a post-doc saying please reveal the 47 papers because I could be working on those papers. We couldn't do that because the investigators constrained us with confidentiality agreements in place. What I did do was took the top three papers off my disk at that time and analyzed them as examples of what we're seeing in the literature and these are examples but my allegation that if you pick up any issue of Nature Science on sale you'll see these every week. So the first question is were the studies blinded? The graph on the left shows pyknotic cells. This is a highly suggestive

assay simply counting the number of dead cells in the hematoxylin and eosin stained field. It's impossible to do this without it being blinded. It's not surprising then that the investigators found the difference they wanted, the orange bar is different to the other bars. The graph on the right is from a cancer cell paper in 2012 and they were counting metastatic deposits in the lungs. You see two arrows but we've got no idea whether the others areas that were obvious to all of us were counted or not. This is an example of something that has to be done blinded. In the absence of blinding it's not surprising that they got the result they wanted which is shown on the graph on the right. This is from a paper where I've just simply taken two different figures from the same paper. This is a western blot. If you look, for example, at the top panel the graph on the left is supposed to show the EGF receptor the panel on the right has got at least three or four bands. This is from the same investigators, the same lab, using the same antibody. If you look at what's highlighted at the very bottom you can see that the graph on the left, the doublet is of similar intensity, the graph on the right, the bottom band is less intense. And if you look up and down fourth from the bottom you can see on the left there's an only singlet. On the right they are doublet. These are from the same antibodies by the same investigators in the same paper. And it's hard to know whether or not these experiments were ever repeated. There are no size standards on the gels. We don't know if the antibodies were validated and so on. This is from the Nature Genetics paper that was on the top of my desk. You can see on the graph. This was a Kaplan-Meier analysis. In the blue, four animals that died in the green five animals have died. Of course if you actually look at the data that means two and a half animals died at about day 45. (laughter) two and a half animals died at day 50 and one has to wonder whether or not this was actually read by the co-authors, of which there were ten, the reviewers and the editors. This is published in nature genetics the same paper. It's in color so we know it's true. You can see here that the blue curve and this is a xenograft study, they are measuring tumors. And this looks good until you actually look at the axis. This is tumor size percent. What we would typically expect in an assay like this is a three or four fold change not a 40% change. If we had a 40% change between animals I'd be happy. The errors here are frankly unbelievable, it's impossible to measure tumors with that degree of accuracy and we don't know how many animals were actually examined. Here's another example the graph on the left shows another subjective assay counting viable cells with the crystal violet assay, the errors are almost invisible. Frankly this data is just not believable. In addition, this is cell proliferation which should be expressed on a logarithmic scale not a linear scale and if you do it mentally probably all the curves won anyway. The graph on the right shows another example from the cancer cell paper of tumor growth and, again, the errors are frankly so small it's unbelievable. This data just cannot be believed and we don't know how many animals per cohort. This is from the Nature paper in 2012. There are four figures labeled in this paper showing there's 100 fold increases over control. Clearly what the investigators meant was 100% but it doesn't matter that says 100 fold. You take this at face value there's 100 fold difference in these curves. And again you have to wonder whether or not this paper has been read. This is a perennial problem in the literature. This is from a group that used the Pfizer molecule that's a MET and an ALK inhibitor. They acknowledge that it's an ALD inhibitor but say that data is not important they don't need to show it to us they fail to acknowledge there are 16 kinases inhibited by this molecule and the entire focus of this paper is the MET because that supports the hypothesis they want to support. This is immunohistochemistry from those same three papers. The figure that you can hardly see is from the cancer cell paper. That's supposed to be showing cell surface expression of MET. I frankly can't believe that was published but that's from the paper. The Nature Genetics paper below, the second panel from the bottom shows a polyclonal anti-peptide antibody with no controls demonstrating that that's legitimate. The Nature paper claims on the left that they were using KRAS, a specific antibody against KRAS and that was the basis for their immunohistochemistry. Actually, to their credit the Santa Cruz data sheet said this should not be used for immunohistochemistry. Were the statistical tests appropriate? This is another tumor xenograft assay. Again, this is unbelievable. This data is just not believable. But the point here is the statistical test should really look at the area under the curve rather than comparing these two points. So my conclusion is we have a systemic problem. Our system tolerates perhaps even encourages these behaviors. The principal responsibility rests with the investigator and the host institution and to my mind I'm disappointed that the host institutions have failed to take this on aggressively. Patients deserve and certainly expect more. I would like to acknowledge the people that actually did the work that wasn't me and a number of colleagues that encouraged me to proceed with these and my recommendation is that investigators and institutions who bear the principal responsibility should insist that

studies within their departments are blinded, all results are shown, we shouldn't have western blots where we crop out the majority of the gel and only show the band that you want to show. Experiments should be repeated. We should see positive and negative controls and reagents should be validated and we should have appropriate use of statistical tests. Thanks very much.

Eric Lander: Great. Thank you. Next we'll turn to Donald Berry Professor of Biostatistics at University of Texas, MD Anderson Cancer Center who, I suspect, will also touch upon things having to do with cancer but probably more broadly.

Donald Berry: Thank you, Eric. And thanks to Glenn for saying some things that I don't have to say. I'm going to talk about a number of things, and I'll just hit the highlights in my ten minutes. Regression to the mean. These papers, most of the examples of irreproducibility have to do with regression to the mean. So let me say a little bit about that. This is an artificial distribution. It's a bivariate normal sample for a bivariate normal illustrating the effect. The X axis is the first experiment. The Y axis the second experiment. They're marginally the same. They're correlated, correlation is 50% in this example. And you think, well, if they're the same distribution, just flipped, that the line of confirmation equals initial must be the line that we should focus on, but that's not the line that we focus on. We focus on the regression line. The regression line is what do you expect in the second observation given the first observation? And that's always shifted down. You see rather dramatically in this example. If you focus on the three statistically significant first experiments, they're not statistically significant in the second experiments. There's nothing, no fraud, no nothing, this is just the ubiquitous issue of regression to the mean. So just to show that a bit further, if this is the ellipse of data, regression is not the main axis of the ellipse. It's where you touch the vertical line that is it's the expected value of the second observation given the first, and it's a ubiquitous effect. It's the real placebo effect. Almost every example I know of the placebo effect this is it. If you're into baseball, the sophomore jinx is simply the regression to the mean. This is an example, and there are other examples, this guy made his career in the depression in looking at companies and seeing that companies tended that were doing well tended to do poorly or less well the next time and vice versa. And so he assumed that that meant that we're converging. Carl Galton who discovered regression to the mean who looked at the heights of fathers and sons so tall men tended to have tall sons but not as tall as themselves and correspondingly on the short side. That would mean we're all converging into a particular thing. Remember, I said that the marginal distributions were the same. It doesn't happen. So this book, the triumph of mediocrity in business, was just completely wrong.

In cancer, we have the worst phase three record. We build -- well, none of the records are that good across the therapeutic areas. These phase three studies are built to have power 90%. You think maybe 90% success. Not so. The failure phase three is usually is at least partly in every case because of a bias selection of phase two trials. We take those trials, those drugs that look good on the basis of phase three, and go into phase two and go into phase three. And there's an inevitable regression to the mean.
Eric Lander:

Eric Lander: When you're using the word bias, because there's lots of people watching this on the Web, you mean not intentional bias but it looked good in phase two and it was a little bit lucky it was looking good in phase two because of the distribution, you took it forward and it was actually worse in phase three.

Donald Berry: Yes. It's not intentional. I agree completely with Glenn. These are almost always, this is just ignorance. But it's possible to understand that bias. And here's an example. This was a stroke trial. The only information I had is my prior distribution of stroke therapies which if you know anything about stroke they've been greatly unsuccessful. If you take that into consideration and read the New England Journal of Medicine article on something called Saint One. This was an experimental agent in stroke, it's statistically significant P value .038 and powered the next study whose power is 80% to see the same result that they saw in the first study. And they increased the sample size a bit. And it was powered at 80%. My probability of success based on just this New

England Journal of Medicine article in the first study was 10%. 10%. In comparison to the 80%. And, of course, it failed. And one of the things I'll tell you is that in any kind of study, the things that are untold are probably more important than the things that are told. And you may ask yourself so what is Berry not telling me now. Here's an example where I've been successful, but I tell you there are many examples like this. This turned out in a second study much to the chagrin of stroke investigators as well as the company. placebo actually beat the experimental agent. Multiplicities. So I want to give you an example. I took a study and looked at 20 markers, 1500 women with no positive breast cancer. I then -- so 20 doesn't seem big. But if you take the number of possible combinations of the 20, there are over a million. And big, sometimes is two. We don't know how to handle even two markers in many circumstances. But I took, I had a protocol. I took 20 markers and I selected on the basis of their prognostic ability, and I built the recurrent score using Cox regression and I got wonderful results, statistically significant highly so. If you further divide into quartiles and you see many papers that do this very thing, I got great predictability. Highly prognostic. The punch line is that my markers were complete random. White noise. I just made them up. I just listed distributions and I picked the ones off and I'm able to get this amazing predictability. Of course, in any confirmation study it's going to be not borne out. The good news is I had a protocol. I could go back and I could say how likely was it I could see this; I could adjust my P values accordingly. I want to say a little bit about the process of learning when journals consider papers, for example. So this is -- this is my prior distribution. It's in some circumstances it's appropriate. Somebody starts out looking at a particular phenomenon like MET, for example. My prior probability distribution of their effect size, there's a big spike at zero. And the investigators would acknowledge this. But some hoped it would be possible. The X axis is arbitrary. Suppose you do a modestly powered study. The probability in view of that spike at zero, the probability that it's going to be positive is small. 5% in this particular example that I did. But if it's positive, then this is the distribution that you see. This may be surprising to you, you get a statistically significant result and the probability dropped from 90% to 45%. It's still huge. The probability that it's null is still huge. That's one of the reasons we insist on confirmation. If you do confirmation, if you do another study the same, there's some good news here. You do mostly get rid of this null hypothesis probability and there's a shift in the effect size. So it's looking more positive. That brings me to my final slide, I wrote a paper in Multiplicities in the JNCI and I ended with a list of recommendations for investigators and for journals and leading the list is you have to have a protocol. Stat analysis plan where you say this is what I'm going to do and you do it. You indicate to the reviewers and the readers any analysis you specified that you didn't do. And you keep a log of the actual analysis. Sometimes you have to go beyond. You can't be completely constrained by the protocol but you have to say, you have to tell everybody what you did that wasn't in the protocol. And why you did it. And keep a log of the actual results and there are a number of other things here on which I gave advice. But my ten minutes are up so I'll stop there. Thank you.

Eric Lander: Now our third speaker. This is great, we have very different views about what are the questions of irreproducibility, what are its causes and Daniel MacArthur is an Assistant Professor, a geneticist at Harvard Medical School at Mass General Hospital and an associate member of the Broad Institute. Daniel.

