GENOMIC DATA SHARING-OBJECTIVES, MECHANISMS AND POLICIES

DR. ROYAL: Okay. Thank you, Steve.

Well, first I must say thanks to Sarah and Cathy, who are the ones who have really been working hard on this but we're going to—

(Slide.)

I'm just going to give a very brief overview because we have a wonderful lineup of speakers this afternoon. And in starting out I just want to remind us about why we are even talking about this and sharing of genomic data is important for advancing the agenda of science but the sharing of this data has the potential to have all kinds of ethical implications that are associated with it.

(Slide.)

Another issue that is raised is the potential blurring. We are not saying that this blurring is being caused by genomic data sharing. Certainly the blurring of the line between research and clinical practice has been happening for a while and the question is whether genomic data sharing is going to

increase this even more.

And then the questions about informed consent: Are we going to need to think about new approaches to informed consent as we move ahead with widespread genomic data sharing?

So what have we been doing so far?

(Slide.)

In December of 2008 it was decided by this group that genomic data sharing was an area of priority. In March of ‘09 there were briefings on the IOM report and from some other advisory committees. In September of last year, the Lewin Group got a contract to draft a report on genomic data sharing and to work along with SACGHS to do that. The project is a yearlong project that the Lewin Group is working on.

(Slide.)

And in our meeting in October we discussed this and we formed a steering group and some volunteered, some were volunteered, and the group, as you see here, and we met and talked about what we were going to be doing today. And in our meeting in October we did decide that it might be great to have a session

at this meeting and on our conference call we sort of fine-tuned that session in terms of what shape it was going to take.

(Slide.)

The Lewin group has actually started working on the project and they have been doing the background work and done some lit searches. And the questions that this report is going to explore are whether there are new issues regarding privacy and discrimination that we need to address, issues about consent, how can the process be improved? What are the benefits and risks of population based registries and how can researchers and policymakers address the issues related to indigenous groups and what we tend to call sometimes special populations who participate in this research?

(Slide.)

So what are we going to do today? We’re going to have a group of speakers who are going to really talk to us about the models of genomic data sharing. We are going to gather some information from them and we’re going to figure out--the next thing we're going to do is talk about the information that

they give us, what issues are raised, and try to think about where we might go with this.

(Slide.)

The presentations: We're going to start out with an overview by Laura Rodriquez from the Genome Institute and she's going to give us an overview of federal policies on genomic data sharing; then Joyce Mitchell is going to talk about future directions in terms of health information technology; and then we have speakers who are going to talk from different sectors about governmental, healthcare system, academic, commercial and consumer controlled policies for genomic data sharing.

(Slide.)

As we listen to those talks, the things we want to think about are what—as we listen to the models that are present, think about what are the implications for informed consent and what are the things that are common and what are the things that are different in terms of consent, in terms of the storage of data, issues regarding access and secondary uses of data, privacy, confidentiality, protection in terms of re-

identification and de-identification of genomic data. How do we handle sensitive data and then the incorporation into electronic health records?

(Slide.)

And at the end of the talk we are going to have a discussion and the key things we're going to focus on in that discussion are what are the elements that have worked well in these models that have been presented and what are the ones that haven't, and are there issues that could benefit from more policy discussion and development?

And then we'll try to think about what a next step should be? Will there be a need for us after we hear all this information? Will there be a need to identify best practices? Could SACGHS contribute to this? Or should we wait until the Lewin Group's report is done? The report already is generating some information about what is happening out there in terms of the literature that they are gathering. Should we wait until the report is complete before we decide what we should do? Or in the interim should we just plan some additional individual sessions to try to explore

this a little bit more?

(Slide.)

So we're going to ahead and move right into the discussions because, as Steve said, we are going to run a tight ship this afternoon.

So, Laura, please come and give us a talk about the federal policy.

REVIEW OF FEDERAL ACTIVITIES

RELATED TO GENOMIC DATA SHARING

DR. RODRIGUEZ: Okay.

Well, I would like to thank the committee for having me come and speak today on behalf of all of the ex officious and then just clarify pretty quickly here that I'm just reporting on what is going on from all of the different ex officious. I'm not an expert on many of the things that I'm going to talk about today so I'm very happy to see them all sitting around the table so that they can answer questions that you might have as we go through this information.

(Slide.)

Charmaine has already gone over some of the information in terms of why we're having this

conversation this afternoon and the goals from the committee but just to reiterate some of the rationale that I understood from the committee's discussion in terms of why they wanted to look at genomic data sharing was, in fact, the potential for this kind of data sharing to facilitate very important research and things that were going forward at the moment in a very rapid way and in a way that raised questions that we wanted to be very deliberate about in how we handle them going—as we moved.

Additionally, too, something that Steve mentioned already, is the fact that these kinds of data are blurring the line between research findings and clinical care, and so that's something that we also wanted to think carefully about. And, in doing so, the number of different ethical questions that are raised by not only the potential applications of these data but also how we manage them and what are different protections we have put into place for the individuals whose data we are looking at that is generated in large volumes around some different and new areas are the kinds of data that we are gathering.

And also again something that Charmaine mentioned is the fact that genomic data by its nature is challenging the traditional paradigm of what was de-identified or autonomous and how is that going to change the way that we needed to think about managing the data in terms of providing appropriate balance between wanting the research to go forward and also maintaining and protecting the interest of the participants from whom these data are derived.

(Slide.)

Again so we're—in terms of what I—what I understand you all wanted to do in hearing from all of the different feds and what we were doing in this area was largely because of the amount of money clearly that the federal government is putting into this in terms of the national investment in genomics research and, in fact, in building resources for the data sharing going forward.

And in doing this, the government is not only playing a role in the research as a funder but also is providing some leadership to the community in thinking about how to go into these new domains and, hopefully,

after today you'll have a little bit better sense of what kinds of things that the different agencies are thinking about as they are doing this.

(Slide.)

So the survey that was put together by SACGHS staff included ten questions that focused around the issues listed here trying to find out what research programs, if any, existed within the various ex officio agencies. And if they did have research programs or did not, did they see genomic data sharing as relating to their agency mission in any way, and how so?

And, again, assuming that they had programs for genomics research and some expectations of data sharing, how had they developed policies to try and implement these expectations and how did they incorporate elements concerning the different ethics questions into those policies.

And then, finally, again going back to the concept that these data are blurring a line between research and clinical information, was there any allowance within the policies or expectations within the policies to provide interconnectivity between the

research data and electronic health records?

(Slide.)

So this survey was sent to all of the ex officios, as well as separately to USDA and NSF since we knew that they included some genomic information, just to find out what kinds of policies, again, they were putting forward.

(Slide.)

We had responses from 12 of the ex officious, plus NSF. And as you will see, we had a range of feedback in terms of how this related. So there were four of the groups that had no genomic data sharing activities and they did not see it being relevant to their mission at all in terms of the purpose of the agency. And I think from looking at who these different groups are, it is not surprising in some regards how they were doing it.

DR. : (Not at microphone.)

(Laughter.)

DR. RODRIGUEZ: I am just reporting. I am not going to take—

DR. : (Not at microphone.)

DR. RODRIGUEZ: Then there were also several others that reported that they didn't have any genomic data sharing activities and weren't conducting any research in this area but they did see it as being relevant to their agency mission in some way.

(Slide.)

And, again, this is OHRP and this is in a general way. As we'll talk about later, they have some policies that overlap in the realm of general research protections and then the specific considerations around how genomic data sharing is done; OCR with their responsibilities for implementing HIPAA and also involvement in GINA, et cetera. We can see where these come from.

And then there were five other of the respondents who did have genomic data activities—data sharing activities and research programs and they did see it as directly relevant to their mission and, not surprisingly, these were the ones that were more research based and would be—as the dominant activity for what they did.

(Slide.)

Looking at those five even, they are still very different across the board. Of course, NSF and DOE, not surprisingly, largely deal with plant genomic activities. So that wasn't something that was particularly close to what this committee was thinking about but, even just looking at the VA, CDC and NIH, there were still very different states in terms of their thinking and activities in this regard and how they were approaching it.

So the VA at this point is still--they have n expectation for data sharing within their research programs but as they are an intramurally based research program, the expectation for sharing is within their own system.

The CDC has several different programs included in their genomics portfolio and the policies for data sharing vary among those programs but they tend to be based on how they have set up their traditional sharing and how they have set up their traditional sharing structures and how they have interpreted looking at the genomic data and other systems that they have. Their sharing tends to work

through direct collaborations or through coming to particular CDC research sites and doing research at those sites.

And then, of course, the NIH has invested significant time and energy in building database repositories and in having the broad sharing take place in a way that is much more indirect and through central resources.

(Slide.)

So I have tried to provide throughout the rest of these slides links. Actually I—Symma has provided links to many of these programs.

(Slide.)

Going forward, and just to highlight again one that CDC--the N-HANES program is a large cohort study that has had multiple different visits. They are—in the last few years one of the—they did a genomic data collection and they have thought a lot about how to move forward in the area of genomic data sharing. They've hosted several meetings and workshops to think about that going forward what would be appropriate within the particular structures for N-HANES, whether

there is actually legislative language that structures how they can move forward and how they can share their data.

The VA has a large genomic medicine program that they have been moving forward again within their intramural program and they have been very proactive in thinking about this. They formed an advisory committee in 2006, which has met on a regular basis to think about the different questions regarding how to appropriately share genomic data and how that relates to the particular participant population that they would have at the VA, including commissioning a study to look at participant attitudes specifically from the veteran’s community and asking how those veterans felt about different aspects of data sharing. Again, what would they want to get out of the data sharing in terms of return of results and other things.

So, the policies themselves at the VA are still under development but they have been very active in thinking through the issues and imposing the questions and serving as a forum, too, for discussion in a broader way than just their own agency.

(Slide.)

NIH, as is listed here, has been very active in putting forth policies. We have multiple different policies at the NIH level. One which I'll talk about later this afternoon focused specifically around genome-wide association studies but we also have specific policies for other projects that involve sequence data. We have roadmap projects for the microbiome, for instance, that have a genomics program. All of them have their own policies for what the expectations are for sharing of genomic data.

(Slide.)

At the IC level for the different institutes we have some more policies and I haven't listed them all. So we have been very prolific at putting together different ideas of how they apply to our specific programs and what the expectations are for genomic data sharing. Ideally these are all consistent from one to the other. We have tried to work very hard at doing that but we, of course, are still working on that.

(Slide.)

Coming back again to the agencies I mentioned

where they have policies or areas that touched on their mission that did not directly involve genomic data sharing, if we look at OHRP, the policies that they mentioned, not surprisingly, have to do with coded specimens, again, which pertains directly as to how the genomic data sharing is conducted in terms of is it human subjects research or is it not considered human subjects research and what are the regulatory implications then of that kind of determination, engagement in human subjects research. And so clearly these are very relevant to what is going on in the research programs themselves.

(Slide.)

Several other—the ex officio agencies, the EEOC, and OCR, in addition to OHRP noted that they have overlap in this area with regard to GINA and as there are regulations for GINA developed and are implemented.

And then, lastly, I really just wanted to mention that NIH is now going forward and extending our existing policies to putting together a trans-NIH policy for sequence data and related genomic data, such as epigenomic data going forward. And we see this as

extending what we have done in the past for GWAS and all of these different individual project by project policy development activities but, hopefully, will provide a way that will be consistent for investigators, institutions, those investigators that want to use the data, as well as those that are submitting the data and, of course, the public so there is a common expectation of what the NIH is doing and being a steward of all of these data that come into our resources.

(Slide.)