Daniel MacArthur: Thank you very much, Eric. My background, as Eric mention is human genetics and genomics. In this presentation what I wanted to do is draw some lessons from the history of my field to the more general issues surrounding scientific reproducibility. In particular what I wanted to focus on was the history of the field of the genetics of complex traits and complex diseases; these are things like Type II diabetes and rheumatoid arthritis where many different genes and environmental factors converge on causing risk of these diseases. I think it's safe to say that prior to about the mid 2000s the genetics of these traits was more or less a scientific baseline in the sense we had done thousands of attempted association studies and each case starting with targeting testing of a variant that was considered to be biologically plausible normally, testing those in hundreds or smaller numbers of disease cases and control samples and looking for frequency differences. Again many thousands of papers have been reported showing associations between genes and these diseases and almost none of these actually consistently replicated. So the challenge that was facing the field was firstly, and this I think was perhaps not widely understood within the field, was that most allegedly biologically plausible variants have absolutely no association with disease and this arises from the fact that our intuitive understanding of biology is actually not that

great. So we have very weak ability to generate a confident prior probability that a given region of the genome is actually associating with disease. And that combined with a number of challenges in study design to produce this plethora of false positives. Firstly, as I mentioned, the studies had typically small sample sizes, sometimes in the dozens, typically in the hundreds of samples and that increased the probability that false positive associations may arise by chance. Secondly, they're often unmeasured biases so systematic errors arising within the data that were unable to be measured using the techniques that were used in these studies and these would involve things like population structure where there was a difference, for instance, in ancestry between the disease cases and the controls. Typically these studies had no requirement for independent replication built into the system. So there was no confirmation, which was mentioned is a critical step and actually validating that a result is correct. And there were many opportunities for what I call cryptic multiple testing, In other words there were many researcher degrees of freedom in which there were opportunities for researchers to try many different protocols of many different analysis plans and stumble on one that was accidentally statistically significant and publish that one. And of course, these combined with publication bias which was mentioned, the tendency for only those studies which came up a positive result end up in high impact papers to produce this astoundingly large number of completely false positive associations. So in the mid 2000s the situation changed in this field, and this change was driven by firstly a technological advance. The development of cost-effective high through put, genome wide genotyping arrays that for the first time allowed us to look at tens of thousands or hundreds of thousands of common markers throughout the genome and genotype those in hundreds or thousands or even tens of thousands of samples. And secondly we began to understand the patterns of human genetic variation through projects like the Hapmap project led by Eric and others. And these two advances combined suddenly allowed us to begin doing what we call genome wide association studies. So these are scans in which we take many different common variants, typically hundreds of thousands common variants scattered throughout our DNA, we then test these in unbiased fashion typically looking at thousands of cases and controls at least in the beginning moving up relatively rapidly into tens of thousands of samples and over the last couple of years into hundreds of thousands of samples. These association studies have been astoundingly successful at least in terms of identifying variants, albeit typically a very small effect that are robustly associated with these complex diseases. So the replication rates for genome wide association studies turns out to be extremely high. This is a very different situation than what we were in in the first stages of this field. So I wanted to illustrate the successes of genome wide association study using a particular disease. This is a slide I borrowed from Mark Daly. This is an inflammatory disease called Crohn's disease what this plot shows is time on the X axis and on the Y axis we have a number of specific or independent regions of the genome which have been shown through genome wide association studies to be associated with this disease. And as the size and scale of these genome wide association studies increased we were able to dramatically increase the number of independent regions that were linked to this disease so here starting in the low thousand of cases and controls to begin with moving up to high thousands and finally in 2012 a very large meta analysis led by Lukia Austins that resulted in the discovery of over 160 different regions of the genome associated with this disease. So what this graph shows is that, firstly, we can actually find many alleged regions of genome associated with these diseases using this approach. But most importantly these results are also extremely robust and replicable. So if we go back for instance and take the 32 variants that were reported to be associated with this disease in this first 2008 study by **Jeff Barrett et al.** and look at those same markers, look at those same markers in our much larger 2012 data set of 75,000 samples, every single one of these markers is confirmed and all of them are more significant than they were in the original study. So this process is extremely successful at actually finding true and replicated associations. So I wanted to draw some more general lessons I think from the success of GWAS to ways in which we might be able to approach improving reproducibility in science as a whole. The first is, and this has already been mentioned, is the importance of designing and performing very well powered studies, studies where we actually have sample sizes where we have some expectation that these actually give us the power to identify a variance with a realistic effect size and this requires doing formal power calculations up front. Secondly and to some extent this shows my bias here I believe where possible it's often advantageous to start with unbiased scans of data. These might be looking at all the genes in the genome or all the common variants in the genome or many different chemical entities and then pursue the results that stand out at the top of those distributions for further analysis. That comes across with extreme caveats which I'll add in the next slide. But this result is often much

more successful in finding true associations than entering a data set with very strong and well framed biological hypothesis. Another lesson is that it's important to leave very little room for post-doc analysis of data. We need to establish consensus on statistical approaches and very rigorous and well laid out protocols for performing analysis that gives no room for the researchers to redefine their analysis post hoc and find analysis that gives them significant results. In the GWAS field it's also critical that raw data, and the software that's used to generate those results are available to other researchers. And finally I think an underappreciated benefit of the GWAS era is that it pushed the sheer scale of the numbers that were required to find associations pushed us all to work in consortia and as a result of work by many different groups together first we had access to many more samples and secondly it was now no longer the case that a result that was pushed out to publication had only been seen by a single lab. Now there were many analysts who had looked at the raw data, there were many different labs that discussed that data in conference calls and that was an opportunity for internal peer review to identify systematic errors and misinterpretations before they reached publication. I mentioned the value of doing unbiased scans, I think it's worth emphasizing these unbiased scans raise a new set of statistical challenges that we face in the genomic era. When we don't start with hypotheses but rather sift results from complex and noisy data sets we face the challenge that it extracts outliers from such a noisy distribution and will also enrich for biases and artifacts. So understanding and removing these artifacts from data firstly requires a deep understanding of the classes of error that arise from any technology that we are using in that particular analysis and will often require manual spot checking and many different approaches to visualization and secondly this has already been flagged. The critical importance of validation using an independent technology. So we can't rely on the first results that emerge from any genome wide or large scale scan, we need to go forth and validate those approaches using a separate and orthogonal technology. I'll skip through this slide in the interests of time but this slide focuses on the need I think for a more nimble approach to discussing scientific results. And in particular moving towards approaches such as preprint archives and four that allow opportunities for post publication peer review and give us an opportunity for robust and real-time discussion of controversial scientific results. So I'll finish then with specific recommendations. So I'll finish then with specific recommendations,touching I hope, on some different recommendations than has already been mentioned here. The first is I think there are many fields that lack the strong statistical consensus that's emerged within the space of complex disease genetics, I think there certainly is room for standards workshops that bring together domain experts as well as statisticians to define consensus protocols that can be adopted for widely used techniques. The second is funding creation of software that makes it as easy as possible for people who are relatively new to large scale data sets to actually deeply understand the error modes that arise within those data sets to perform quality control and to understand when significant errors have emerged. The third point which has already been touched on is to mandate across the board the availability of the raw data that was required to generate those results as well as all the protocols and the software and code that's actually used to generate those data. Second to last, there's a huge importance to large scale collaborative scientific approaches to understanding large scientific problems and I think this comes with benefits both in terms of scale and also the internal peer review that I mentioned. And finally currently there's very little career incentive for scientists to engage in post publication discussion of papers in an open forum. I think anything we can do to provide career incentive for scientists who are actively engaged in discussion of controversial science will benefit the community as a whole. Thank you.

Eric Lander: Let me thank the whole panel. That was an awful lot crammed into about 30 minutes, which was what the plan was, and thank you very much for the careful preparation and wide range of subjects. Our pattern on PCAST is that people raise their flags to be acknowledged to talk and I'm looking around and I see Dr. Press has his flag up. Jim, I've got you. Bill.

Bill Press: So not coincidentally, the three panelists are all in biomedical-related fields and that's where I think this problem has shown itself as most acute. But it exists in other fields and other methodologies. So I wonder if the panelists could sort of go up a level and think about a couple of different examples. One example might be a field that is nowhere near establishing standard protocols and in which the use of statistics is largely exploratory. I was much influenced years ago by John Tukey, who was a very unconventional statistician, but who firmly believed that

it was possible for an investigator to be sufficiently self critical and self honest to be able to not specify protocols in advance. And I wonder, was he kidding himself or is there room for that. As a different question but also looking at different loci, what about computational experiment, where the issue is not reproducibility in a narrow sense that if I get the exact same code from you and I run it I'll get the exactly the same result, but the argument is about modeling errors in the code that can be quite deeply buried and quite sophisticated. So any thoughts on either of those issues?

Donald Berry: With respect to the last point, the famous Duke example that did involve fraud but the code that was used was then used by the NCI to try to duplicate things. And it turned out to be that every time they ran the code they get different answers. It was written in a very unstable sort of fashion. So even if you get the same code, you can get difference answers. But I disagree with Tukey. I think he's a great statistician and great scientist. But unless you -- in my old age I'm convinced that unless you write out what you know in advance, you're so -- when you see the data, you're taken by the data and you think this is real. The data are telling me it's real. But it's probably not. And you've got to put it into context. And in GWAS, you've got to put it into context or you're not going to get anywhere because of multiplicities and the huge data are going to kill you. So I think you have to specify in advance what you know and then, of course, do the things that Tukey would do to do the exploratory stuff. So he named the field exploratory data analysis, which is incredibly important but it's exactly that. It has to be confirmed and exploration is not sufficient.

Eric Lander: Other flags. While people are thinking, I will ask, there are a the range of different topics that have come up. There's true irreproducibility. You do the same experiment again. Maybe with the same reagents and the same cells. And you don't get the same answer. I'd love to use the word irreproducibility for that. There's non-generalizability. That's a very different sort of thing. The inhibitory RNA that you used worked just fine but the next three don't. That may be your ill-advised to use just that one or two but your result was reproducible it just isn't generalizable, or it worked in this cell line but doesn't work in five other model cell lines. That's a series issue, it deserves its own name. Very often I think some of the points being referred to is that the conclusion that's being drawn is not justified. So you think this is a specific antibody that does something, well, it doesn't it recognizes 20 other proteins. So it seems to me that a taxonomy of what the problems are could be really helpful because these are all important problems. And simply whether the results are right or wrong, whether the data are available, just in ancient days when we had a limited amount of space in the journal you might trim all those blots and now with the Web all the data could be there. So I wonder if you guys can help us think about for clear thinking on these various fronts we see an example, which is the best structured case, genes predicting risk of disease, where there's little risk that the causality runs the other way. The genes are there before the disease develops. In most other models it's going to be much messier as to which way causality is even running how various problems could screw up your model. How should we taxonomize this space to be most helpful in thinking about solutions? All these things deserve solutions I'm sure.