The themes that came forward, I think, through looking at all of the information from the different agencies that responded are clearly that genomic data coming from many different individuals will include sensitive information. I don’t think there was really any question about that from anyone.

And, also, that broad sharing of the data does enable an acceleration, the potential for acceleration of scientific research.

So with those two principles accepted, then the consequences are that policies that are developed

must ensure privacy and confidentiality of research subjects, and this was mentioned even by those agencies that said they had no activities and no relation to their mission. They were still concerned about the ethics of this kind of activity going forward. That protection mechanisms were needed against unauthorized access and that there needed to be careful attention to the distribution and use of genomic data.

And, of course, that the LC issues regarding their management, their distribution and their collection needed to be very carefully considered, and that they must remain relevant and timely to the technologies that are being used as well as to the public conversation that is taking place around this kind of information and how we are using it within research and within society more broadly.

(Slide.)

Potential gaps that were identified through the survey: Informed consent was the most frequently mentioned place where more guidance was needed or best practices were needed.

Again, not surprising, I do not think but it

was again something that was mentioned even by those groups that didn't have any activity in the area so something that is really permeating the discussions.

Additional consideration around what access participants may have to the data itself, either through the databases or to results, their own results, from their participation, and return of results is something again that's a very hotly contested issue with strong opinions on both sides of it in the community at the moment.

And, again, there was a recognition that while policies at the moment don't preclude any incorporation of this kind of data into electronic health records as those initiatives go forward, there really aren't at the moment any clear structures that will make the inclusion of the data into EHRs something that is easy to see how it will happen or feasible and so more attention to that area was another place identified as a gap.

(Slide.)

And with that I'll just, I think, come back to some of the same questions that Charmaine put up

again in terms of questions for you all to consider. So whether there is a need for additional policies for genomic data sharing in this area and, if so, from a federal perspective, does the committee have thoughts about whether or not an agency specific initiative or the way it should go or if they should be coordinated in some way.

And, also, is there a need to try and deliberately raise public awareness around the importance of sharing genomic data and the inclusion of these kinds of the data in their electronic health record in a targeted way?

With that, I will, again, thank Symma, Cathy and Sarah for all of the work that they did to put the survey together and to pull these slides together as well, and take any questions.

Yes?

COMMITTEE DISCUSSION

DR. WILLIAMS: So this is a question for you but probably more broadly to Charmaine in terms of the task force. Was--how much time was spent looking at not so much issues of privacy and confidentiality

related to sharing but to actual physical aspects of sharing data like use of standards across different organizations. Is that in scope, out of scope of the task force? Is that something that you addressed in your surveys to the various groups?

DR. RODRIGUEZ: So I can say it came up minimally in the survey and again the questions weren't structured to draw it out but I think only one or so of the answers that I saw come back in mentioned the standards for that kind of thing and with regard to the scope of the task force that is definitely a question for Charmaine.

DR. ROYAL: Marc, I think we sort of put that under the HIT umbrella and so that wasn't part of our—yes.

DR. RODRIGUEZ: Any other questions?

Okay.

DR. DALE: I'll raise a question then.

In the data sharing area there is the—you talked about mostly data sharing within the government agencies but there is--when the NIH or a government agency sponsors a study then it is really governed by

the IRB and usually governed by the local IRB. My experience has been there is huge differences between the IRBs and how they look at this issue. So we have this morass of different feelings, let's call them, about data sharing.

No one has corralled the wild horses in a way and it is very confusing if you are a researcher in this area. You spend a huge amount of time trying to share data.

DR. RODRIGUEZ: So I will agree with all of those statements. I am not sure what to do in terms of solving them. I think as we'll talk about, at least I will talk about more in my talk later about the GWAS policy that NIH developed, while the decisions for whether or not data sharing is appropriate still resides with the local institution. NIH has tried to put forward an infrastructure for some consistent protections to be in place and mechanisms to be in place for how the data are shared and what the considerations are in making decisions about sharing data to try and bring a little bit more ease, I guess, to the process of doing it both for investigators,

again trying to access the data, and those submitting the data.

And, ideally we have tried to provide some help to the community in thinking about these issues but obviously this is moving very quickly and we are not anywhere near a consensus on how to do it.

DR. : Can I ask along those same lines, are there standards that either cross the various agencies—standards or guidelines that each of the institutions, at least NIH or NCI grantees, have to use in any of these key issues, let’s say informed consent?

DR. RODRIGUEZ: So across NIH with the GWAS policy, that is a trans-NIH policy so there is a consistent threshold that is supposed to be used. We are a very large agency and sort of interpreting that policy, of course, is always somewhat subjective. So we have done a lot of work since the policy came out in terms of trying to develop rubrics and SOPs and other informational material for our staff to try to bring them up to a consistent level but that is taking time as we are all learning to go about this and, you know,

everyone has different ways of doing things. So, ideally, there is consistency within NIH but I am sure it is not perfect.

And in terms of other agencies, you know, I think every agency again is trying to do this on their own and we're talking to each other, and there are some general consistencies and principles but I think how each of us are deciding to do it is still evolving.

DR. KHOURY: So a clarification and a comment. AHRP does not have any activity on genomic data sharing but we do think it's relevant as is any patient specific clinical information is to our agency.

Going to the point that David had raised, this issue has been discussed in many different settings in terms of coordinating the different IRBs, which is a bigger problem not just in genomics but whenever we have any consortium of research in different centers, and there are different policies, how do we coordinate this.

So there has been some talk about coming up with new policies for multiple IRB or a blanket IRB or a minimum threshold but I do not know if that has led

to any conclusion or activity within the NIH. As I said, we're discussing with AHRQ of how to approach this area.

DR. RODRIGUEZ: And we are discussing it, too, but, no, I wouldn't say that we are at a conclusive point.

Okay.

DR. ROYAL: Thank you, Laura.

Now we'll have Joyce Mitchell, who is going to talk about the future for HIT.

FUTURE DIRECTIONS IN HEALTH INFORMATION TECHNOLOGY

DR. MITCHELL: Thank you.

I am assuming that you will get it so it shows up here?

(Slide.)

Oh, there it is. It wasn't there a second ago.

I'm delighted to be here to talk to you about existing and emerging technologies affecting the genomic data sharing. It is a very large topic to cover in a short period of time. And the tactic that I have taken is to be more broad in terms of general

areas and some trends that I see that are emerging.

(Slide.)

This group clearly knows a lot about genomics but I felt like it was useful to give a very broad brush overview. There has been huge progress made in the last decade or two decades certainly. And the broad overview of the whole thing is here at the bottom. There are 5,000 genomes available online, incredible information available online, and public data repositories are routine, and that is different than the situation would have been even 20 years ago or 10 years ago.

(Slide.)

There are lots of genetic tests which are available today to anybody in the physician community who wishes to order them. There are almost 1,900 SNP chips that are routine GWAS studies, expanding gene expression studies are impacting clinical care today, and next generation sequencing has arrived and has taken all of our small-scale single gene experiments into some things which are enormous as we try to do with the average variant trial for a human sequence

from 3 to 4.5 million SNPs and if you had insertions and deletions you end up with 10 percent larger than that. So it is a fairly large data problem.

DR. : (Not at microphone.)

DR. MITCHELL: Yes.

DR. : G2P is what?

DR. MITCHELL: G2P is genotype to phenotype. Sorry for the jargon.

DR. : (Not at microphone.)

DR. MITCHELL: Genotype to phenotype.

(Slide.)

So at the same time that all of that information is out there and available and expanding, then consumer demand for genetics is exploding. And I take you first to the genetics home reference. This is a site that I have a particular—it's particularly dear to my heart. I was a senior scientific advisor on this site from 2001 to 2009. It was the first site which actually targeted the public and said that the public would like to know how to bridge their consumer health questions with the bioinformatics data coming out of the genome experiments. And we started out—when people

were saying, you know, how are you going to do that.

(Slide.)

Here is our website, which you could go certainly explore at your leisure. It sits at the National Library of Medicine as part of the Natinoal Institutes of Health and currently it has about 500 health conditions and about 700 curated gene summaries, and another 1,800 automated gene summaries. And what I would like to point out is it has 215 million hits per year. It is never advertised because they can't advertise it, and that's 215 million hits per year from the public and from clinicians who go there a lot when the public hits them with questions about diseases and disorders that they don't deal with on a routine basis.

(Slide.)

And then, of course, the interesting phenomenon of direct to consumer genetic testing—I know it's controversial and I do understand the issues behind that but it is a huge force and it is there and it is happening daily. It's changing the pace and the standards for data exchange in genomic medicine and doing it in some interesting ways.

(Slide.)

First of all, it ends up in the fashion and style section instead of the scientific section in the middle of the *New York Times*.

(Slide.)

There are three major companies and a lot of other companies that deal in direct to consumer testing, and this one is 23 & Me. Just to show you a few things, and I'm sure you have all seen it before, this is a clinical report, which over here is the clinical report. There are other research reports. Disease risks, there are 11 of them; carrier test, there's 21 of them; 10 traits; and 7 drug responses.

(Slide.)

Let me just show you very briefly cystic fibrosis as an example of a carrier trait and down here the Warfarin/Coumadin sensitivity as an example of a drug response.

(Slide.)

And there is lots to say about all of these various tabs. Tell me about your data, how it works, the timeline, et cetera, but I am just taking you in

this room to the technical report of cystic fibrosis carrier testing. It tells you a lot, and this is data sharing. This is for the person who paid for the test giving them complete information on the test. It has not only the 23 & Me name. It has other names. There’s deltaF508 as the most common variant certainly that you would be looking for in cystic fibrosis. It tells you what you’re looking for. It tells you your genotype.

And down here at the bottom it does not have any of 31 CFTR mutations. So it tells you the gene and has a reference where you could find more out about the gene. It most likely has no disease, not a carrier, may still be a carrier due to other mutations in the CFTR gene not recorded it. And I would say that that’s a pretty sophisticated kind of report and a direct sharing of data in that particular regard.

(Slide.)

And then here's another one where they actually tell you some clinical information for pharmacogenetics. This one, in fact, is looking at the results of two genes, the CYP2C9, where this

particular person is a *2/*2 homozygous for that allele and the vitamin K regulator gene, VKORC1, with promoter mutation.

(Slide.)

And here is the result saying increased Warfarin sensitivity may require decreased Warfarin dose. Now, there are lots of folks who were dealing with patients who have questions about all of this but, on the other hand, I have said before and I say again, it is out there and it is ours to deal with as the profession and there is a lot of data sharing that’s going on with this and a lot of curiosity and willingness to get these results and to investigate more.

(Slide.)

And the most telling thing about the data sharing is that you can actually download your entire set of results and go investigate them yourself if you're so inclined. So now it’s large files and you have to do some learning in order to have to figure out how to deal with it but that is what the public is doing at this point.

(Slide.)

Now let's talk about genetics, genomics and the EMR.

(Slide.)

This is the growth of laboratory tests. Now these are single gene tests for specific syndromes and specific mutations that are known to cause disease or be associated with diseases. This takes you to 2008. If you go to the end of 2009 that is where you get the 1,900 of these tests available and the purple is the laboratories, 600 laboratories around the world.

(Slide.)

But in addition to those single gene tests there is also a number of gene expression tests which are growing rapidly. I give you two examples.

(Slide.)

The first one is the Mammaprint. It is used to do a gene expression profile on the tumor and it is for prognostic purposes so that in 2007 the FDA cleared for marketing this test. It determines the likelihood of breast cancer returning within five to ten years after the woman’s initial cancer. The first cleared a

molecular test profiling genetic activity so it is a gene expression microarray test. It's a 70 gene profile and it is patented and available commercially, and you can send off a sample and get the results back. It is used routinely in some places.