Glenn Begley: Can I comment? I agree with what you've said. Personally I don't have a problem with the result not being generalizable. My perspective is discovering drugs to help people. So it has to be a robust result upon which you can build. But if the paper says two I and look at another 20 that's okay I don't frankly have any problem with that. It's not robust enough to build a drug discovery program but there's nothing wrong with that, to use sort of a moral term. So I'm happy with that. I don't like titles of papers that are overreaching. So very common the investigators use a single cell line and the title of the paper says we've cured ovarian cancer. That deserves a separate name, too. That's scientific exaggeration to achieve publication or whatever it might be. But that's actually, there may be nothing wrong with that either in the sense that the data for the single cell line is correct but it doesn't help the world to exaggerate and blow that out of all proportion. It's certainly not good for patients because I've sat on the other side of the table on a Monday morning when on the set over the weekend these papers were published and the patients come in and they think their ovarian cancer is going to be cured. Patients deserve better. So I think what you've said is exactly right. And irreproducibility is different, it comes in many different flavors. And each of them I suspect, likely has a different solution. If I could choose one solution, I

think it's got to be education. Because I fundamentally believe that people want to do the right thing and every investigator with whom I've spoken wants to do, they want to contribute. They want to see cancer cured. And oftentimes they just don't know that the data selection that they're using is not legitimate. People have told me things that if they thought it was really wrong, they wouldn't have told me. That's not the case. I think training is very important in terms of helping the next generation understand what the issues are. And I agree we shouldn't lump them all together because there's a spectrum of irreproducibility.

Eric Lander: Other panelists here.

Donald Berry: In terms of taxonomy, I don't have a great handle on it. Some of what you described, Eric, is heterogeneity and understanding heterogeneity so the different cell lines you expect to get differences and we know in cancer, half of all cancer papers start out cancers is a heterogeneous disease then they assume it's homogeneous. (laughter) So I think it's critical to have that understanding of heterogeneity both with respect to within cells but also going from cells to people. In terms of taxonomy, I think the first use of the irreproducibility, i.e., reproducibility issue was a professor at Stanford who had post-docs and he had a post-doc this year that had to spend the first year reproducing the last year's post-doc's results. So writing down what did you do and how did you do it so that you could have, if not the code, at least the algorithm. So starting there, in terms of statistical reproducibility, I've learned through the last few years that when I go back and try to reproduce something I did. I can't do it. Why did I take the logarithm? What else did I do besides the logarithm? So having that sort of record, I think it's really critical and plays a role at least in some aspects of the taxonomy.

Eric Lander: Indeed, there's a movement called reproducible research for being able to embody in some persistent piece of software or environment, a complete description of what you've computed. Daniel.

Daniel MacArthur: I think we are getting close to reproducible research at least within the space of genomics and genetics, it's increasingly tractable there. I do wonder, though, the degree to which that's extensible to many other fields even biomedical research. So actually certainly when in a computational setting where all we have to do is lay out each piece of code in exactly the order in which that was applied to which data set it's feasible when it's a whole series of different cell lines that are stored in different freezers at different times, it becomes more challenging. Nonetheless we have to tackle what we can tackle and given my focus I think if we can push towards fully reproducible code that will go a long way towards solving some problems.

Donald Berry: In terms of that, can I. Let me address Dr. Press's question about outside biomedical stuff. It certainly happens. I gave an example in our exchanges before this PCAST meeting of two labs back in the 1990s that estimated the density of Charon, Pluto's moon, and they came up with greatly different answers but the standard errors were tiny in light of some of the things that Glenn was saying and completely different. And it was all based on their particular measurement process, which was obviously different. It turns out now today we think it's in the middle of those two so they were both wrong.

Eric Lander: I've got Craig Mundie and Ed Penhoet and Dan Schrag and I see Mario Molina is raising his flag.

Craig Mundie: I'd like to ask you to think a bit in the future as it relates to the how the presence of software, broadly defined, is going to contribute to more variability. People in science are often worried about the word of provenance of the experiment and the data but I would argue now that you have to pay equal attention to the provenance of every element of software in the system that you're using so that now that everything is based on micro processors those things all run software or firmware as it's often called. Each of these things represents something that can introduce variability from one to the next. And a lot of the scientists and also your own remarks tend to talk about the software I'll call it at the application level, which is the most direct manifestation of what you know into a codification in software. But you really are dependent upon this incredible array of software below that, the platforms the operating systems, the mathematical libraries, the actual chips themselves and how

they do computation. And as the data gets larger and larger, and these combinations grow, don't you think this is going to become more of an issue? And do you think that rather than just mandating the data and the app have to be made available, you actually have to hold the lab accountable for detailed collection and presentation of the provenance of everything in the system. And because so much of it is soft now, I think it's going to be much more challenging.

Daniel McArthur: This is definitely a manifestation of the generalizability point that Eric raised. It's certainly true we had nasty experiences in this area that the same pieces of software run on slightly different hardware architecture can give disturbingly different results. To some extent this comes back to developing software that is robust to hardware architecture and the challenge here is that this is typically the domain of professional software engineers whereas most software that's written in an academic setting is written by people, at least in my field, by people who come from a biomedical or biological background who entered into the software world without any formal training. I think what's required for this to happen, to really harden software against those type of problems is to professionalize software development to a much greater degree than we currently have that would require massive amounts of training, require making software fully available so others can take the same piece of code and generalize it and robustifies it and then we start moving in the right direction.

Craig Mundie: I agree with you completely. I also think it's more challenging insofar as we don't have a good way to create these cadres of professional people who are going to work in this area. And there's also a sentiment just because people can see the software they can determine its correctness. And I would argue that that's no longer the case, just due to the complexity and subtlety of the bugs.

Eric Lander: We're going to move along because we have a bunch of questions, we're actually about at the end of the period for this but we'll let this go on a little bit because the next session is on the same but we'll keep them concise. Ed.

Ed Penhoet: Two related questions. Your presentation, Dr. Begley, asks for some institutional responsibility for results. And I would like to have you explain a little bit further how you think that might work in an academic environment and it's related to this question about education in the field of statistics as applied, especially to molecular biology and related disciplines where education has been lacking in that field and where most universities departments of statistics and biostatistics are moving to the frontier not dealing with the sort of the more mundane problems of daily living and applying statistics to these areas. Do you have any thoughts about statistical services within the university environment and how you could accelerate the education process if we have to wait to train a whole new generation of scientist it could be a long experience to do that? And is there a model that could be useful to draw on, or Dan, in your case, where did you guys access your statistical power or your statistical algorithms that you used within the university system, outside, how did that work.

Glenn Begley: So if I can comment on the first part of your question. I'll leave the second to the others. The first part of question institutional responsibility. I think that the institutions have to my mind been tardy in beginning to address this problem. I think that there's a significant advantage to them were they to do so. They could save money first by the patents they filed many of which will not stand the test of time. That immediately would save them money. In addition, if an institution could put a stamp of approval on a particular patent, then I know that the venture capitalists and those that would be willing to take that forward in terms of additional discovery and turning it into a drug would be much more confident that that had been independently replicated. I immediately see value in terms of the institutions addressing this. I also think the institution should hold their staff accountable. The institutions are willing to bathe in the reflective glory that comes from a top investigator publishing in a journal like Nature and happy to have the interviews on television so on but when it turns out not to be nonreproduced the institution seldom stands up and says, okay, the person that was promoted to professor maybe we should think about taking them back a step or two. The institution to my mind has got significant advantages if they take responsibility in this area which I think they have. And finally education. The institution

should be committed to making sure that the post-docs and the Ph.D. students within an individual laboratory are being appropriately trained so that they can teach the next generation in terms of the appropriate application of the scientific method.

Eric Lander: All right. I've got next Dan Schrag.

Dan Schrag: Thank you. So I want to focus on, sort of get it to the big data question because what you're talking about, Dr. Begley, is really seems like it's just data analysis. I'm interested in this question of selection bias. There's robust ways of dealing with multiple stages of selection and updating your priors and changing your statistical tools. It's shocking that people don't know how to do that. I'm a geologist and in the earth sciences, there are some analogous problems like looking at 10,000 tree rings and picking the 100 that correlate most with what you're looking at and forgetting to update your priors on that. That's an analogy to what you guys do. But in general when you are looking at the history of the earth you can't do experiments. You've got to read the record and there's a question of quality assurance on your measurements, but that's a separate issue. In general, you can't do reproducibility. You have to look at what's there. And I actually think that's a better analog for what we're dealing with, with big data. Big data you're not doing, you can't reproduce the experiment. You're talking about massive data sets where you are automatically selecting subsets of those data, probably multiple stages. And it seems like there are challenges in making sure that those pathways of selection are carefully described so that the statistical tools you're using are appropriate and not inflating the significance of the results. I'd love to bring this discussion into the big data category, if you could think about that.

Donald Berry: So just a point about can't reproduce the big data. You can. In this statistical analysis sense, for example, it's well known that a standard approach is to split the data in half and look at this half and build a model and try to confirm it on the other half. So I think you can reproduce. The issue of the -- surprised at people who can't do that. And commenting on the education, the problem is really who are the educators. The senior scientists are the problem. They are not the solution. And it's a culture issue with the statistics in getting people so that they learn from this and the young statisticians are just as clueless as the senior scientists.

Eric Lander: And the senior scientists are the problem as I understand it because they're not native to any of this world of big data. They grew up in a very different world where the data is in their lab notebook and lack any instincts for --

Donald Berry: And the utility --

Eric Lander: They believe in reproducibility but good science is mostly about figuring out all the ways you're fooling yourself inadvertently not intentionally and usually you try to get those before you publish it. You don't, somebody else gets them for you. But good science is mostly just killing very attractive hypotheses and if you weren't brought up with big data. I'm just struck across biology that people who are not native to large amounts of data are extremely careful about the antibody but not necessarily careful about how many hypotheses they might have been testing.

Donald Berry: But Eric the utility is so different for senior scientists, they get the paper in Nature. That's wonderful if they have to do a correction they do another paper. (laughter) It's all in the utility structure.

Eric Lander: Maybe. I'm not sure that -- because I watch people be obsessive about the things that they know are problems. So even on papers that would go off to Nature and get a lot of attention, they can be utterly obsessive about the areas where they know there's pitfalls. Mario --

Mario Molina: I recognize I'm sort of part of the problem. But anyhow, my question or my worry has to do with the nature of the scientific method, you make a hypothesis and you test it and presumably you have experiments

that can tell you that your hypothesis is wrong. But to make the point. Let me use an example. It's Dan Schrag's similarity. And this is measuring the temperature increase of the planet. This has been very controversial. Lots of data, lots of statistics. And the conclusion presumably if you do the right kind of statistics that there's a slight temperature increase. But where science comes in, and that's my question to you, what part of that is missing, is to interpret it. First of all, there are other measurements, other than temperature, that support the hypothesis that the climate is warming because ice sheets are melting, all things are happening. It's not just a temperature measurement. But then perhaps more important you have this hypothesis. Why is this happening? Because the amount of CO2 in the atmosphere is increasing, it absorbs infrared radiation and so forth. So you have a whole scientific sort, how I put it, story behind the statistics. Okay. And then of course there's lots of people that try to disprove it, the consequences and so on. But you put the whole story together and there's no reasonable hypothesis that gives the same conclusions. So my question is can you go beyond just the statistics and understand perhaps in biosciences it's not always easy, but to tie it in with an understanding perhaps of the molecular chemistry perspective is this working or not. Does that matter or are you just looking at numbers or --

Donald Berry: No, you can't just look at numbers. In the world of big data, if you don't use the biology or the chemistry or whatever it is, you're lost. And so you have to build it in. But there are inevitably assumptions and modeling that can go into things and people can throw darts at you and say you made this assumption and I challenge that assumption. The big data issue just makes it critical for bringing scientists involved and learning as we're going.