(Slide.)

Here is another example of a gene expression test. It is called AlloMap. It is a molecular expression profiling, it's a little hard to read there in the back but that is for heart transplant patient management. This is a test that is a 11 gene profile. They look at the expression changes related to the immune system. It is used to alleviate morbidity associated with intra-cardiac biopsies. So when you get a heart transplant you have to go in regularly to get an evaluation to see whether or not you are rejecting your transplant or not, and the standard way for doing that is a cardiac biopsy, an intra-cardiac biopsy, which is somewhat invasive in my mind. I've heard surgeons say they are routine but they are not on the receiving end. You have to go in once a week in the initial stages.

What they're doing now with the AlloMap is to take a blood sample and run your leucocytes through an expression analysis microarray. If it looks like you are not rejecting then, in fact, you don’t need the biopsy. If it looks like you have signs that you’re starting to have rejection then you need the biopsy and further tests.

So those are numb to examples that are two examples and they are coming fast.

(Slide.)

So if you look at genetic testing in the electronic medical record you’ve got tests being done in all of these laboratories throughout the world and private laboratories. You have got a lot of test interpretations which are faxed back as opposed to being sent electronically. Tests are not stored in a structured form; not generally available for decision support. If your own laboratory does the test you have a much better chance of making that happen. The interpretation does not give too many details.

A MammaPrint doesn’t tell you the expression of each and every one of those 70 genes. It gives you

an interpretation overall and clinicians are struggling to explain these tests. And I would suggest that the rest of us are trying to figure out how to interpret them as well and explain them to the patients.

(Slide.)

And the business models of most of the laboratories doing the testing include the gene patents in many cases or the patents on these expression profile tests. They don’t necessarily have a business model that promotes data sharing. They make money on doing the test and not on sharing the data. And to compare and contrast this with this direct to consumer data sharing policy where they say we do a test, we give you the test, we give you the raw data, we give you the interpretation and it’s yours. So comparing that is a fairly major deal.

(Slide.)

So for electronic medical records, it has got implications for all of the component systems of electronic medical records. Certainly the laboratory exams are the ones that are impacted first and foremost but the rest of them as well.

(Slide.)

And one of the big things in standards and data sharing is messaging and vocabulary standards. There is HL7 as one of the standard methods by which you exchange data between systems. There is a clinical genomic standard which has been approved and has been started to be used or tested, and that is a big step.

(Slide.)

Here is a screen shot of Intermountain Healthcare, LDS Hospital, saying scientists clear major hurdle in genetic medicine. Dr. Williams was involved in that along the way sharing genetic data using the HL7 clinical genomic standard.

(Slide.)

At the same time genomic data is in all of these other information systems, especially public health systems. It certainly is represented in some form in newborn screening, tissue and organ banks. Department of Defense requires DNA samples of all new recruits and the identification of World Trade Center victims was a hallmark in the tools methodologies and techniques by which you could identify or re-identify

people based upon small bits of tissue which you find after a bombing or a Trade Center collapse. And those tools/techniques are used on a routine basis daily throughout the world with the suicide bombings and the terrorist attacks that happen.

(Slide.)

And at the same time you'll hear a little later today about some use of this genomic data looking at infective agent identification and the origin and spread. This year of H1N1 is big; SARS before, but the data is clearly there. It is not necessarily represented in a way which is standardized yet.

(Slide.)

There are definitely strategic information issues which have not been solved and they are being discussed. How to represent this data in electronic medical records is a large question. There are some systems that do that already. I'd point to the Helix Molecular Biology Subsystem within Cerner. How to send structured genetic data between systems is still being worked out, although the HL7 clinical genomics standard has started to solve that problem.

(Slide.)

How do you make this understandable to providers and patients, I think, is going to be a problem for some time because it keeps emerging as more and more information comes along. It’s not a settled issue and we learn more all the time so that’s an ongoing issue to be dealt with.

And then how do you keep all of this knowledge up to date is a problem for all of us. This is all emerging and what are the implications for healthcare and providers and patients, and how is it that you notify people of appropriate information and in all places in the world.

(Slide.)

You can have once again some examples. Here is an example of a genetic test, the CYP2C9 test, which can be represented in a pharmacogenomic decision support system but these are examples in single cases and not generally available throughout the world.

(Slide.)

What is coming? All of this stuff is coming. You cannot just settle in and think that you can deal

with what is there today when tomorrow the world changes. Certainly next generation sequence is interesting and here and now. The environmental variables have to be correlated in order to figure out what is the appropriate interpretation on many genetic tests. I think one that is interesting is the microbiome so that not only do I need to know what is my DNA, I need to know all the little critters' DNA that live with me in my life and help determine how I metabolize my food and react to various situations.

(Slide.)

Nanoparticles will then be interacting at the molecular level for therapeutic purposes and all of these things are coming and in some way or another will be part of our electronic medical record as we go forward.

(Slide.)

HIT standards, of course, are hot news today. There was a Technology Standard Panel established in 2005, a public-private partnership enabling the Bush first and now the Obama's vision of this nationwide system of electronic health record sharing by 2014. I

think that’s amazingly close to figure all of that out but things are going along rather rapidly.

(Slide.)

There is an interim final rule on standards specifically, which was issued the last day of ’09 and goes into effect next week. It is a final rule so it goes into effect at the same time it’s still being discussed and can still be altered as things go forward but things obviously are happening at a national level.

(Slide.)

I would say effective data sharing requires standards for data representation and transmission, and all of that is emerging in the genomics world. There are standard that are being discussed and developed. There clinical genomics standard is one. You’ve got a CDA clinical document architecture which is part of the RIM, the Reference Information Model, for test results. That is being worked out.

You have got representation of how to share the data and the gene expression data for the microarrays. You’ve got MIAME. That’s a way to represent the data. Here is a way to exchange the

data.

The same way for proteomics, which, of course, is the standard, which is at the base of tandem mass spec in all newborn screening. You have a ways to represent the data and a week to exchange the data being worked out.

You have vocabularies within the healthcare system, which are—SNOMED has been named as one of the standard vocabularies in the HTSB standards. You have got other vocabularies and representation of relationship between entities that are coming through these various ontologies. If you can’t represent the data and talk about it on a conceptual level then you don’t go too very far.

But I would say all of these are emerging and immature at the present time, promising for the future but not quite there yet.

Thank you.

Any questions?

COMMITTEE DISCUSSION

DR. ROYAL: Any questions for Dr. Mitchell?

DR. : I have a question.

DR. MITCHELL: Yes.

DR. : You talk about of 23 & Me and you used that as an example.

DR. MITCHELL: Yes.

DR. : Are there significant differences in how 23 & Me decode Navigenics and share their data with the consumers? Do you know?

DR. MITCHELL: I have not looked into the details of all of them. I do know that both Navigenics and 23 & Me will allow you to download your complete dataset if you request it.

There are some software packages which are available to an open source, which will allow you to accept that data and manipulate it.

Some of—I'm not sure which one. I think Navigenics says if you are going to do that they would like to talk to you first. So it requires not just a email saying, you know, send me my file. It requires, you know, let's set up a time to talk on the phone. Do you know what you're getting and it's pretty technical stuff.

But that’s a complete data file and there is a community of people who, of course, are doing that and who have their little Facebook pages and are sharing it and sharing interpretation, and 23 & Me, in particular, suggests that you might wish to share your data with the research community and other entities, you know.

It is very possible that 23 & Me may be making money off of various contracts that they have with companies that would like to have this data on people but I am not sure of the details of that.

Yes?

DR. ASPINALL: A very helpful report appeared. A couple of comments: I believe several of the firms have said that they are not sharing any data with any companies or any commercial interests in the midst of any of the genome because that’s part—having done my genome and all of them, that’s part of the agreement going forward that when you do that they have no other commercial relationships in doing that. So one of—oh, I don’t know—

DR. MITCHELL: Unless you agree, unless you

agree to share.

MS. ASPINALL: Yes, not with commercial entities. They have a number of agreements that they have talked about with some of the patient groups that then do it. And that may be another interesting piece about this is the patient groups. Things like—

DR. MITCHELL: Patients Like Me is one that—

MS. ASPINALL: Exactly.

DR. MITCHELL: Yes, looking for people who are like you, yes.

MS. ASPINALL: Yes, so that’s another—just one of the aspects and it was very comprehensive but one of the aspects that’s interesting that I think could be something that is increasing.

The Alzheimer's Patient Family Group has been very active in saying share full genomes. How do we, as a patient group or families of patients group, want to take this and then bring that to researchers who have agreed to deal with that so that is sort of one additional model there.

DR. MITCHELL: Yes.

MS. ASPINALL: I think the only comment that

I would have when you talk about the company piece for some of the companies that have products that are currently on the market, several of them don't have products that are currently on the market, the amount of information you are able to share really depends on whether you are CLIA approved or FDA approved now, and the ability to give that additional information has been deemed on occasion not possible because the regulatory authorities won't let a lab release the additional information if their approval is on a composite that you cannot do that but the genomics companies, the patient genomics, until very recently have been regulated differently and, therefore, have been able to give more information as they so choose.

DR. MITCHELL: Yes?

DR. FERREIRA-GONZALEZ: So maybe just a comment on what Mara just said.

I think that companies that have FDA approval, they can--it depends on what kind of information you are talking about because if you have FDA approval then you can say that your test is just for whatever you have validated studies for and what

you have approval or clearance for. You cannot say, well, you know, there is all of this other stuff that my test can do but I don't know what it means.

DR. : (Not at microphone.)

DR. MITCHELL: Or it is approve for this and it is the combination of these mutations. There are some data that says this one mutation alone is relevant to this other disease. You are not allowed to give that out because it hasn't been approved and—

DR. FERREIRA-GONZALEZ: Unless you validated it.

DR. MITCHELL: Right, yes.

DR. FERREIRA-GONZALEZ: But, you know, just to clarify. For instance, if you have 70 genes, it's their prerogative to say which 70 genes there are or not.

DR. : Yes.

DR. FERRIERA-GONZALEZ: So that's, you know—

DR. MITCHELL: Mm-hum.

DR. FERRIERA-GONZALEZ: --we are not telling them, no, you can't say—

DR. MITCHELL: Well, it definitely makes a

difference if you are providing a test for a medical purpose as opposed to for the curiosity of the person who wishes to pay for it.

DR. FERRIERA-GONZALEZ: I’m sorry, just another point.

That is very relevant. It's also interesting because many of these companies are saying we are doing this just for educational purposes but take your data to your medical provider and then it's a question of, well, what can that medical provider—what they have to do with it—

DR. MITCHELL: That’s right. That’s right.

Yes, I have a story on that, which is I have a colleague who is an emergency room physician, who said that he actually had a patient come in for an ED visit bringing his Navigenics report with him and being very anxious about the whole thing. And so my colleague said, “You know, I could treat you for your anxiety and that would be an appropriate emergency room visit but you coming in to want to talk to me about being anxious about this direct to customer test is not an appropriate emergency room visit. I will refer you

to a genetic counselor and a geneticist but, you know, it kind of has to stop there."

So it is true that you get folks who are starting with these test results out of curiosity and then it does make them anxious and then what do you do about the whole thing? It is a phenomenon and it is impacting the care providers.

Thank you.

DR. ROYAL: Thank you, Dr. Mitchell.

We are going to open up for a little discussion, if we have any, on the two talks and then we're just going to move—we're not going to take a break as we have in our program, we're just going to move in to talking about the different models. So if there's any other comment related to those two talks by Dr. Rodriguez and Mitchell, and then we're going to move on to the next talk.

Any comments?

All right.

Okay. We'll, we'll go ahead.

Steve, where is my program? I'm all confused here. This one. Okay.

The next talk we're going to have is from Dr. Catherine Schaefer, who is going to talk about healthcare systems.

CHAIRMAN TEUTSCH: These are the five models, right?

DR. ROYAL: Mm-hum. Yes, this is our five models that we're going to hear about.

HEALTH CARE SYSTEMS MODEL

DR. SCHAEFER: Thanks very much for inviting me here today to be part of this discussion of these—may we say topics rather than one topic of genomic data sharing. It is a very important set issues and an important series of discussions to have. We appreciate very much being able to be a part of this.

(Slide.)

You asked me here today to represent the perspective of the healthcare delivery system, and I should just point out that being part of Kaiser Permanente, particularly in Northern California, that this is a healthcare delivery system with a very large and active research division that is creating a very large and comprehensive resource for research on

genetic and environmental influences on health and, therefore, may not be typical of all healthcare delivery systems or even those that do research.

But this is the perspective that I am going to be talking about today as the issues that arise in any integrated healthcare delivery system with an electronic medical record that is preparing a very large resource to facilitate research on genetic and environmental influences on health.

The resource that we are developing will link together data on 500,000 members of Kaiser Permanente in Northern California, including comprehensive continuously updated clinical data from electronic medical records, data from participant surveys, data on environmental exposures, including social determinants and built environment, based in a geographic information system database, and genetic biomarker and environmental data from collected biospecimens.

(Slide.)

The purpose is developing this resource is really to enable scientists, including scientists

within Kaiser Permanente, but also the broader scientific community to conduct research on genetic and environmental influences on disease susceptibility, disease course, prognosis and outcomes, and response to treatment as in pharmacogenetics. Our aim is also to enable or facilitate, conduct research to translate findings into improvements in medical care and public health. And from the beginning we have also had the aim of conducting research on the ethical, legal and social implications of genetic research, and the use of genomic information in medical care.

(Slide.)

I thought it would be helpful if I gave you a little bit of background about this resource.

With initial funding that we received in 2005 and 2006 we developed (sic) a lot of time and effort to engaging the membership of Kaiser Permanente in Northern California and the sort of broader organization, providers, staff and so forth, through focus groups, internal communications and media about what we were planning.

And to sample concerns and values and better

understand the perspective of our organization, its membership, about development of this sort of resource, we organized community—separate community scientific and bioethics advisory panels and we spent a lot of time organizing our electronic medical record data by disease groups to facilitate research, creating over ten registries creating over 1.000—sorry, 100 diseases and conditions.

(Slide>)

In 2007, we started with enrollment of a general cohort and collection of survey data through a mail survey to 1.9 million people in Northern California, Northern California members. That is virtually our entire adult membership was mailed the survey, which sought information about demographic and background factors not included in the electronic medical record, health behavior information and so forth.

(Slide.)

About 400,000 people completed the survey over the course of about a year and then beginning in late 2008 we again contacted survey respondents and

asked them to provide written and informed consent, and the saliva sample. As of last month, about 130,000 individuals have provided written consent and a saliva sample.

(Slide.)

Our current activities include continuing efforts to enroll the planned participant sample. We plan to enroll a total of 200,000 individuals by the end of this year and reaching the goal of 500,000 participants by the end of 2013. We are beginning the collection of blood samples, phasing out the collection of saliva, using the clinical infrastructure, and we are continuing work on several funded genome-wide association studies, including a multi ethic study of bipolar disorder that involves 6,000 cases and 6,000 controls, and a study of prostate cancer among African Americans that involves 3,000 individuals.

(Slide.)

We have also developed a collaboration's portal and an access review committee that will be ready to receive applications later in 2010.

(Slide.)

And, importantly, we recently received a GO grant funded by the National Institutes of Health that supports genome-wide genotyping of the first 100,000 or so individuals/participants in our resource by year end 2011. This study was designed to be a resource for the study of age-related diseases, healthy aging and longevity. The average age of this first 100,000 participants in our resource is 65. So we have a large number of aged individuals and a large number of people in middle age whom are perfect for beginning to study factors that affect aging. We will be genotyping 650,000 SNPs as a part of this process and the resulting genomic data will be linked to data from the electronic medical record survey and environmental databases to create this resource.

It will be accessible through dbGaP and through collaborations, direct collaborations with us, and we believe that it will require reconsent for deposit of data in dbGAP.

(Slide.)

So considerations for data sharing in this environment, in this sort of a resource:

First of all, this is a very rich resource that would be difficult and extremely expensive to replicate in another environment or de novo. It is large. It is diverse ethnically and socioeconomically and it's generally representative of the population. The comprehensive continuously updated EMR enables excellent phenotypic characterization and follow up.

(Slide.)

Kaiser Permanente recognizes that the RPGH can make an important contribution and wants to ensure that the best and broadest use is made of this research consistent with its commitment to its members. So our perspective on data sharing is shaped by this commitment to our members. We are invested in them and they determine the future of this organization. So our situation is a little different than may exist in academic models. With this genotyping of 100,000 individuals and deposit of data into dbGAP, we clearly are going to have a lot of skin in the game, so to speak, with respect to genomic data sharing.

(Slide.)

So we are the very interested in and focused

on these issues even as we are extremely committed to the two data sharing, to the advances that we all hope this will bring.

Over 50 percent of our first 100,000 participants have been members of this organization and received healthcare for over 20 years. So both is a very rich—researchers can appreciate this is a very rich source of data since we have data on these individuals, comprehensive data going back to 1995 in an electronic format, and then data going forward as well.

(Slide.)

Trust in Kaiser Permanente by our members enables us to do research and so we are very—it’s very important and we’re very committed to maintaining that trust.

(Slide>)

In terms of factors that affect data sharing, and certainly informed consent and the nature of that informed consent is quite central, we use written informed consent that is broad and includes no restrictions on any kinds of health problems that could

be studied. Health information can be updated from the electronic medical record going forward in time. It contains a stipulation that all studies must be approved by an institutional review board and data can be shared with scientists outside Kaiser Permanente who agree to protect confidentiality and follow rules for use.

(Slide.)

The informed consent stipulates that using and sharing genomic data will be for research purposes only. Research results will not be placed in the electronic medical record and participation is confidential. Genomic data will not be returned to individuals or their providers. Participants, however, may be contacted if information develops that has significance for their health. Participants may withdraw and may ask that their sample be destroyed.

So one of the questions that arises is how we ensure that these latter commitments made in the informed consent are met when data are used through a public database where we have less information or less probable feedback from investigators who use

information that way.

(Slide.)

Informed consent does not historically really address—really addresses issues of sort of individual autonomy but has less to say or has not historically been used to address issues that can arise about social harms that may arise in the process of carrying out some kinds of research.

(Slide.)

So in our environment concern has been expressed about data sharing through a federal database such as dbGAP. Our community advisory panel focus groups that we have conducted and some survey respondents very directly have been concerned about this issue and expressed the idea that the government may “take or misuse” data.

The building of the other federal DNA databases increases the perceived vulnerability of this NIH database to re-identification or misuse at least in the individuals who are concerned.

And the use of DNA to deny treaty rights or label immigrants or other sort of forensic uses is also

a prominent community concern.

There is the concern that research may be done that could be used subsequently to stigmatize a vulnerable group, that there is—once a broad consent is signed there is actually no recourse of the individual other than withdrawing from the resource; no choice is involved about the kind of research that is undertaken then with the resulting data.

And then there is the perception on the part of our members that storage and control of data by Kaiser Permanente, by the resource, sort of local storage and control gives participants better recourse and control over events.

(Slide.)

I just want to mention--I think you are going to hear quite a bit from Dan Masys about this—a later speaker today. But the obvious fact that in most research contexts sharing genomic data means sharing phenotypic data. And so we also need to consider factors affecting the sharing of these other forms of data that may be linked to genomic data.

(Slide.)

Health plans with EMRs, which as is the case for our resource, have huge investments in the EMR data, that is so rich, and then is linked to the genomic data.

The quality of phenotypic data that you can derive from these very high density EMRs is critical to the best use of the genomic data and the resource, and it’s challenging. Let me tell you very challenging to extract high density data that will be useful to all for a whole variety of studies that can then be deposited in a database such as dbGAP.

(Slide.)

The best use of the data depend on knowledge of the system that generated the data and this is often--the meaning of this sort of clinical data, even when standard diagnostic codes are used and efforts are made to harmonize data across systems, is often--it really takes an understanding of how that data has been generated to make the best or most valid use.

(Slide.)

Partly in response to a variety of these concerns, we have begun at this point to perform a

series of stakeholder interviews led by Carol Somkin, who is the head of our Ethical, Legal, and Social Implications Core, with the goal of informing the development of our access and collaboration policies and procedures.

So we have been conducting these qualitative interviews with a variety of stakeholders, as listed here, and with the following sort of research questions, such as what are the specific data sharing, benefit sharing and governance issues inherent in a biobank that is situated in an integrated delivery system?

Well, that’s really what you wanted me to tell you about today and I regret to say that we have just begun these interviews and so I really do not have data that I can present about the outcome but the things that I talk about are the result of sort of earlier focus groups and interviews that we have conducted.

I hope there is a chance actually to come back and tell you a little bit more about the outcomes of these interview efforts at a later date.

Thanks very much for your time.

DR. ROYAL: Thank you, Dr. Schaefer.

Any questions?

Marc?

DR. WILLIAMS: Your institution, like many, has policies allowing subjects of this type of research to withdraw and, by that, meaning have their samples destroyed. Are you familiar—aware of any cases where such a request has been made at your institution?

DR. SCHAEFER: Yes. Actually it happens rarely but it has happened and it has already happened with this particular resource.

But I know of only five instances out of 130,000 individuals participating where that has happened.

DR. : (Not at microphone.)

DR. : Thanks. Could you comment on how the providers in the Kaiser system understand what the heck you are doing and, also, I am curious about the uptake or the frequency of participation by your members. It seems quite high and I wonder whether-- what you have done to foster such high participation.

DR. SCHAEFER: Well, I'm delighted to hear you describe it that way. It has mostly been an effort that has been—you know, the traditional ways we know how to contact people, which is essentially mailing people materials that are descriptive of the research program, the eight page—the consent form written in—consent form is eight pages long if you include the HIPAA authorization. So I think the way we look at it is most of the participants that we have garnered so far are--and perhaps this is one reason why we have very good representation in older age groups--are people with the time and patience to basically make their way through printed material that they receive in the mail.

So our next efforts at enrollment actually are to carry out different sorts of efforts that don't involve only essentially reaching out to our members through a mailed written material format but involve other ways of engaging people.

With respect to providers, our providers are perhaps—well, they--we have had a research division since 1966 so they are familiar with essentially having

patients recruited for studies. The research division essentially operates sort of side by side with the providers but we do not typically recruit through providers. That is we do not ask physicians to obtain—to talk to their patients and ask them to participate in studies. There are certain clinical trials that are exception to that but, in general, for this sort of general research we don't do that.

What do they think about it? They are very hopeful that in the not too distant future we will begin to be able to do translational studies that will involve them more directly and that will fulfill the promise that is held out there that this kind of research will result in things that directly improve healthcare.