I want to thank our three panelists here first.

[Applause].

Improving Scientific Reproducibility in an Age of International Competition and Big Data II: Editors

Eric Lander: Now we're going to turn to the next panel. We'll quickly switch over we hope you'll stay for that as well because clearly if there's a problem our next panelists are responsible for solving it, because they put their reputation -- if an institution matters, it's their institutions that matter the most because they're saying they stand behind papers they're publishing and they've been adequately reviewed. Not to put too fine a point on it there Marcia and Phil. But we have the distinguished editors in chief of perhaps the world's two leading general science journals, Science and Nature and we appreciate that both of them have made time to come here and talk to us about their perspective on this issue. Their journals have been sifting scientific claims and papers for a very long time. It's always an evolving process and we're curious to hear how evolution is going. Phil is joined by Véronique Kiermer, who is the Executive Editor and head or research services at the Nature Publishing Group. Thanks

Marcia McNutt: Good morning, thank you for having us here. Thank you Bill for prompting PCAST to look into this very important issue. Let's see I guess I have control here. Okay. So by way of my background, I am not a biologist and I'm not working in preclinical studies. I'm a geophysicist, which is an observational field, primarily not lab. I spent 15 years on the faculty at MIT. I was also president of a 40,000-member scientific society. And I spent 12 years at an oceanographic research institute leading that and four years heading the U.S. geological survey. So I have some views on the various roles of university agencies funding agencies scientific societies and journals as to how they can variously contribute to the solution to this very important problem we're facing today. And I'll comment on that towards the end of my remarks today. So there's a spectrum of reproducibility and there was a special issue of Science that was published in December of 2011 on data replication and reproducibility, which I recommend to you all. And it talked about at the low end repeatability which is exactly what Glenn was talking about this morning, and that is can another group simply access the data, analyze it using the same methodology and obtain the same result. And that's one thing. And that is sort of what one end of what we'd like to see happen if the experiment is carefully documented, and then Glenn was also talking about the high end, the gold

standard of replication, where the study is repeated, start to finish, including new data collection and analysis using fresh materials and fresh reagents and obtain the same result. And he also talked about cases where they even used the same lab with the same reagents but were not able to obtain the same results. That would be replication not quite the gold standard. So you can see how you could get from a replication that wouldn't be quite the gold standard, but still not be able to do it. So everything from repeatability to total replication and all of that is within the spectrum of reproducibility. Now, as I said, I'm a geo scientist, and geo scientists have a particular challenge in this spectrum of reproducibility. We're never dealing with the high end, the gold standard of replication. You take something like the Tohoku earthquake which happened recently. This was an earthquake that, of course, was not only very tragic, but an important learning experience to learn about tsunami generation processes and about major earth processes. We can't replicate that earthquake. At best we think something like it happened a thousand years ago. We're not going to wait another thousand years to have an earthquake like that happen again. At best what we can do is take the data that was recorded and hope that another group can get similar results in terms of the magnitude of the earthquake, the suspected slip on the fault, the moment magnitude release of the earthquake, et cetera, but we can't repeat the earthquake. We don't have multiple earths that we can look at that we can try to have a control earth that didn't have that earthquake, and the earth that did have that earthquake. So in the geosciences, replicability is always a challenge. What we can hope to do is take the same data with the same algorithms and get the same answer. Now, there are many reasons for lack of reproducibility and we heard about some of those, too. One can be that in publishing papers, information is simply withheld. And there can be reasons for that. One might be that there's not enough space in the paper. Now, journals like Science and Nature have worked very hard to remove that as a barrier. Even though we have limited space within the print journal we allow unlimited online supplements for that. But sometimes the author simply doesn't deem it important. That's where the reviewers and editors need to step in to make sure that the authors understand that it is important. Another point that Bob Shrock raised with me recently is that sometimes the authors have tacit knowledge that the practitioner doesn't even know he or she possesses, sort of like the master welder or the master artisan, where they do something a certain way because they've always done it that way, and they don't understand that it is critical to the outcome that they got. Of course we heard from Glenn earlier that sometimes even the scientists who originally got the result can't reproduce it. So that doesn't explain all of the irreproducibility. But sometimes the best thing that can happen is someone from another lab comes in and observes your lab protocol, and says, oh, you do it that way. We never did it that way. And sometimes you learn about variability and experimental protocol which can be very important to understanding differences in reproducibility. We also heard an example from Dan this morning about systems not sufficiently known. So that not all independent variables are controlled and that can be a reason for lack of reproducibility and then of course the well known case of false positives where unfortunately I think that because people want to produce papers that are provocative, that are interesting and have results that are exciting. If they get a result that reinforces their bias as to how they think the system actually behaves, and it's a false positive, are they going to repeat the experiment? Unlikely. They are probably going to rush to publication before someone gets in ahead of them. If it's a result that they don't believe doesn't reinforce their bias then they'll probably repeat the experiment and then they notice it's a false positive. And so I think this problem with false positives is always about getting a handle on your own biases. And not allowing, if you're a laboratory scientist, your biases to get in front of your rush to publication. Because that's when the false positive is likely to get in the way of reproducibility. So what is science doing about this? Well, I published an editorial not long ago that announced a set of new initiatives to increase reader and reviewer competence in studies published in science. First of all, we're adding additional members to our board of reviewing editors from the statistics community with the help of the American statistical association. And that will help us get at this issue of has the data analysis been done appropriately, particularly in cases where there are particularly gnarly problems with the data analysis so that the board members can help us find good reviewers and flag those papers that need special help in that area. We're also going to be for preclinical studies which are the cases that Glenn talked about which have been the topic of recent concern, we are going to be asking all authors to assure us whether they have followed the NINDS guidelines which were published in Nature not too long ago. And those guidelines basically ask whether the author had a pre-experimental plan for handling data, not changing the rules of the road on the fly. Whether there was a sample size estimation to assure

appropriate signal-to-noise, did you need 50 mice, did you need 200 mice, et cetera. Did you have randomization in your sample treatment and was there blind conduct in the experiment. So that's fine for the preclinical studies but we want to go beyond preclinical because we honestly don't believe that problems with reproducibility are confined simply to preclinical studies. So for all studies, when papers are accepted to Science, we're going to ask our reviewers and editors to flag the papers with unusually excellent treatment of data and samples to have the authors volunteer to write up their approach in its general terms as reasonable. Hopefully that will actually be in the paper itself so it won't be a burden on them. Collect a compendium of treatments across all fields of science that will provide input for NINDS style workshops later in 2014 in selected areas. And those areas will be decided upon based upon the input we get from the authors. It might be field studies, lab studies, modeling, et cetera. We can sort of decide once we see what these best practices are. So we can hopefully come up with some best practices so we can propagate these NINDS best practices to other fields other than just preclinical and maybe even improve on the NINDS best practices as well. Now, going quickly through the various roles of these different groups prestigious journals have some leverage in enforcing standards because scientists want to publish there. But traditional journals are failing more competition from new publishing models. Open access, preprint servers, not all of which had the same requirements on authors for reproducibility. Journals are likely to be the first to know when research they published is not reproducible so they have an obligation to alert the scientific community. Science has always had a policy of publishing technical comments whenever we're alerted to problems with our papers. And we are going to continue to do that, and one of the topics that I want to explore in the future is more online access to longer technical comments that would actually be negative findings that would be adjunct to papers to allow longer commentary on negative results from papers we published. Role of universities. They're responsible for training future and current researchers in the scientific method in best practices to produce reproducibility. They can reward researchers who produce reproducible results and withhold rewards from researchers who produce non-reproducible research. I actually haven't seen the latter happen. And I'm not sure I've seen maybe the former happens, too, but it seems like there's more emphasis on publishing in high impact journals than what actually happens afterwards with the results in that. So anyway. Role of funding agencies, make panels alert to criteria for reproducibility at proposal stage as it needs to be part of the experimental plan. And has budget implications. You know, 50 mice versus 200 mice. Consider whether reproducing key experiments is worth funding, and preferentially support PIs that produce reproducible research. That should be very important. And role of scientific society. Consider honoring those who consistently produce reproducible research, devote special sessions at scientific meetings to the topic of best practices in reproducibility and adopt reproducibility guidelines for society publications like triple AS (AAAS) is doing right now with its publications and this needs to be a team effort very much. That's it.

Eric Lander: Marcia thank you very much. Phil.

Philip Campbell: Thank you very much for the invitation to talk about this very important topic. My predecessor, John Maddox, was once asked by journalists, I think how much of Nature is wrong. And he said all of it. You might just want to think about that as a starting point for this discussion. It's not a recipe for complacency but we know the papers are contingent. We know that journalism contains nuances on the truth so on. Sp we're very alert to the fact that what we do does not necessarily represent objective reality but an approximation to that. Ever since I've been associated with Nature, which started as long ago as 1979 we've done the same thing when selecting papers. We've simply tried to find those papers that we the editorial staff find the most interesting. That hasn't changed. What's changed over the last 30 years or so you guys, by which I mean the scientific community, have given us power that we don't necessarily want. And the reason I emphasize that is the big part of this problem that we're talking about is the incentive structure in science. Glenn mentioned that. And I think to look at the details of what we do is important. Because there's no sense in which we are complacent about our role in not helping the scientific community enough but I do want to make that general point. That the culture of science, the culture of science funding and the culture of institutions needs the attention of a body like PCAST just as much as the journals do. So here is one of the triggers of the problem in our minds the growth in formal corrections in our pages, note that there are retractions idling at the bottom but it's the corrections that showed to us that we had a

problem in sloppy science as we saw it. And that plus a conversation I had with Glenn Begley in 2012 after he had published his first article in Nature with the Ellis, gave to me a sense that we really have to look at this seriously. I want to acknowledge also Véronique Kiermer who is with me who won't be giving a presentation but she's a biologist and I'm not. And a lot of this is biology, and she has played a very important role in our implementation of the policies. So I thought it would be good for her to be here in case that was helpful. So the growth in formal corrections, I'm not going to go through this list because actually we've heard a lot of the problems already discussed but you can be sure that our corrections reflect the problems you've heard about. I also want to make the point that although we place a lot of emphasis on mandates for data deposition, etc. and on reporting standards, the fact that you do that isn't sufficient. And this was just an example published by John Ioannidis who did a huge amount of good work, I think, in highlighting the problems of the sort we're talking about in Nature Genetics, which just showed how unrepeatable, even though the data was apparently mandated it was following the mandate in terms of deposition. In fact, when you got into the details it was problematic. So we have over the last one and a half years or two years taken some steps. So we did quite a lot of awareness raising, we participated in meetings the NDIS meeting. I called a meeting and given talks at the Academy of Medical Sciences, the Royal Society. There's a meeting coming up of the collection of science research councils in Europe, Science Europe where they'll be discussing this issue. We've published awareness raising articles you'll see the list in a minute I won't dwell on it. More importantly I think, we removed length limits on the online methods sections completely and we have found, incidentally, that the length has grown by something like 30% since we did that. That was this April of last year. We substantially increased the figure limits in Nature, there really was a problem as Marcia has referred to already and we improved the access to supplementary information. We agreed with a company called Fig Share that is co-owned by our parent company to present data behind the figures as much as we can. And we are considering developing the author contribution statement, which I think can point the finger if I can put it impolitely of the responsibilities of the individual authors more than it does at the moment. We've appointed a statistical advisor and hired statistical referees and we published this editorial we did in 23rd of April, the checklist for authors, editors and referees just to show you quickly these are the articles we published that reflect the importance we gave to the various dimensions to this problem. I would draw your attention to the last one, which is one of the few that actually is cautious about this whole concern about replication. It's referring to the project that has been set up to deliberately replicate papers published in high impact journals and calls attention to problems when you set out to do that. That, I think, is a salutary thing to read. That's the editorial. You can read it for yourselves if you haven't. Those are the editorials that point out at the same time all of our relevant journals. And here's a checklist which I'm not going to go through but the figure legends statistics and general methods are dealt with, reagents, animal models, human subjects data deposition. Is this a complete answer from us to the problems to which we are contributing in the reproducibility question? Certainly not. We'll review the impacts of these measures we've taken later this year in spring. And exactly how we're going to do that we're still carefully thinking about because we do want it to be independently minded that review and also trying to look at some further measures, some of which have been referred to by the previous panel that we should be taking. We found the implementation of the reporting checklist to be onerous. The authors, referees, copy editors we've all had to spend time on this, both developing it but actually implementing it. We're not sure yet to what extent referees are paying a lot of attention to the list. That will come out in a survey we hope to do. Authors are submitting some papers with a checklist with our prompt many have embraced the source data opportunity we give them. And we have found an improvement in reporting but at the moment I say we haven't systematically studied it. For example, one of the measures we introduced in relation to animal experiments was researchers had to report what they hadn't done as well as what they had done. And so you can see in Nature Neuroscience in particular, reports of randomizations, blindings predetermination of sample size that's not been done to increase transparency. That I think is what we can claim to have achieved. We have begun to increase the transparency in terms of what the researchers did and didn't do. This is an example how we're making source data behind the slide available. Using making Excel data available so other people can go immediately and replicate what they're seeing at least at the surface level of the paper. This is just showing that we're now taking more trouble to show the uncropped image of blots. This is an immunoblot, but it's just one example, we haven't the survey how the practice is changing as a whole since we brought in an insistence as much as possible we do show these images.

There is attention needed on the question of cell line identity. This is a particular initiative that's out there by the community to try and require authors to validate the cell line that they're using. We are in the middle of -- we've certainly recommended that, and advocated that as something that authors should do but are not yet ready to mandate it because we're not convinced how practical that is, but as long as we're declaring the fact that that hasn't happened then for now we're considering that to be adequate. The author contribution statements informal and unstructured. They're non-templated. Could we change that, could we go for more detail, that's just a question we're looking at. We're working with organizations like the Welcome Trust and looking at that question. So now I come towards the end of the presentation, which is what are the underlying issues. So all the ones that are not in bold are things you've heard about. You've also heard about the ones that are in black. But those are the ones I think a lot much attention needs to be given. I'm delighted that Francis Collins and his deputy published in Nature this week in the latest issue the NIH response to this problem. With that there's no question that looking at what funders choose to fund and for them to be sure that what they're funding is a realistic deliverable, as it were, on the problem which can often be extremely fuzzy and unconstrained. That attention is crucial. We published the endpoint of the process, but the publication bias and where you can publish refutations are issues which are we do need to look at and we are looking at. The issue for IP confidentiality that afflicted if you like Glenn's study where he wasn't able to report the papers that he actually looked at in his initial paper. That's something that we can't do much about, but it is out there. And lab supervision I was tempted initially when all this started to relate this to lab size. There are certain laboratories out there that put a limit on how big or rather certain institutions which put a limit of something like ten on the labs they will support because they believe that once up get to a size beyond that it becomes very difficult to give adequate supervision and validation of the data that are coming out. But what I'm not saying is that this is a problem of lab size. But that can contribute to the problem. Lab training, there's no question if you go to some prestigious institutes there are people not adequately trained for some of these problems. You've heard all about the pressure to publish, to what extent we at Nature and at Science and other journals are responsible for that pressure is open to discussion. But there is no question that it is contributing to this problem. And the worst way in which it contributes is summarized in the line that was just said to me by a scientist at a meeting but seems to be believed by many people now when you heard someone referred to it in the past panel that you get your paper in Nature and it may be wrong but hey you've got your tenure or your grant by the time it's shown wrong. That's paid for you to be sloppy because we don't have what it takes to detect that. The only thing I can say we've taken some steps to try and clean up what we publish and we welcome any further critical scrutiny of what we've done. That's just a slide about the replications challenge. These are the issues that I see facing universities and institutes: data validation, lab size and management training, publication bias, access to the data in the institution and access to the reagents. Thank you very much.

Eric Lander: Great. Thank you very much. Complicated set of issues, what a varied set of issues. I should note, though, with regard to the last point I'm not going to name the journal but at least 25 years ago it was broadly said about a journal which was not either of your journals. "Just because it was published in X doesn't necessarily mean it's wrong." This was a broadly, and it was one of the most distinguished journals in the field and what it meant was the journal really specialized on pretty out-there, exciting, interesting things and therefore many of those things that were really interesting turned out to be wrong because there were some artifact of some sort over interpretation and at the cutting edge one often selects for those along with often exciting things that are right. It's not the first time we've had the discussion although I think it's different for a variety of ways you guys have highlighted. I think it's important it comes back and again. I'm going to ask two very quick questions and then turn it over to my colleagues. There's a tension I'm picking up between, on the one hand this online world of getting things out quickly and letting the community decide and higher and stronger refereeing standards to demand all of these things that we want to see in the paper. Are these intentioned with each other or not, the idea that referees can be gatekeepers for two years over a set of very unreasonable demands and they are blocking the scientific community from hearing one person's opinion or one person's view and it would be better to get it out there or it would be much better to have much greater gold standard gate keeping. That's my first question. My second quick question is you referred to gold stars for really good papers. I'm just curious whether it would also be useful maybe with the permission of the authors to have postmortems of randomly selected papers as to what their

deficiencies are with regard to these things. It's good to put up those good models but to the extent we recognize we all want to do better it might be good to see some less good examples randomly chosen as well. Anyway, that's how I see those questions.

Philip Campbell: I'll take the first one. Although there's the tension between speed and, if you like, validation, All I can say is that we track the speeds as a measure of performance so we do want to provide a good author service, but we will despite claims or rather despite protestations about from outside we'll go back again and again if necessary in order to get the paper to a level we want it to be. So we put the primacy on getting that paper as valid as it can be.

Eric Lander: You shouldn't and we should get them out quickly. But you would not buy in to that?

For Nature, we're conservative about the standards of what we're trying to publish despite everything that you've seen.

Eric Lander: Very good. Marcia.

Marcia McNutt: Yes, on the second issue, I do think that we probably will have in these workshops some, the postmortems on some examples of some papers perhaps some retractions, not retractions but some that had technical comments that came in on them especially ones that attracted several technical comments that had problems with them. We're seeing them and it's not just the preclinical area, we're also seeing many of our papers in the social sciences that have had problems particularly with the statistical analysis.

Eric Lander: Could I urge you not to do just the really good ones and the retractions but the random sampling because I bet a random sampling will turn out not to have a lot of the data available, not to have the figure fully shown, where we have no reason to think it's right or wrong in an unbiased way we can ask is everything we want there.

Marcia McNutt: Well, Eric, we could do a random sampling but let me tell you my experience with random sampling. Back when I was at MIT I taught a course in spectral analysis and time series. And the way I taught the class was the students were responsible to read the book on their own because we had a really good textbook. And during the class we all came into a lab that had workstations, and I would give them a randomly chosen paper from a top journal that had a data set in it, that was illustrated a topic from whatever we had been reading about. Maybe it was co-linearity in data or statistics of small sample sets or something like that. And the students would take that data set and they would analyze it using the principles in the chapter. And as I say the papers were randomly chosen. There wasn't a single paper in which we could reproduce the results in the paper.

Eric Lander: Thank you. Bill Press.

Bill Press: On the subject of technical corrections or feedback to the community. I think it's very interesting as we've evolved to, through the Internet to a sort of a world scientific conversation, at a level of formality that exists underneath the more formal level of published papers. And Marcia knows where I'm going because we've corresponded on this. I have a lot of sympathy for Glenn Begley's point that if you flip through a random issue of Science or Nature and, as this point has just been illustrated, you'll see a lot of small things wrong or missing or impossibly small error bars or "N not stated". And it doesn't rise to a technical correction which is authored and goes to the author and you're going to affect your scientific reputation or in my case I had to decide whether it was a good thing for my students to write a technical correction. I decided it was not. Why don't we have informal comments the way everything else on the Web has probably moderated by perhaps volunteer moderators, but something where an individual can just suggest, would it have been good to know what N equals in this paper or that kind of a thing.

Philip Campbell: When Plus One started out they made a big thing of the idea that that was going to be a feature of that journal. At the same time, roughly we put comments as a facility on our papers and in both cases comments didn't really take off at all. And that's to do with incentivization, as to why would you do this, why would you put your head on the block, as it were, you're not going to get the paper that might not come if you do formal correction, et cetera. That's my interpretation of why that hasn't happened. One thing we can do and we are doing more of now is referring to discussions going on elsewhere in the social media, for example, where people feel freer to discuss those sorts of points. I'm not saying that's an easy thing to do but actually I do think it's one way forward. because there's a lot of, as you rightly say, discussion out there.

Eric Lander: Jim Gates.

James Gates: Thank you Eric and I'd like to thank both sets of briefers for their very erudite presentation. This has certainly been noted as the century of biology, as a decade of biology. As far we can see the promise for what the biological scientists will do for people's health, for the quality of life is, the potential is enormous. And this problem that you've discussed has mostly been within the confines of the biological sciences, but, of course, we have examples, for example, the cold fusion of physics two decades or so ago, or similar things in other parts of science that shake the faith of the public. And in fact this discussion to me has been sort of within our tent of scientists, but we also have an outward looking face and a responsibility to interact with the public in a way to confirm their support for the scientific endeavor because at the end of the day we all rely on that support of the public. And so as you folks have discussed this and certainly even in the way that Bill, I don't know if you shaped the title of this session, but it says science, as if it's a problem we see across all fields of science at this magnitude. My question to the panelists is there, in your opinion, is there a way we can discuss this so that we are responsible within our community but which also allows some of the nuance to be seen outside of our community as we put our public face out?

Marcia McNutt: So let me see if I understand where your question is going here. How do we -- how do we phrase this as more than just the life sciences?

James Gates: That's one way to interpret it.