DR. : An interesting project. You may not know the answer to this but I am wondering whether there has been any—it may be obvious why I'm asking this question--training to the healthcare providers, the physicians or nurse practitioners who are the primary providers, not the researchers, to address issues if their patients come maybe having read

something in the press about a genetic study or maybe knowing more about this study. Who do they go—who do your participants go to for sort of small questions? Not informed consent kind of questions but health related questions? Do you know what I mean?

DR. SCHAEFER: About this study you mean?

DR. : About genetics in general. I would assume that just being—reading through the consent form, the eight page consent form, would make the participants a little more alert to things in the press about genetic research in general, that they might then go to their providers with general questions and is there any training for those providers within Kaiser for this?

DR. SCHAEFER: I don’t think as yet that there actually has been any provider training in how to respond to sort of general questions about this kind of data in particular or these kinds of large scale genomic studies.

The providers have what we have provided to them and that as yet is not a great deal of information.

We do have a strong medical genetics department that is distributed across the region. And those providers, themselves, for example, have organized, about the time that, for example, BRCA-1 testing became available, we anticipated that there would be a lot of general interest in this even though the test was really not in the appropriate for women who by virtue of family history might have a low to moderate risk of inherited susceptibility.

So the genetics providers actually pioneered a class that any woman could come to and that would sort of explain what BRCA1 is, the sort of role of family history in risk of breast cancer, and then women could self refer then for genetics counseling if they thought that—and then subsequently for the test if they thought that this was something they really needed.

So we have a little bit of a model of how to handle a situation with--where there is sort of general interest in something but, in fact, there may be—and that’s a case where the test is really only appropriate for a relatively small number of people.

DR. WILLIAMS: So a couple of brief

questions. One is a follow up to Michael's question, which is if somebody leaves Kaiser to go to another payer, how do you handle people that leave the system in terms of participation in the study?

And then the second question is just to resolve what, to me, is an apparent contradiction but probably just represents a lack of information, which is given the age distribution and the membership length, how representative is your sample actually compared to the rest of the Kaiser membership specifically and maybe the population of California, in general?

DR. SCHAEFER: Let's see.

We actually do not have a good solution in terms of continuing someone’s participation if they leave Kaiser Permanente and go to another system.

Essentially, we have then no way to really follow up in the same sort of way their medical history at the point at which they leave us. So the informed consent gives us permission to use the data that we do have the system or died, for example, but right now, at least, we don’t really have agreements with other

systems for sort of continuation of observation.

And with respect to—

DR. ROYAL: Okay.

DR. SCHAEFER: I'm sorry. Can I answer his question about representativeness or are we—

DR. : (Not at microphone.)

DR. SCHAEFER: Okay.

DR. : (Not at microphone.)

DR. ROYAL: We were going to take one last one but—

DR. WILLIAMS: Well, she had a second part I had asked.

DR. ROYAL: Oh.

DR. : (Not at microphone.)

DR. ROYAL: Okay. Yes. Okay. And then we will move on to the next speaker.

DR. SCHAEFER: Well, the answer is that while we have good representation of different groups, it is actually—I would not say that it's exactly representative of the population of Kaiser Permanente, which is generally representative of the population of Northern California. So we have--now the way that we

have been enrolling people, for example, it's older, more female, more White, and better educated than our general membership is.

DR. ROYAL: Thank you, Dr. Schaefer.

We'll hear from Daniel Masys from Vanderbilt.

Dr. Masys is going to talk about the academic model and then we're going to take a break.

ACADEMIC MODEL

DR. MASYS: Thank you. And since I stand between you and a break, I will move with all due expediency through a presentation that shares a lot of the elements you've just heard from Cathy in the sense that these are phenotypes derived from electronic medical records combined with genome-wide scan. And I'm doing that in my capacity as the principal investigator for the National Coordination Center for a consortium called eMERGE, the Electronic Medical Records and Genomics Consortium.

(Slide.)

Three topics in the next 15 minutes:

First, what eMERGE is; lessons that we are learning about data sharing; and then I'll focus, in

particular, on where we are with respect to the science, the emerging science of data de-identification and re-identification.

(Slide.)

This consortium grew out of a request for applications by the Genome Research Institute in 2007. The key element of which is highlighted here in red that, in essence, was support for investigative groups affiliated with--you had to have an existing biorepository and then you had to have the ability to extract phenotypes from electronic medical records.

The consortium: Members of the awardee institutions are five that you see listed on this slide and they have both a geographic distribution and a pretty wide distribution in the differences with which they acquire both their biobanks and their clinical data.

(Slide.)

Here's a map that shows the primary phenotypes that were part of the original project submission. So as part of the grant submission you had to propose what phenotype you were going to do a GWAS

on as the anchor for participation in the network but we have since expanded that with a number of cross network phenotypes.

So you see here that they range, and I will actually note as well that there is a variety of sizes of biobanks, so in the upper left hand corner then, the Pacific Northwest Group Health of Puget Sound, had essentially a dedicated Alzheimer's research cohort that was linkable to Group Health data and their biobank was about 3,000 samples.

The cataract primary phenotype for Marshfield represented probably the most mature health system based biobank, one that many have been built from, which was a prospectively consented cohort of about 22,000 individuals in Northern Wisconsin.

The Mayo Clinic one was again about 3,000 samples focused in peripheral vascular disease built from a research cohort as its anchor.

Northwestern was looking at Type 2 diabetes with a general purpose biobank built from all comers into an internal medicine environment with a prospective consented biobank participation.

And then Vanderbilt was looking at a—is looking at a continuous trait of the QRS duration as a predictor of future cardiac events and the Vanderbilt model is a non-human subjects, that is a de-identified biobank built from discarded blood samples where the DNA is extracted unless patients have elected to opt out of that mode3l. And Vanderbilt is also the site of the coordination center for the network. Vanderbilt started their biobank in about 2007. It is just a little north of 76,000 samples now growing at about 500 per week.

(Slide.)

So, again, the features of our network, each site having DNA linked to the corresponding electronic medical record data. An important component and requirement of the RFA was community engagement and so investigation into models of consent and re-consent. Two of the five members have had to re-consent their members because of the last condition shown on this slide, that is the submission to dbGAP as a condition of NIH funding. It was part of the—already part of the model of consent for others and so it was not the case

that all of the groups had to do that.

The core was a 3,000—roughly 3,000 subject GWAS study that gave us roughly 20,000 genome-wide scans that we could then not only do the primary phenotype associations but mine the associated EMR data for other opportunistic, if you will, phenotypes and I will show you a little bit more data about that.

(Slide.)

We have since received supplemental funding for additional new genotyping for our cross network phenotypes and that work is in progress now.

(Slide.)

This is an example of conditions that were not part of the original proposal but they do represent data that because it's so commonly acquired just in the natural course of people having routine testing and electronic medical records, it gives the network the opportunity to share samples, and you see here that by and large they range for most conditions in the thousands of samples for which the genotyping is essentially already done because of a—we—the basic platform is a 600k Illumina genome scan, although the

African-Americans—we have about a one million SNP chip on about 2,000 samples across the network.

So this ability to look at red cell and white cell indices, diabetic retinopathy, lipid levels, GWAS studies on height, which by and large are replication studies at this point since they're already published dedicated research cohorts and glomerular filtration rates are emblematic of the fact that it is a data rich environment, such as Cathy described, and it allows us to begin with a genome-wide scan and then look at various aspects of the phenotype.

Now what we have along the way is—in fact, the inside joke in eMERGE is that any fool can get a genome scan and many do; the real hard part is the phenotypes. And so the informatics issues that we are engaging are essentially because we are going to pool and do meta-analysis of phenotypes, how comparable are the patient populations who walk through the doors or sign up for the cohorts in these five different health systems because if biologically there is some inherent bias in the nature of these patient populations then pooling their genomic data may mislead us with respect

to statistical associations.

We have discovered that genotypes are pretty easy to share by virtue of the NIH supported genotyping centers and so sending samples and receiving datasets, not unlike the ones you can download from 23 and Me or Navigenics, is actually the easy part in terms of it but there is as yet no set of standards that represents the kind of genotype/phenotype package where you can send the whole thing in one electronic envelope, and so we are developing in association with NCBI sort of standards for how clinical data can be put into a format that is useful for association studies.

Clinical data has a number of features that make it different than the classical cross-sectional research cohort that has been published in the GWAS literature up to the current time. One of those is we can't predict how many times some measures will be done. For example, if you have a diabetic who has blood sugar measurements, there may be thousands of them in the record so how do you decide which ones to include in a research data submission, as well as the feature of EMRs that clinicians are absolutely

comfortable with the notions that they are—that are clear that some people have definite diseases, others may have probable or possible, and that notion of uncertainty that lives comfortably in the clinic is not well suited to this research environment that looks more like a case report form that has a sort of dichotomous representation that you either have a condition or you don't. So one of the things we're working with dbGAP is the assertion of whether a condition is present or absent and our level of comfort that exists in EMR.

(Slide.)

I would like to focus on the last thing that we are making progress in the network on, and that is this re-identification potential that arises particularly out of clinical data and those phenotypes associated with the genetic samples, with a general model that we would of course like to maximize scientific value while complying with the federal privacy policies that Laura Rodriguez has mentioned and will mention in greater detail later in this session.

(Slide.)

This is the screen shot from the dbGAP data submission policy, and I've highlighted in the little red box there that says if you're a submitter to dbGAP you have to send your phenotype exposure and genotype data without identifiable information using a unique code and such.

And so the question is when you have got clinically derived phenotypes, how do you do that? It calls to mind, I think, an important set of vocabulary because IRBs always get balled up with this about the notion of, well, is it anonymous or not anonymous, or what do you mean by de-identified. So I think in this regard, at least in the computer science and informatics community, we regard anonymous as this definition that things are not traceable to an individual and that it was a concept prevalent from about 5000 BC, the time of Hippocrates, to about ten years ago, and it was generally thought of as a dichotomous variable, that is anonymous. The data was anonymous or it wasn't, and the IRB was happy with that assessment.

But what we know now is that we had to

replace that with something that looks like kind of a slider bar that it has replaced anonymous because we recognize that biologic data is so inherently rich in attributes that its re-identification potential essentially never goes to zero.

(Slide.)

And so it’s a continuous variable whose properties can be calculated for some but actually not all types of health data. The primmer on re-identification is a simple one, and that is if a dataset has ostensibly been de-identified then the way--the pathway to trying to find out the identity of the individual from who it is derived requires two conditions.

The first is out of many records getting a unique set of attributes, what computer people called a logical unit record associated with one individual. So you’ve got to get to uniqueness first.

And then that's necessary but not sufficient. And a lot of the U.S. population, whose understanding of genetics is mostly informed by the OJ trial and CSI, believe that DNA is inherently identifying as if you

found a poly-vial of it on the carpet here you'd actually know who that person was. And so what you need in addition to the biology of uniqueness is you need a naming source. You've got to be able to intersect that with a person's demographic information.

So as a result de-identification methods basically are aimed at either preventing you from getting isolated to a unique record, that there's always more than one that satisfies any set of characteristics, or you might be able to get a unique record but what you can do is block the linkage to a naming source.

(Slide.)

This is a graphical view of this from work—and I'm going to present to you a couple of slides from Brad Mullen, who is a faculty member in our department at Vanderbilt, who is a data privacy guy. In fact, your briefing materials have one of his recent publications.

And, in essence, it shows on the left hand side that—and actually in all of—all three of these conditions have to be satisfied. You have to be unique

on the left-hand side with respect to your de-identified dataset, you have to be unique on the right side in terms of named data such as a voter list or a health and vital statistics registry, and you've got to get a firm linkage one to one between those two models.