Marcia McNutt: Because I think there certainly are examples of cases where there has been irreproducibility that has been more than just for life sciences. And whereas we've certainly seen previous decades as where there have been huge advances in the life sciences. I think in the future we're going to see many problems with water, with environment, with climate change. And we've seen many examples of groups that have questioned the reproducibility of climate change records and have wanted to know more about whether projections for climate change models can be reproduced, whether some of the energy projections for fossil fuel reserves and things like that can be reproduced. So there are a number of examples. One that I know very well was an earthquake prediction algorithm that was, came out of Cal Tech decades ago that was based on a model that was using travel time differences between P and S waves that was actually creating the anomaly the way the code was written. And so it was finally noted that there was actually nothing happening in the earth but that the code itself was creating the anomaly, and so the earthquake was actually never going to happen. But there's all sorts of examples from other sciences as well. So I think that this is an issue that it's much broader than just the life sciences, and it's something that will impact all aspects of our quality of life. And safety.

Philip Campbell: Just a quick response. Two points, I think. One is I've seen it work in a sense, that we published a climate paper, a skeptical data analyst who runs a website that actually systematically tries to criticize the literature did find errors in the data. This was about temperature in Antarctica. And the authors corrected the data. So that sort of public engagement, when it comes to very challenged science, actually can work. And I mean we'd be happy -- I will just put in a plug for gold open access here whatever my publishers might say I am going to

plug it for climate change because if ever there's a discipline where you want this stuff out there, that's one of those disciplines. So that's one point. And the other point I would say is as an avenue for public engagement in the area we've talked about today, patient groups I think can be a very interesting avenue for that. You know, what do they think of these problems, to what extent are they taking this Antarctica literature in a naive way and how dangerous is that?

Eric Lander: I'll just chime in favor of Bill's point. Discussion of papers, post publication. It might not have worked. But the richest exchanges are these really great collegial post publication exchanges. I didn't understand what you did there, could you elaborate. What is the end. And if we haven't got quite the way to make that work yet, where it's a considered -- phase four after the phase three trial there's post marketing surveillance or something, it's okay. It's part of the whole thing. If we found a way to say that the discussion, once the paper comes out is really important and it's not challenging necessarily, it's not undermining one's integrity, it's collegial discussion. And if at the end of it you have a collegial discussion section you expect that for a year a record of collegial discussion would attach to a paper. Maybe that could be an interesting thing. I just urge you to experiment and try ways to draw us out because smart young students who are tweeting about things, about papers, they know a lot. They see a lot. I read some of the blogs in genetics and I'm very impressed that papers appear, I won't mention, but even sometimes in your journals when they appear and within three days the young bloggers know the paper is wrong and they can tell you it's wrong and you guys can correct it eight months later.

Marcia McNutt: Yeah, Eric, the only thing I will say about that is there are some journals that have had open blogs for discussion and then closed them down, and the issue often is just a financial one because you have to have a dedicated staff person monitoring them because, believe it or not, there are a lot of kooks and crazies out there. (laughter)

Eric Lander: No. (laughter) The NIH might be interested in supporting some experiments now based upon this recent paper of Francis and his colleagues. I know we've come to the end of the time because we do have a break that's scheduled here for our next. Is there any other burning question. Does anybody on panel one have a burning question for anyone on panel two? No. Okay, I want to thank the panelists on the first and second panels. This is a big complex question. We've seen all the little pieces of it, it's a diverse set of problems but it's great to see the energy that's being devoted in the best scientific tradition to self-criticism of the fields. It's great in this first panel that people have taken time to actually write papers about this, think about good models. It is a thankless task being journal editors, sifting through all this stuff. We can barely come close to thanking them for what they have to do and what they are going to continue to have to do. But were you not doing it the scientific literature would not be as tremendously powerful and marginally reliable as it is. So we're going to thank you although it is very small compared to the thanks you probably deserve for it. But it's all you get. So thanks much. [Applause].

Challenges for the 21st Century Enterprise: Leveraging S&T Across the Department of Commerce

John Holdren: Okay, welcome back from that break. We're now are privileged to have a session on challenges for 21st Century enterprise, leveraging science and technology across the Department of Commerce. And to talk about that we have Pat Gallagher, the Director of the National Institute of Standards and Technology, which sits in the Department of Commerce, and he is also Under Secretary of Commerce for Standards and Technology. And we have Larry Strickling the Assistant Secretary for Communications and Information in the Department of Commerce, and Administrator of the National Telecommunications and Information Administration, NTIA. Their biographical details are in the materials that the PCAST members have. So I won't take time to recite their many distinctions, but gentlemen we appreciate very much your being with us to talk about this big topic and the floor is yours.

Patrick Gallagher: Thank you John. It's great to be back. Good to see everybody again. I know originally Secretary Pritzker was hoping to be here today to talk to you about this. My goal is to wear my third hat, John, which is the Acting Deputy Secretary hat and channel Penny (Pritzker) a little bit for the group and talk about the department's agenda and the overlap with science and technology and you're going to see a big data and information thread running through this. And then I'll elaborate a little bit on some on the NIST side and let Larry talk about what's happening as well in NTIA and I think that might be the best fit for the agenda you've had today in this session. So secretary Pritzker has been moving very fast as the new Secretary coming in at this point in the administration and reframing the priorities of the department. And remember the Department of Commerce is quite broad with activities from fisheries and earth observations at NOAA to standards of measurement science at NIST to trade and investment issues at ITA, enforcement, a whole set of issues. And the result of a very extensive listening session with both with staff as she came on board and talking with the bureaus but also going across the country and meeting, I think over 400 CEOs across the country, and really listening resulted in us actually starting over and rebuilding a strategic framework for the department that we have put together and have been not just putting a pretty report together but we've actually been baking this in to the way the bureaus operate. So this has been a very different kind of exercise focused on real action plans, real priorities we've been using it to drive budget discussions for '14, 15 and now '16. And it's a very different kind of activity than we've done before. The framework is what I call the four plus one framework it basically organizes our priorities into four programmatic, thematic areas and then one that's the verb, it's how we execute, it's called operational excellence. The four programmatic areas are trade and investment, innovation, environment and data. And what I thought I would do today is give you a sense of how we're thinking about this but I in particular wanted to basically draw two distinctions and I'm going to use the data angle as one to focus on because it's quite interesting. It's the one that's getting, I think, some of the most excitement from outside government in terms of reaction that we're hearing. But the distinction I want to draw for you is the following. The role of the department, when you look at the mission, we are using our science and data in two distinct ways. One is to reinvent how we operate within the department. And the other one is basically the department as the enabler business community in the United States. In other words, enabling a data-driven economy. And I want to sort of separate those. They playoff of each other but for the sake of today I thought I would actually separate them and talk a little bit about those two activities. So let me start with the second one first. One of the strengths of the Commerce Department is an ability to engage and partner with business to drive change, to act as a convener, to act as a facilitator and to partner. And there's a number of vehicles in which we do that. We could talk about the NIST MEP program, we can talk about things that are happening in the manufacturing area where we're providing seed funding for industry to come together and form consortia, the Semitek type models we can talk about and Larry will probably talk about things we're doing to work in the spectrum area to support interdisciplinary work on spectrum measurement and testing. But one area in particular is in cybersecurity. PCAST has written a report on cybersecurity. NIST was given a specific responsibility under an executive order that was issued last February to develop a set of practices for the critical infrastructure of the United States that, if put into practice, would result in improved cybersecurity performance. The way we did that was not to have NIST go off and create some body of work. It was actually to have NIST act as a convener and have industry create a set of practices that it could use itself. And so we turned the sort of traditional tables around and used an approach that you might remember we used in smart grid standard settings and others which embraces the fact that in the United States, when we tend to do these things, and this is somewhat unique of this country, the government isn't the standard setter. Under the National Technology Transfer and Advancement Act, in fact, the government is specifically told it should seek first to use private sector led standards and only develop its own as a last resort. And so we really embrace this model, and began with a request for information. We went out to the public. We held workshops across the United States. And I think two things have happened. One is the framework that has evolved is quite different than the one that probably would have evolved if it had been written by the government. It's one that is designed to align or maximize the alignment between business operations within a company and cybersecurity. This is, after all, a risk management activity. And so it seeks to basically bake in that risk management approach. And this should sound very familiar to PCAST since this was the essence of your recommendation. And that has been crucial, not only because we think it's going to result in something that has a higher probability of working as it's

not some static list of controls. It's an approach that seeks to engage a company. It also adopts a maturity model thinking. So that companies starting out may look like a set of controls that it generates and that's how you begin, but a high mature organization is basically going to be proactive. It's going to be adopting to new information, it's going to have the approach to do that. But the other reason this has been so important, though, is that when the executive order was released, the debate we were having in the country was whether to regulate or not. And by having companies sort of embrace the management of risk themselves, in other words if you think of regulations as addressing a market failure, what we're trying to do is do everything we can to make the market work. And that has been enormously important in getting very broad buy-in. What we're hearing from companies is that they are already doing the due diligence and internal process to adopt and use the framework and in a couple of weeks when this comes out I think I hope that this is, has a sense of inevitability and that this framework process becomes a dynamic and ongoing process. So that's an example of sort of the Commerce Department acting in this very different role of convening and supporting business-led activities in response to a particular need. And the elements of that approach are to put the public need in front of the business community, because it's often a collective need that's there. And that was certainly true in cybersecurity. This is a global technology infrastructure that ties us together. It was important to have somebody lay out the common goals and to provide a venue where this collection of private sector entities could work, contribute best practices and put something together. And I don't believe we've overused that approach yet. And I think that will be emblematic of a lot of other things that we do. Before turning this over to Larry and then stopping so we can have a discussion with you and hear any questions you may have, I also wanted to talk about how we were thinking about data within the Commerce Department. One of the interesting things is that, you know, folks coming into commerce, and I'm sure Penny Pritzker heard this when she was going through her confirmation was that it's a holding company of all these eclectic things. And it certainly feels that way to a lot of folks. But the truth of the matter is, there's something that is very central to everything that happens in Commerce, and that is that every single one of the missions of all of these bureaus is fundamentally enabled by data and information. The predominant thing that the Commerce Department is doing is providing technical assistance to the public and to businesses through its activities. And that means that it has a rather robust mandate. Its core competency, if you will, is going to be in areas of data collection. Look at what NOAA does in the area of earth observations. Look at what the Census Department does in terms of collecting statistical information from the United States. It's a data collection and analytics engine. And on the other side we build services that deliver, that analyze, consume and adjust that information and provide targeted services to parts of the business community. So if you think of the Economic Development Agency, yes, it provides financial assistance, but what it's really been doing over the last five to 10 years is providing through a network to the state economic development agencies information on what makes regional economic development strategies work. What is it about a hub or cluster that makes the hub or cluster work so that when you make that investment through a grant on an infrastructure project it's in the context of something that makes sense. That's true of NIST in terms of providing technical information on how to do a measure, how to support adoption of standards, and that's even true of our Minority Business Development Organization which is providing minority business owners with the targeted information they need to advantage them and to help them succeed. So we have sort of, for the first time, embraced the idea that rather than a federated approach to data the one thing we're learning is the data has to be done at scale, the real power of data comes from integrating it and doing it in a compelling way. So for the first time we're actually having, under our data priority, a rethink about the data infrastructure that the department itself runs. Instead of census, for example, being a collection of data collection on a particular survey and analytics and producing a report, and by the way that's how census is built today. Each one of these things is thought of as a separate activity. What if you create the data analytics and collection engine and then build the surveys on top of that. The same thing with NOAA, you really have a global observation data system there that's measuring the earth's atmosphere and oceans, let's focus on developing that at scale infrastructure and in this case and hybridize it with private sector because it's the real activity in NOAA now, this is happening globally, and then look at the things that can be enabled when you do this at scale. And provide the potential opportunity for machine learning and advanced data analytics and the type of scope. So we've been thinking about it and where do you create the organizational home for this core competency to be developed in the department. And we're quite excited about that, and I think it's an area where we will want to

work closely with PCAST in terms of thinking about how that is enabled. With that let me turn it over to Larry to let him tell you about some of the exciting things happening on Internet and data.