(Slide.)

So, let's look first at uniqueness.

Well, if you take clinical data—this is our own cohort of the 2,500 Vanderbilt patients that are in our genome-wide study. One can say, well, how alike are they with one another based on common measures that are in clinical data? One of the common ones that he used is ICD9 disease coding.

(Slide.)

To cut to the chase, out of our—using it as a reference population, it's now about 1.9 million records in our EMR, about 97 percent of people are out of the box unique. It's just their--the combination of their age, their gender and their ICD9 codes, if they have--on average we have about 12 codes per person. It only takes about five codes and all of a sudden you are in a box where *n* is one in the cell. And, so, it would

seem that we are in very good condition to be able to do--sending out the entire detailed set of rich phenotypic attributes representing even ICD9 codes.

Now, we work in a world that’s governed by HIPAA. And so it has two nominated standards for data sharing. The more stringent one, as you probably know, is called Safe Harbor, and it allows you to release race, gender, only year of birth, not date of birth, and only state as the smallest geographic entity in most cases. And then there is this called a limited dataset, which allows you, in addition to those, to increase the specificity so you can release the actual date of birth and you can go down in most cases to a county level thing.

(Slide.)

And then the question is what kind of linkage does that--potential does that give you for identified data sources?

(Slide.)

So here is Brad's work on the pooled U.S. census data from the year 2000 that shows you the fraction of unique individuals under HIPAA safe harbor,

that is you're releasing only year of birth, Sex and race on the left. And the important thing here is it's not zero. The HIPAA safe harbor standard has roughly about a 10^{-4}, unique, and it still exists in that, and it depends upon the states, how sparse or densely populated your state is.

You'll notice that there is a dramatic increase in the number of records that become unique when you go to the actual date of birth, the sex, the race and the county. So now we're in the range of about 30 percent to almost 100 percent of individuals can be uniquely isolated inside of a clinically derived dataset.

Well, how about the naming sources? If the issue is that lots of things are unique in an EMR, here the story is very highly variable across the landscape, both of information resources on the Internet but, importantly, across states because a common re-identification source is to use either health and vital statistics registry or voter records. So these happen to be the state policies and the data items available for the states of the participants in the

eMERGE network. They include Illinois, Minnesota, Tennessee, Washington, and Wisconsin. And you see that authorized users include in three of the states anybody in Minnesota. You have to be a Minnesota voter. You have to be a political person in Illinois but isn't everybody.

(Laughter.)

And so you can get the stuff—we can get this stuff on a disk and it ranges from $20 to $12,500, and you get a variety of different data elements, including date of birth. You always get name and address. So, the question is what are the other things that map, for example, to those HIPAA limited dataset items?

If you then—as a result of that availability, and we've done this for all the states in eMERGE, you take—that graph on the left is exactly the one you saw before. That's not identifiability. That's just uniqueness.

(Slide.)

So what happens to uniqueness when you merge it with a naming source? And you see on the last—on the right-hand side that the number drops but it

doesn’t drop dramatically. So, in essence, NK here is, by the way, the cell size. Because you could say the re-identifiability doesn’t begin at just a single record. Maybe it begins when it’s only—when you’ve got a pool of five records. That’s close enough where we could then use other methods to try and zero in.

So you see at even a K of one that where 30 percent were unique, it drops to about 15 percent but that means 15 percent of that entire population you have a name, address, all—you’ve successfully and fully re-identified the individual from the de-identified data.

(Slide.)

So as a result of that ability to do a quantitative analysis what we found in the network is that the clinical data that we are sharing with dbGAP is going to necessarily need to be a subset of those present in the full clinical record, specifically by removing uncommon codes that support elevated risk re-identification risk. And when we say “elevated,” we mean elevated above the HIPAA standards so we can quantitatively say what the HIPAA standards are and

then we can mathematically meet those same standards by a variety of methods.

(Slide.)

Now, between the members of the network with respect to an academic sharing, we actually all have lawyer-approved data sharing agreements at the individual record level. So that works fine among the consortium. It took us 18 months to get, you know, n by n, and way we did that was everybody made an agreement to share with the coordination center as opposed to having to do four other agreements with four other institutions.

(Slide.)

The eMERGE coordination center in its capacity as a data quality and analysis center is providing data privacy consultation to the network members, including quantitative assessment of the re-identification risk of their datasets before they go to dbGAP because they vary on the different disease populations. You might imagine the Alzheimer's disease population is highly skewed to older individuals so it more impacts the HIPAA standards about people that are

ages 90 and above.

And the good news is that we're also just about—we have a couple of manuscripts in review and I'm just about to release some tools that will be open source, usable by mere mortals for actually determining the quantitative risk of the demographics of publicly submitted datasets and how you can, in essence, trade off for scientific purposes the granularity of one item so you can kind of smudge the zip code, if you will, if it's important to maintain age because that's an important dependent variability in the analysis.

And that's the sort of good news about the statistical standard is you can do various permutations of the data in order to meet the formal federal standards and you are losing some content but if it's content not important for the key scientific hypothesis then it's still kind of a whim.

So that's where ewe stand. It's a work in progress and we'll be happy and will be reporting in the literature and I'm happy to report to you as we make progress on these issues.

Like all good networks, we have a URL with

the unpretentious URL of GWAS.net and so as all of our publications and our white papers for practices within the network are posted on that website.

And, with that, I'd be happy to answer any questions.

DR. ROYAL: Thank you, Dr. Masys.

Any questions?

Marc?

` DR. WILLIAMS: So on one of the first slides where you talked about the RFA, I think there was a reference in there to the use of natural language processing.

DR. MASYS: Yes.

DR. WILLIAMS: And so I am just curious how is that working out for you?

DR. MASYS: Yes. So what we discovered is that-- well, some people have said—I mean the null hypothesis for the whole network is that EMRs are so bad you couldn't use them for anything. Right? It's just a mess. So what we have discovered is that in order to get a positive predictive value of a phenotype definition, what works across multiple EMRs, you need a

combination of structured items, including codes, the ICD9 codes, labs, specific lab values, and importantly medications because in a sense medication is the sincerest evidence that a clinician thinks you have a disorder. And, in addition to that, that journal gets RPVs in the range—it depends upon the condition but roughly only about 65-75 percent. We have to use natural language processing, that is teaching computers to identify concepts, diagnostic concepts, and whether they are asserted or negated in the record to get RPVs in the 95 percent, and that's most of them are in that range.

The good news is we use experts to do that. Basically scoring how good the algorithm is.

It generally takes about five iterations to get it right. Then when one of our institutions gets it right we can actually—we found we can actually transport that across the network and, with relatively minor modifications, most of the PPVs only fall a few percent when they are re-used as selection logic in very, very heterogeneous EMRs. So that's the unexpected big win here is that if one group does the

work of creating the phenotype selection logic and we're going to build public libraries of these, other institutions that want to use these to find cohorts of interest, either for administrative purposes or for research purposes, can reuse that without a lot of having to redo the wheel.

Yes?

DR. ROYAL: Questions.

Andrea?

DR. FERREIRA-GONZALEZ: It's a very impressive presentation.

I was curious to see if you can elaborate a little bit more about the process of the informed consent of the patients. What I understood you were talking about is they use procedural specimens and they will be discarded otherwise. But you have mentioned also that unless the patients opt out of having that—so what is the process of the informed consent or there is a blanket informed consent as they come through the Vanderbilt institution that they will be enrolled in this unless they actually specifically—and how that process works.

DR. MASYS: So Vanderbilt is the one member of the network that has—that works in a de-identified space where both the records and biological samples are de-identified and we cannot construct identities to go back and contact individuals.

The general model has been published, and I would be happy to sort of provide it as the reference but the short version of OHRP-approved is that in this nonhuman subject space the federal regulations would have actually allowed us to view this as existing tissue in data without notifying anybody. Our ethics board and our IRB said, “It doesn't sound right.” And so in the conceptualization and implementation of biobank which was preceded by a number of surveys of patient attitudes and such, we added this component of a very extensive public notification campaign. The fact that people re-signed a consent for treatment and right above the signature line the only bold-faced type in the whole thing is a big box. In bold face type it says "I understand that Vanderbilt extracts DNA from leftover blood samples and I should check this box if I don't want to have my samples used for that research."

On average, now having run for about the last 30 months, we had a predicted opt-out rate of five percent, and that's exactly what we are observing, right at about 4.9 to 5.2 percent on that. And, generally, broad acceptance of Vanderbilt patients based on what Cathy said, and that is that while our patients--they may not trust the government but basically they trust the institution that they are getting their healthcare from. So that they are willing to let Vanderbilt do this kind of research.

And, as I say, to not turn this into an hour-long discussion of this model, and we can maybe come back and give you the full soup to nuts, we have published it and I will send you the URL.

DR. ROYAL: Okay.

Any other questions?

No?

Thank you, Dr. Masys.

We'll take a 10-minute break so we'll come back at ten after 3:00.

(Whereupon, at 3:00 p.m., a break was taken.)

CHAIRMAN TEUTSCH: I'll turn it back over to

Charmaine.

DR. ROYAL: We're going to hear from Laura again, Laura Rodriguez, Dr. Rodriguez from the Genome Institute, talking about the government model.

GOVERNMENT MODEL

DR. RODRIGUEZ: Okay. So I would like to thank the committee again for having me to talk to you all again, and I promise this will be the last time this afternoon.

(Slide.)

So now I'm going to switch to talking to something that I do know much more about than talking about all of the activities across the federal government, and that is—I'm not sure that I would say it is the government model for genomic data sharing but it is one of them, and it is an NIH model that we are seeing become increasingly consistent, as we talked about before, across the NIH and across the different institutes going forward.

(Slide.)

There were several questions the task force asked the speakers to try and address in their

questions, and so I’m going to do that through the course of the slides. And for that reason get down to some of the nuts and to answer, I will try to get down to some of the nuts and bolts about the process for how this works to try and address things like informed consent and responsibilities for different aspects of protection along the way.

(Slide.)

So as you all know, data sharing is nothing new to the NIH. There has been a longstanding tradition of sharing resources and tools, and of having large policies for all of our extramural grantees on the expectations around how they share data.

And, traditionally, this has come forward, I think, in one of the more major statements in 2003. For any grant over $500,000 to have the data shared at the time or post completion of the study the data was to be shared broadly and made available.

So what's different--one of the things that’s different about GWAS as we move forward was that GWAS began to merge the genomic traditions of making data rapidly available prior to publication into the NIH

realm of the broad data sharing.

And the reasons that we did this were partially—largely, of course, based on of scientific opportunities that were coming forward as the technology became accessible to do whole-genome scans to look at so many different points of variation across the genome and actually be able to try and tease apart the genetic underpinnings of common diseases which have been so difficult to address through standard genetic mechanisms and strategies in the past.

And so the opportunity to do this and the breadth of different institutes that were disease focused that were wanting to try and take advantage of these new strategies were something that were very—a very strong force for the leadership at NIH to say that we needed something that went across the board, across all of the institutes so that there was consistency in expectations for the investigators, and again for the public in terms of what they would understand about the data that was out there and what would be in place for the protections for how these data would be shared.

And, of course, again, I think you are well-

aware of the power that the genome-wide association data had in terms of the richness of genotype and phenotype information, and the ability to ask many different questions of the data, and thus all supporting the reasons to have as many different investigators have access to the data as possible so that they could ask as many different questions as possible.

(Slide.)

And this brings us back to the guiding principle and really the foundation upon which everything came from as we constructed the policy and all of the different elements within the policy. And that was to try to achieve maximum public benefit from the federal investment and from the wealth of information and data generated through the different studies that NIH was beginning to fund.

(Slide.)

The policy itself was broken into three primary sections. The bulk of the language focuses around data management, and that speaks to the importance that the NIH put on both standards and

expectations for data submission and data access but also for the protection of the data. And of the interest of those individuals whose data was within the resources that the NIH was creating through this policy.

(Slide.)

The way that the process works, of course, is that everything is built upon primary research studies, which take place perhaps outside the realm of an NIH funded study. It took place sometime in the past where there was a relationship between research participants and an investigator that is structured around an informed consent discussion and agreement between those two individuals about how the data will be used and whether it will be shared, et cetera.

And at the point in time that an investigator decided that they wanted to apply to the NIH for funding for the genotyping is the point--was the trigger point for the GWAS policy. And at that point then there would be an expectation and is an expectation now that the data will come in to a central repository that is housed at the NIH and that

repository is the database for genotypes and phenotypes. It was newly constructed at the time within the NCBI and there's a lot of information. These different screen shots simply portray the range of information about different studies that are available from the protocol and the survey instruments that were used to averages of the phenotype data, as well as different views of the genomic information from the genotype data so that people can zero in on where they might want to look.

(Slide.)

The other advantage to having the central resource, besides being able to make documents such as these examination procedures, which for many studies has been in existence only in the lab's file cabinets, to now available and be searchable through open access pages on the web, is that people could find new collaborations, they could preview studies and also try to find out if they were relevant to the kinds of questions that they were going to ask before they ever attempted to request access from NIH.

(Slide.)

All of the data that come into dbGAP are de-identified and the way that we define de-identified since, as Dan mentioned, this is a variable term, was to look to the HIPAA standards and the 18 identifiers named within the privacy rule, and use that as the basic rubric by which investigators were asked to de-identify information, to hold the key to the code within their institution and not to share it with the NIH so that when the information came to the NIH we did not have any way to link back to the code for any of the individuals within the datasets.

(Slide.)

The third and final phase, of course, is to make this data available to the secondary investigators. And this would be through a controlled access process, which I will talk about in a moment, and again they are only ever getting access to coded information, and they would be—they would request the data for a specific research purpose and project.

(Slide.)

Coming back to measures of protection, one of the things that NIH did in this policy, which departed

from the basic regulatory requirements, was to attach an expectation that the informed consent of individuals, and those agreements that may have been made in terms of how data could be used in the future or how data could be shared, would remain attached to how data were distributed through the resource. So, again, all of the data that since they are de-identified and, therefore, don't technically represent human subjects, data, once they come in to the NIH or for use by the secondary investigators, we still maintain that the informed consent was an ethical principle that we wanted to have follow the data as it went out and was used by others.

(Slide.)

In terms of implementation for the policy, the local institution, consistent with general practices where the IRB is the authority for any study that happens, is asked to provide a certification to the NIH which stipulates that the dataset and all of the data within it are appropriate to come into the data repository and to be distributed to secondary investigators.

They are specifically asked to have an IRB review elements of the informed consent and state that the consent is consistent with use coming through dbGAP.

And also, again, assertions that the PI will remove all of the HIPAA identifiers so that it can meet that standard of de-identification set forth in the policy.

Any limitations on future data use are also requested through the certification. This can speak to issues around informed consents so that if data were collected under an agreement where the data would only ever be used for cancer research, the NIH is aware of that and can only ever release it or distribute it to secondary investigators also doing cancer research, but also for other issues where IRBs may have concerns in terms of the particular data elements or how things are going forward so that we can respect again the decisions of the local institution coming into the NIH.

(Slide.)

In order to try and find or provide some information to local institutions on these new

responsibilities for the data that would be coming in to the resource, we did craft a points to consider document that discussed all of the basic elements of the policy, as well as some of the overview of the science. The audience for this points to consider document really was intended to be the IRBs who might not necessarily understand or have the background in the science at the time they were first seeing this come through and being asked to provide the certification.

The points to consider walks through many of the elements within informed consent that the NIH felt were important for institutions to take a look at within the informed consent documents but it is not intended to serve as a checklist. And so it still leaves to the discretion of the institution what is appropriate and what would not be appropriate based on their own deliberations relative to the particular population in a given dataset or relative to the institutional policies at their research institution.

(Slide.)

Data access, as I've already mentioned, was

two-tiered. So there are public access pages that are available for anyone to look at to get basic high-level information on the studies within the data set. One, to understand whether or not the dataset might be interesting and relevant to the questions that they would like to ask but, in order to get to the individual level coded data, it had to come through a controlled access process where every investigator seeking the data will need to submit a specific research use, proposed research use, that would be reviewed by a data access committee and the decision would then be made.

And the point of the specific use was in order to have--for the data access committee to make a determination about any limitations on data use provided by the local institution at the time of submission.

(Slide.)

To try and gain some accountability for investigator practices, once they had the data, every request for data must come in co-signed by institutional official. So that the institution at the

secondary site is taking responsibility and sort of vouching for the credibility of the investigator that's coming in, and acknowledging they know this investigator has the data, they know that they are intending to use the data, and that the investigator is in compliance with any of the local policies that they have put in place for how data use of this kind of genomic data, whole genome data, is used at their institutions since, again, different institutions have different policies about how they review the conduct of research with whole genome information and coded specimens information, in general, at their local sites.

(Slide.)

I think I have gone through some of this already.

The data access committees in terms of who they are—because all of the data reside within a government database, they represent government records and, therefore, only federal employees can make decisions about access to the data.

So DACs are consisting only of federal staff

but are able to consult with anyone in the process of reviewing a document. So they can bring in an expert in a particular population if they have concerns about potential group harm, for instance, or they can bring in a scientific expert if they are not sure if the particular proposed use actually fits within the use limitations provided by the organization.

The other function that the DAC has in addition to reviewing incoming requests is to track the data use by those users that they have approved within the database. And so annual reports come in for all users where they talk about any significant findings for the work that they have had, any publications coming out of it, any IP, et cetera, that may have been noted. And that also provides a way for the DACs to go back and make sure again that they are only working on that proposed use that they submitted for approval at the time and not doing something else with the data.

(Slide.)

The agreement between the secondary investigator, his or her institution, and the NIH comes through the form of a data use certification agreement.

We have now created one common model template for all of the data access committees to use for every dataset that comes through the NIH, which was something we didn't have at the start so that is an improvement and, hopefully, it will make things easier for institutions and investigators to understand what they are agreeing to.

(Slide.)

The terms and conditions—some of them are fairly obvious in terms of being responsible for compliancy with federal and state law to only use the data for those things that they said they will use the data for. There was a promise not to attempt to identify the study participants either based on the information that they receive from the NIH or by combining that data with any other dataset that they might have access to, public or otherwise.

(Slide.)

And, importantly, too, as a measure of transparency, everyone that requests access to the data also agrees to be identified on the dbGAP homepage so that when you look at the dbGAP homepage for any given

study you can see every approved user, their institution and what their approved research use is for that data so that the public can also see what's being done with the data and how it's being used.

(Slide.)

The final two elements of the policy speak to issues of scientific publication and intellectual property. They were much more straightforward to write. They are not necessarily any less controversial.

For scientific publication, again the concept of this pre-publication broad access to data was something new for GWAS in terms of moving beyond the genomics community. And in order to respect the time and energy and intellectual contributions that these PIs will spend many times over decades to develop cohorts that they were now wanting to do GWAS on, there was a publication embargo period that was put on to the data so data was expected to be submitted as soon as quality control was complete, and be made available for investigators to begin analyzing but there was an agreement that only the PI and their direct

collaborators would be able to submit publications or any other form of public dissemination about their work for the first 12 months that the data was available.

(Slide.)

And this was implemented trying to highlight this embargo policy and the dates attached to different versions of datasets, again on the homepage for dbGAP.

(Slide.)

During those 12 months, however, anything else was appropriate to be done. So you could investigate it thoroughly, you could write your paper, you just could not submit your paper until after the 12 months had expired.

(Slide.)

In terms of intellectual property, we were a bit limited in terms of what we could do and wanted to stay within the bounds of existing NIH policies and respect the Bayh-Dole principles for this but there was a broad consensus, both internally as well as through some consultations that we did with external experts in the area, that the basic GWAS findings that were going

to come out of first round genome wide association studies really were pre-competitive and should remain in the public domain so that everyone would have freedom to develop around and innovate would have freedom to operate around and develop and innovate around those basic findings.

To try to substantiate that policy there are automated calculations around those statistical values of the genotype analysis that are made available in the database so that again everyone can have them and they're out in the public domain to try to substantiate the fact that patents shouldn't be filed on those first round findings.

(Slide.)

This has been further emphasized within the policy statements as well as within the data use certification where investigators acknowledged this intent for the NIH and this principle that the data remain in the public domain as well as their institutions going forward.

(Slide.)

Something else that was important and I think

has proved to be vital to GWAS management, and even within discussions in the community over how to go forward with genomic data sharing and biorepositories at this point, is a governance model.

(Slide.)

This model is both simple and complicated, depending on the level that you are working at, and I think that was part of the design and has been helpful. So at its core there's a senior oversight committee which reports directly to the NIH Director, and they make all of the policy decisions in terms of changes that might need to be made, as well as managing at the highest level how the policy is implemented across the NIH.

The committee is chaired by Dr. Green at NHGRI and includes other IC directors as well as senior staff from the NIH director’s office.

But for day-to-day issues, and to help make this a manageable task for the senior oversight committee, there are two steering committees which sit under the SOC and they are made up of senior staff and focus on two specific realms of issues. The technical

standards steering committee focuses on scientific and programmatic issues, as well as technical issues around dbGAP and security standards that would be important for that, and the participant protection and data management steering committee as constituted from the various data access committee chairs, as well as other experts at NIH in human subjects research protection and bioethics. And that is where really I think the core of the policy development and practices have developed as the DAC chairs try to learn how to do their jobs together and develop more of the framework for how NIH is going to do this across the board.

And will definitely inform and interact with the senior oversight committee as issues arise so that we have both leadership at the highest level making decisions, as well as those staff who are on the ground trying to implement the policy on a day-to-day basis informing what the decisions are.

I will stop there and just point to our GWAS website that is under review but, hopefully, will be a place where we can have some of this information and again increase transparency on what we are doing and

what the practices are going forward for everyone that needs to interact with the policy from the investigators to just members of the general public who hear that we have this database full of genomic data o thousands of individuals.

And I will stop.

DR. ROYAL: Thank you, Dr. Rodriguez.

Any questions? Marc?

DR. WILLIAMS: I may have just missed this but based on what you were talking about in terms of the oversight and that, it sounds like that if you are an investigate that wants to use existing GWAS data and you go through the data request and approval, and all that sort of stuff, then you're able—it sounds like—to download the data on to whatever your local resource is and use it under the terms of the agreement as opposed to the data residing within dbGAP or that database then being manipulated there by investigators as opposed to—where it really wouldn't move to a local type of server.

Clearly the advantage of having it centralized is that you can develop audits and can

automatically make sure that people are staying where they are supposed to be staying. But presumably there was a decision made as to why this model versus another model was used.

Could you comment a bit on that?

DR. RODRIGUEZ: There was a great deal of debate as to what model to use going forward and the final decision was made because the statistical geneticist and many of the people that would want to analyze the data would be writing their own programs, and so it was not something that could be done effectively within NCBI space.

And so instead the decision was made to create something that could be securely transmitted to the local site and to put agreements in place in terms of what security standards should be in place at that site for the data.