Lawrence Strickling: Thanks you Pat and thanks to the members of PCAST for the opportunity to visit with you all again. I think not all of you may be familiar with NTIA and our mission. We have four program areas. I don't intend to talk about all of them today but just as a point of reference, our four program areas are Internet policy, spectrum, broadband specifically, the management of the Recovery Act Broadband Grants Program, and then last a new edition to our portfolio has been the development and deployment of a National Public Safety Broadband Network, which is called First Net, which is operated under our auspices. But the two areas I wanted to touch on in these beginning comments are spectrum and more importantly privacy is an aspect of the Internet policy work we're doing and again in anticipation of your upcoming work in the big data area. So on spectrum, in fact I think the last time I was with this group was at the rollout of the PCAST report at the White House. So, what, Mark, about a year and a half ago, I think. And again, thanks to the committee and all of you who participated on it because that report has provided a tremendous boost to changing the paradigm of how we go about allocating spectrum for commercial use. And quite honestly, and this was recognized right from the first paragraph of your report, we're in a situation where the traditional clearing of spectrum to make it available on an exclusive use basis to commercial users, we just can't do that much longer. And what's happened since then, just to give you a real quick update and be happy to talk more about it in the question period is that the FCC is now on its way to auction off later this year 40 megahertz of spectrum, which will have sharing arrangements in it, largely sharing with satellite earth stations. But even as a result of your work and additional analytics that we were able to do in an unprecedented industry/agency cooperative process last year we came up with a new way to think about how the commercial sector can co-exist with satellite earth stations. Traditionally these huge exclusion zones were created around the earth stations with no commercial activity allowed within them. But we have now developed the tools to allow much more commercial use within what were formally exclusion zones and we're now calling them coordination zones to allow commercial activity, but with more cooperation and a set of rules in terms of what happens to the extent interference might occur. Additionally, the report had focused on the 3.5-gigahertz band, the FCC has launched a proceeding in terms of how to look at making that band available for unlicensed use, again, sharing with the federal use, which is largely, large radar systems, many of them naval radar systems. And I think there's a tremendous opportunity for some experimentation in that band which the FCC is already pursuing. I think pursuant to your recommendation, we are in the midst of preparing and issuing a plan to evaluate another 1,000 megahertz of spectrum that we think can be made available for sharing with commercial users. Just last year, Pat and I announced that we were going to take our assets in Boulder Colorado in the spectrum area and come up with a new organization, the Center for Advanced Communications, which will have as its focus doing the necessary measurements testing, research on spectrum sharing methodologies as a way to provide a place for both industry and government agencies to come to be able to get expert neutral input and output in terms of how we continue to find ways to share the spectrum. And all this is going to lead into a work effort on exploring exactly how to go about implementing the suggestion of the committee to create one or more, we call them test cities, but Mark told me beforehand they're called model cities. We can talk about it more in the Q&A. But what I'd like to turn to now is the privacy area. And just to give you a frame of where the administration, and in particular the Department of Commerce, had been on consumer data privacy issues prior to the President's call a couple of weeks ago to take a much deeper exploration of the implications of big data collection and use. The starting point for all this, and I brought several copies, which will be available first-come, first-served was the release in February of 2012 of the President's blueprint for consumer data privacy, and that's detailed in this report here and it's available online if I don't have enough hard copies for those who are interested in it. And this really lays out a framework based on three key approaches to privacy. The first was the establishment of the Consumer Privacy Bill of Rights. So there's seven fairly general principles laid out in the document, modeled on the fair information practice principles that have been with us since the 1970s. But in an effort to update them to be more appropriate for the more digital age, those are laid out in the document. Then as a departure from the more traditional then let's go out and spend years to write regulations to define all these terms and impose straitjacket rules and regulations on people. As an alternative to that what the blueprint lays out is a process of having industry working

with other stakeholders including consumer groups in civil society work to develop what we call the multi-stakeholder codes of conduct, to actually implement in specific industry context and on specific issues the rules of engagement for how companies can deal with consumers on a particular issue. We started and have concluded the first effort to develop a code of conduct in that manner. This was on mobile application transparency. We started the process and invited anybody to it who wanted to come. And we attracted 400 people to the first meeting. Over time, the group winnowed down to roughly 50 to 70 people who are very dedicated, both from industry, civil society, academics who wanted to participate in the process, and we successfully concluded a code last year. That code is now being tested by the two primary trade associations of mobile app developers. Individual companies are also looking at exactly how they would go about implementing it. So it's too soon to call it a great success, but it certainly, I think, has advanced thinking in terms of how do you go about on that small screen giving people enough information so they understand how their information is going to be used as an alternative to the multi-page privacy notices that people used to get or traditionally have gotten when they attempt to use an online service on their computer. Next week we are going to launch the second of these processes. And this effort is going to focus on facial recognition technology. So the first group or the first group meeting will be I think February the 6th next week. All these meetings are streamed online. So as any of you who are interested in taking a look at this I certainly encourage you to have a look at that. We will continue to look for other opportunities to develop these codes. What we see in the advantages of this are the ability to bring all interested parties together, reaching consensus, taking advantage of specific technological issues at a particular point in time. But more importantly having the flexibility to be able to modify these codes on an ongoing basis to the extent that technology changes. Our concern has always been in this particular space that attempting to regulate a code by the time you've completed it, even if it was perfectly appropriate to address the problem you started out to solve, that problem is now probably 2 to 3 years old, and it's been totally redefined. So we believe we've demonstrated that this is a superior approach to doing it. One of the things we'd like to see is, obviously, an ability to start doing these faster and being able to take on more of these at any particular point in time that will come as people get experience with this. Frankly, this was a new way of doing business for the consumer groups, for civil society, for companies and we went through the first several months of having everybody looking at us expecting us to give them the answer. And I think everyone was a little frustrated by our refusal to do so. And frankly it took a lot of discipline not to be injecting our view of how it ought to come out, but if these processes are going to succeed you really have to force the group to find its own way to reach consensus without interference from the governmental authority in the room. This is a matter that we're very interested in now going back and reassessing how well what we've done for consumer data privacy, is it applicable to the new world of big data which goes beyond the kind of the paradigm of a company collecting information about you to then do targeted marketing back to you or to have third-parties come back and do targeted marketing. So we're very interested in, and want to continue to work with you all as you sort through the implications of big data in terms of whether we ought to go back and recalibrate what we've been doing in the consumer area to collect and be able to respond to some of these new trends that are emerging today. So, again, I'd be happy to talk more about that in the Q&A but I think at that point maybe we should stop and go to questions from you. Thank you.

John Holdren: Well thank you very much, Larry and Pat. I'm going to take the prerogative of the Chair to make the first comment/question. And that is, I'm struck by the a possible parallel between the Commerce Department's role in cybersecurity and its potential role in climate security. You've made the point that the private sector is increasingly at risk from breaches of cybersecurity and what the Commerce Department has been doing in that domain includes engaging with business to get them to better link their business operations to their cyber risk management. And it seems to me there is great potential, given that the private sector is also increasingly at risk from climate change and from lack of preparedness and resilience for the impacts of climate change to engage the private sector in better linking their business operations to climate risk management. And in fact, PCAST has been talking a bit about what mechanisms we have to link President Obama's Climate Action Plan more effectively to activities and actions in the private sector. the Interagency Council on Climate Preparedness and Resilience, that is co-chaired by OSTP, CEQ, and the National Security Council staff has also been thinking about this question. It just seems to me that the Commerce Department has a big potential role, given both, obviously, the fact that NOAA

resides there, but also the fact that you have long been in the business of getting the private sector to think about questions like linking their business operations more effectively to constellations of risks they face. And I'm just wondering to what extent you've been thinking about that or will be willing to think about it.

Patrick Gallagher: So I think that's very much the focal point we have as well. That it's great to have earth observation and weather and climate prediction capacity but the real public benefit of that knowhow is translating that into usable information by the business and by the public. And there's a lot of examples. Whether that's information going into improving building codes and standards so that the built infrastructure is more resilient, whether that's information, I mean, there's fascinating work going on in taking that kind of information, translating it into decisions that, for example, farmers are making with regard to crops and some interesting aspects where that's actually tied to the crop insurance markets. And so we agree and so the focus, if you were to talk to Kathy Sullivan, is to have sort of a, there's really two aspects to this. One is creating the data capacity that can support this and the other is creating the types of targeted agile services on top that really provide that value. I will tell you the one area, though, where I think you can make a big contribution, and this is a big data problem. The magic of making that happen is going to come when the data that we collect can be mixed. So the real power, for example, is taking the climate and weather data that we collect and also looking at the business data that we have on the census side. And to the extent that policy and structural approaches keep those separate, which is the current state of affairs today, we basically deny ourselves the ability to use that. So the real magic comes when you can put it all together. Now you've got the problem we're facing. Once it's all together how do you ensure that it's used for the purposes for which it's taken. So you're going to have to put the right kind of policy controls on the usage of this very powerful combined data in a way that this is something that I think everyone's struggling with, and what we're hoping as we struggle through this even within the Commerce Department, which is in our management span of control, that we can tackle this. And that's a place where I think this group will have a lot to say. It's the analog of exactly everything else that everyone is struggling with here which is how do you bring a pervasively sampled world and combine that data and yet still have the integrity of process. But John we agree to you that's where the value is.

John Holdren: Good. Thanks. We have now a hair over 15 minutes for the rest of this session. We have four or five flags up so I just note that the questions and answers combined have to average about three minutes each. Jim Gates is the first flag I saw.

James Gates: Thank you John. Thank you to both Pat and Larry for the briefing. Yesterday we had some folks from the Bureau of Standards, Kristina Bartsch and James Franklin give us a brief on some things that relate to statistics that BLS collect. We in PCAST have relied on BLS in a number of our reports. And one of the things that I think that was clear from yesterday's presentation is maybe a missed opportunity and in particular Pat I would ask you to think about this. You mentioned NOAA in your comments. And we all have seen the synergy that arisen between NOAA and outside and creating things like opportunities for the Weather Channel and Weather Bug, I mean there's a whole ecology around that data set. So during the discussion yesterday, the opportunity seemed to somehow present itself that something like that with BLS data might be something that you would think about and what would your view be about that, because we think there are impediments at least some of us think there are impediments, and I would just like to hear your thoughts.

Patrick Gallagher: So I think that's the holy grail. I mean the value of this data is enormous. You're talking about data that's actionable by business. Whether it's employment data or it's trade data or whether it's economic activity data, whether it's climate and weather data. And yes, we wholeheartedly agree that you've got to create the service layer that this can take part. The challenge is going to be struggling through what is the right way to provide that value that businesses can use. Remember, in the case of business data, this is a lot is geo-located business data that has profound privacy implications on it. And that's why the BLS authorities and census authorities and others have been constructed the way they are. It even makes it very difficult for them to share information with themselves across government to government, federal to federal. And so the real issue is if

you're going to create these high value data platforms we've got to find a way where we understand the rules in which that can be released and used and built upon, but we actually agree. And that's the real power, the real big opportunities are where you're combining different types, taking CDC data with weather data. Taking this is where some of the real power comes from in terms of the public benefit of that information.

John Holdren: Good. Maxine Savitz is next.

Maxine Savitz: Thank you for both coming out. Following up on the big data issue. You have very disparate groups, as you say. How are you both involving the private sector and sort of developing this infrastructure and then are there going to be barriers to implementing it, because the government says you can't do this kind of procurement or that kind of procurement, how are you going to overcome it so you really have a successful system and census will go on and not be postponed and NOAA will still be able to get the data.

Patrick Gallagher: So the short answer is our approach is basically to have two things happening simultaneously. One is we have to figure out the actual vehicle where you can create a common data infrastructure within Commerce, right. So we're looking at the approach that we can use so that we can actually create a common data infrastructure within our agency. Commerce has an interesting angle in that we have the National Technical Information Service within, most of you may not even know that exists but it was created in the '60s, and its specific mandate is to be the repository for the government's technical and business information. Now, at that time it was to be publications and paper. But in fact from an authority perspective it may be the right vehicle to create a shared data infrastructure, the data layer. And then I think the problem solving unfolds in two ways. One is the private sector is way ahead on how to do this so what we'll be doing is reaching out to the leaders in the field and they will be part of what does this look like how does this work. On the other hand we'll face problems in combining data that are unique to government. So what we're also setting up is the team that can sort of tackle those problems as well. So if you're looking at Commerce today it's a lot of business outreach to a list of companies that you would widely recognize and the other side of this is a group within government that's looking, that includes attorneys and others, that can help us sort of manage that. But that's why I mentioned this technology enabling is going to be very important that we have the capacity.

John Holdren: Pat, I just have to mention that some of us in the room are old enough to remember relying completely the NTIS for our access to federal data and documents in the pre-Internet pre-World Wide Web era. Next is Bill Press.

Bill Press: Pat, Larry, thanks for doing this. My question is, in this intersection of big data with privacy that both you and we have been asked to think about. But it's also, I'm struck by the experience base you're developing with this convening power with bringing the right people into the room and, as Larry said, not dictating, which is hard or harder than dictating. But one of the properties of big data is that it kind of interchanges the rows and columns. Let me say what I mean by that. You may have in the room as you say the stakeholders in mobile or in Larry's second example the stakeholders in I would generalize it to video processing or something, and these are the sort of collection modalities and it's relatively easy to say who are the stakeholders who are producing the products in those areas, but when we start correlating big data we're sort of cutting across those and we're looking at things like identity across all platforms or location of an individual or association, who they're associating with or remote sensing of their health. And I'm a little bit lost as to how I would apply your process if I need to go across the rows, who would I get in the room who could speak any common language across these, all of these stovepipes?

Lawrence Stickling: So in our case we don't make those choices. Our processes are open to anybody. And then you're dependent on how well have you publicized what you're doing and been able to frame it in a way that you attract the people that have an interest in the particular topic. If the specific example you give where maybe you want to take a broad topic and look at it across six or seven different actual products, that might be something we might not take on at the current level of maturity of our process, because it might just overwhelm one's ability to

get enough focus on what you're trying to deal with. I think eventually one could, as you build up more experience and more of a book of precedent in this area. But I think we have found this is an area as we do this it's so new to people that we've pretty much had to crawl before we can walk, and we're going to have to walk before we can run. So I think how you select the topic and how you publicize it matters in terms of one's ability to get the group of people in the room that can talk about it and reach some consensus about it.

John Holdren: Good. Next is Susan Graham.

Susan Graham: One of the things that's going on right now is the emergence of sensor networks and the so-called Internet of things. And I'm wondering how you're doing with that and partly how you deal with new technologies in a reasonably rapid way but that's also a source of big data issues, it's a source of potentially issues for standards. It's got its privacy concerns. Where are you in regards to that topic.

Patrick Gallagher: We can both probably give quick answers this one. You're exactly right. And one of the things that we were doing at NIST in this space is we were recognizing that the cyber physical approaches were being done in industry either within company or within sector approaches. And they have to because the speed at which they're moving and deploying. And the issue, though, is as these new technologies unfold, even before we figure out how we are going to manage it, they have to be enabled to protect information. And the question is do we reinvent the wheel on how that will happen in each of these sectors or not. And what we've done already is convened cross cutting groups of companies that are operating in cyber physical. So we've been working with OSTP to basically try to form a horizontal consortia of companies where you pull in auto and others and ask them the question is there a common stack, if you will. Is there a common approach that can embed the capability and then allow you to develop the sector specific functionality on top of that. The reaction from industry, by the way, has been very positive and things are moving. But that's kind of a classic approach that we've been taking.

Lawrence Strickling: I think we're at the front end of this in terms of understanding the privacy implications, that's the first challenge, which is how is privacy implicated in this. And then beyond that what are the expectations of consumers as it relates to that? One of the things you'll see with regard to the bill of rights as it exists today I think it's based, I think, fundamentally on a transactional or contractual approach, the idea that you're going to, in effect, give up some of the access to your information in return for a benefit but that people can articulate what that is. So you go on Facebook and you get the benefits of being part of that social network. When you start getting into the Internet of things that might apply to some aspects but take an area where your utility basically says we're going to install in everybody's home something to monitor furnace usage or your refrigerator or things like that so to the consumer they may not see an obvious benefit to that. I think there may be a question right on the verge of this which is what consumer rights are there in terms of being able to say no to that if they find that to be an intrusion. So that needs to be defined. But then as that information is collected in the aggregate that could be tremendously helpful to that company and to society in terms of getting better understanding how to regulate energy usage but at the same time if you haven't thought about how the data can be matched up to particular consumers it could be an opportunity for somebody, a mischief maker to be able to tell when you're out of town because you've turned down your air conditioning or turned down your furnace and that might be information that a consumer might be uncomfortable knowing that a utility could identify that about their behavior. So there's a whole host of issues that we need to address at this but we're really on the doorstep and the Federal Trade Commission, I know, is thinking long and hard about these matters as well.

John Holdren: Great. Well, we seem to have exhausted the flags if not the subject. We're close to on schedule. We turn now to the public comment session, but first let me thank once again Pat Gallagher and Larry Strickling for their presentations and interaction. We really appreciate you being here. [Applause].

Public Comment

John Holdren: Let me now call on our one of our two vice chairs Maxine Savitz who will moderate the public comment section today.

Maxine Savitz: We have two people who are going to present public comment. If the first one would come to the front who is Bethany Drehman, policy analyst at Federation of American Society For Experimental Biology. Again we always welcome public comment and we also have copies of your written activity and you have two minutes and I will give you a 30-second warning.

Bethany Drehman: Thank you for giving us this opportunity to provide comments. I'm seeking on behalf of the Federation of American societies for experimental biology or FASEB 26 members of societies that collective represent over 115 biological and biomedical researchers. FASEB supports status and increased status sharing contributes directly to scientific research and have the potential to increase reproducibility. In commentary on reproducibility published this week in nature, the NIH director stated that many of these failures have simple and practical explanations different animal strains, different lab environments and subtle changes in protocol. Although data science and data sharing could help identify many sources of experimental variability this would require collection of large quantities of metadata for data sets and even individual data points, which is often impractical or impossible. FASEB has described the need to collect and report and share experimental data and metadata. They also recognize that there are many barriers to effective status sharing. The majority of effort and cost associated with development of data sets are borne by individual investigators, with the subsequent users being the primary beneficiaries. Standardized citations of data sets the recognition of effort involved in the development of shared data sets and application reviews and continued investment towards the development of data repositories.

John Holdren: You have 30 seconds.

Bethany Drehman: Will provide greater incentives for researchers to share their data in useful formats the lack of harmonized regulations and guidance regarding data deposition and fragmented IT infrastructures are additional challenges that remain to be addressed. We encourage the federal government to support efforts that would facilitate data and metadata collection and sharing and looks forward to opportunities to work with PCAST and OSTP on these important issues. Thank you.

John Holdren: Thank you very much.

Maxine Savitz: Our next speaker I apologize if I mispronounce your name Mahantesh Hiremath**,** who is currently in ASME congressional fellow at science based technology. Speaking as a private citizen and we appreciate all the work you're doing on the hill.

Mahantesh Hiremath: Dr. Holdren and members of the council, I wish to thank you for giving me the opportunity to talk to you the topic of the comment is who is doing the innovation. President Obama in his State of the Union address described how this moment of the late '50s unleashed a veil of innovation that created new industries and millions of new jobs. Innovation he was referring to is not a simple action, rather than a total process. It starts with an idea. It's followed by a discovery that describes scientific observation. Next comes the invention which is engineering way of using the scientific discovery and finally it becomes innovation by manufacturing in large quantities for a wide market. Innovation is incomplete without mass manufacturing and it is irrelevant without a market. Today the idea and the discovery phases are considered domains of the national labs. Their inventions are supposed to happen in the universities and the universities and the transfer technology office mechanism is supposed to lead to innovation. Let's look at the reality. The scientists at the national labs are mandated to concentrate on basic research, acquisition of knowledge for the sake of knowledge. Those who venture into

applied world are frowned upon. It doesn't get better at the universities. Young aspiring professors seeking tenure.

Maxine Savitz: 30 seconds.

Mahantesh Hiremath: Required to do fundamental research. The papers are considered relevant to one basic research so who is doing innovation? No one. As a result, much of the research is sitting in the brilliant minds of the scientists and on the shelves of the lab. There is a basic disconnect. I urge the PCAST to eliminate this disconnect. Those problems secretary [indiscernible] said it right that the national lab should do more business with significant teams focused on extended time on an important problem. Those problems can be identified by close interaction between the lab scientists and the industry and would lead to many more innovations. Thank you.

Maxine Savitz: Thank you very much. I turn it over to John.

John Holdren: Well, that concludes our agenda for this meeting of President Obama's PCAST. As always, I want to thank the PCAST members for their attendance, attention and participation. The OSTP staff for their support of this operation. I want to thank as well the members of the wider community who joined us here, including in the overflow room and including watching these proceedings on the Web. We really appreciate the interest of the community and the activities of this group. And that seems only to be growing for which we are grateful. Eric, do you have any closing comment?

Eric Lander: Just to echo your thanks and we look forward to our next meeting which will be in April.

John Holdren: Thank you all.