And actually the IT officials for the institution are now required—one of the required signatures, though it doesn’t actually get implemented that way, but then they are supposed to be aware of every request for access as well so that they’re

signing off again that they have the capacity to protect the data in the way that the investigator is agreeing to protect the data.

DR. : So just to follow up on that. Are there—obviously, you're requiring a report to come from the institution to say, yes, we have been behaving well, we're using the data the way we're supposed to, and here is the results of that. Is there any opportunity to—for NIH to audit or if people suspect that something has not been used the way it's supposed to, that you would have the ability to go in and say, "Could you show us exactly what you're doing?" You know, if—I guess it's sort of an IRS model, which was, yes, this is what you told me on your taxes but is that, in fact, really what your income was for this period.

DR. : So we also talked a good deal about setting up some type of audit program and the issue at that point—we looked at several different models and the cost was quite significant and questions of who would absorb that cost and what the return—the benefit of that return would be on instituting such a

policy was such that it was determined that we would not go with an audit model to start with unless, you know, we saw that we had problems and, in fact, so much of what NIH does operates on this assurance model with the organization and that we will trust that you will do what you had agreed to do and, if you don't, then there will be consequences.

DR. ROYAL: Gurvan?

DR. RANDHAWA: Yes. I'm interested in the clinical phenotyping of the subjects. Are there standards for that? You know, one of the problems we have are diseases that predominately affect one organ system but also affects something else and where the clinical phenotyping may be partial the cause of the observer who originally created the dataset. How are you addressing that issue?

DR. ROYAL: So dbGAP set itself up so that they could accept any measure and however it was reported to be open because there was such a variability across the measures.

DR. RANDHAWA: Right.

DR. ROYAL: What we have done is, again, by

putting the protocols on line for every study, you can see exactly how the blood pressure, for instance, was measured in one study and know whether that's going to be comparable to a blood measurement and another study.

And, in terms of going beyond that for standards development, NHGRI has a program, the Phoenix Program, which is looking at building some standards for phenotypic measures across the board but there are no requirements for that at this point within dbGAP for GWAS data.

DR. ROYAL: Jim?

DR. EVANS: I was just wondering how your deliberations, your model and all was affected by the Jacobs' *Nature Genetics* paper about inferring genotype and phenotype, and inclusion, and GWAS. Is that—

DR. RODRIGUEZ: So—

DR. EVANS: The fact that it's possible to analyze the aggregate data and infer phenotype.

DR. RODRIGUEZ: Right, the Nils Homer and David Craig paper from 2008 or Kevin Jacobs' had a paper recently.

DR. EVANS: Right, the subsequent one. Yes.

DR. RODRIGUEZ: So our policy hasn't been changed at all relative to Kevin's paper from this fall. We are having ongoing internal discussions about at what point do we re-address the situation and is there any level of data that might be possible to be made public that we haven’t yet come back to have the formal discussion with Kevin or his group that we’ve—you know, we’ve definitely looked at the papers and the groups.

DR. ROYAL: All right.

I have a question, Laura.

So you guys are in the process of changing or modifying the GWAS policies for sequencing data, right? You’re still in the process or you’ve done it.

DR. RODRIGUEZ: We are just starting to actually do that. We did some internal data collection to go out to our extramural program staff and get information on what policies already existed for sequence data, how they would describe a sequencing project that would or would not be subject to such a policy. The sequence projects are a lot more complex with GWAS but it’s pretty straightforward what it is

when you have it. And so we have that information now and we are beginning to look at the different policy scenarios that we might put together around that, as well as some of the technical issues because it's a lot harder to transmit all of that sequence data and decide when is an appropriate point to release that because the sequence data again comes in, in a very different format and different timeline than the GWAS data does.

DR. ROYAL: Do you have an idea in terms of when you might roll that out?

DR. RODRIGUEZ: Not officially. So we're working on it right now and we hope to have a draft ready for leadership to consider by the spring but I would never predict what the leadership will say about the draft.

DR. EVANS: Can I ask another question, and that is use of genetics and genomics for prediction necessitates having information over time. Have you planned for that in this database, that is, observational data that shows what happens?

DR. RODRIGUEZ: That was again another reason why having coded data was thought to be so useful,

because there can and there have been updates to different datasets. Framingham, for instance, has had several updates and there are different versions of the data that are available for the Framingham so that when there’s a large cohort that has another round of visits, and another round of data collection, you can go back in, and associate that with the data that you already had so it can be a dynamic resource.

DR. ROYAL: No more questions?

Thank you, Dr. Rodriguez.

Now we’ll hear from Dr. Hoffman from Cerner, who is going to talk about a commercial model.

Dr. Hoffman?

COMMERCIAL MODEL

DR. HOFFMAN: Just as there's no single academic or government model, there is no single commercial model for data sharing.

(Slide.)

My intent today is to provide a few examples of things that we are doing in the commercial electronic health record environment to set the scene for the more effective exchange of genomic data and

then some examples from other domains outside of genetics that I think will serve as relevant examples of future trends for how to facilitate data sharing.

(Slide.)

We will begin with a comment that might seem out of center field at first and then, hopefully, when I come back to it you will see why I am saying it. There are more virtual farmers in Facebook Farmville than there are real farmers in the United States. So I'm just going to leave that out there and then come back to it. It is a little bit provocative though.

(Slide.)

The topics that I want to really hit on in my conversation today is that, first of all, how can we generate high quality data during patient care to facilitate both data sharing and decision support?

Then, secondly, we will talk about a few examples of what I call the data sharing ecosystem. There is no single model for how you go implement data sharing and there's actually strengths and weaknesses to a couple of the models that are out there. So I'm going to share a couple of efforts in each of those.

(Slide.)

Within Cerner we are working on both sides of this puzzle. We're working on the patient care provider side in terms of how can we enable genetic testing laboratories to capture data discreetly but then we're also working to facilitate research using our deep knowledge and understanding of clinical processes and of clinical data architecture.

(Slide.)

So to summarize at a very high level some of the key attributes of the electronic health record, and I should also point out there is no single electronic health record. There are multiple implementations. You can go to some of the organizations that are very prominently represented that have homegrown EMRs, then there are the multiple commercial electronic health records, each of which were designed around different principles but, I think, most would agree that capturing information during clinical processes is fundamental. Simplifying data retrieval, queries and analysis is a key goal of moving to electronic health records. Automating processes so you reduce the

opportunities for error, providing decision support capabilities, create efficiencies, and generating a body of data that can then be analyzed, whether for administrative, operational, clinical or scientific insights.

(Slide.)

There's often some blurring between the electronic health record or electronic medical record, and the personal health record, which to me, the medical record is a legally binding system. So if a physician is part of a malpractice suit, the assumption is that there will be high-quality data in the system that can be extracted and then utilized in the discovery process. Whereas, the personal health record, there's probably quite a bit more blurriness around the obligation there. There's often the expectation that the two should be one. I think, likewise, the expectation that you can have one system that’s an EMR or PHR and a research system is something that needs to be scrutinized much more carefully. It's not a perspective that we try to promote. We do think there should believe there should be fire-walling

between the systems.

(Slide.)

In informatics, I think, sometimes we want people to think that you couldn't do research in this model where information is stored on paper but, the fact of the matter is, that there is still a large amount of research that’s done through manual chart abstraction, and the privacy issues there are very similar to those in the electronic world where you have--in many ways they are even more challenging because you have human beings pulling the paper charts out, they see the names, and then re-enter that information into other systems.

What we are trying to move towards is to a fully automated digital system where clinical information and eventually genetic information is stored discreetly in a mineable fashion.

(Slide.)

A couple of the other presentations have referred to standardized vocabularies and ontologies. We have been very active in developing and deploying what we call the clinical bioinformatics ontology.

This is a vocabulary that’s available through an open content model and can be downloaded, and creates standardized concepts that can be used to codify findings whether for molecular diagnostics or cytogenetics or other testing methodologies.

(Slide.)

Just to give one example of probably the orphan topic in genetics discussions, and that’s cytogenetics. If there's anything that would put fear into the heart of an informaticist I think it would be a karyotype. And we have actually put quite a bit of effort into making the karyotype a mineable resource because let's say that you are interested in a condition that’s tied to band 21.2 in this example. If you were to do a purely text-based mining of that karyotype you would never find this patient's result. We drop out discreet concepts into the database from that, one of which is a concept related to 21.2 because the beginning and the end positions of the abnormality documented here, and that creates a mineable resource, whether for research or decision support.

(Slide.)

So these are just some very high-level examples to show that within diagnostic labs we are working towards systems that create that granular body of information so that as you get into data sharing your data is ready from the point of capture and you don't have to re-enter into another system.

(Slide.)

The second theme that I want to cover is some representative data sharing models. One model that’s very familiar is what I would call the centralized data warehouse model. Increasingly, things are moving towards distributed models. From the electronic health record supplier perspective, we have a common architecture that’s in use at thousands of healthcare delivery facilities and believe that architecture alone positions things to be used creatively for collaborative work and so I will share an example of a project based data warehouse

(Slide.

We also have brought in technology to provide a consent-based—a web-based consent-driven system, and I will show that briefly, and then I will return to my

social media comment.

(Slide.)

So the data architecture that--there's a couple of options embedded within this picture. Within an organization, the clinical care data is embedded into a database or the EMR, and that information can then be used by the physician or the CIO or CFO to make observations about how they are running the organization and so forth.

It's also very feasible to migrate that information into a larger meta-data warehouse, and usually that process involves scrubbing the data of all HIPAA regulated identifiers.

It's also normalized, so a key part of any data merging activity, especially among non-affiliated organizations, is mapping to a common vocabulary. And so that is a key part of what many aggregate data warehouses offer.

(Slide.)

Moving to distributed models. The—so if I—if I just summarized the data warehouse model, the distributed model is that instead of pulling data in,

you push queries out to the user—to the sites. So in IT systems we think of operations jobs that run at midnight. So the impact is minimized. And those will be routine processing but there can also be queries that evaluate the data within that site. And then summaries of the findings of those queries, instead of the actual body of data, can be sent to the organization that’s managing the distributed project.

(Slide.)

So we at Cerner are deploying what we are calling our research network where throughout our client base we can push packets of queries. So if you are interested in cystic fibrosis patients, we can—and you work with us to sponsor a project and push these queries, the data remains at the local site. We don't really want the data from these types of initiatives in our hands under this model.

And then the--I really don't like to make the comparison but helps it click. The analogy that resonates is Matchmaker.com where we see our role as matching a trial sponsor to sites that have a candidate group of patients, and that that has value to the

process so that if you are looking for trial candidates you are not mining in territory where you are never likely to actually find candidates.

(Slide.)

Then more recently in the public health domain we have taken that model a step further and worked with the CDC, state and local health departments, on an influenza surveillance initiative where we reached out to the entire client base and said that we’ve got to work with you. You will get a daily view that’s updated every day showing how your organization compares to your state and national peers in terms of positive flu results, influenza-like indicators, and so forth.

We, in three months, rolled this out to 780 facilities throughout the U.S. so it’s present in almost every state. We had, I think, 23 million records that have passed through the system for surveillance. The CDC gets updated information every day with state and local stakeholders. It provides GIS level mapping and trending. It's very feasible to use this push model in a very rapid approach and I think

that a commercial company has the agility to do this type of thing very quickly.

(Slide.)

As an adjunct derived benefit, there’s also a lot of healthcare information that comes out of this so one of the parameters that’s tracked is emergency department utilization. So I don't know if there’s—I think there are some people from Tennessee here but if we compare Tennessee to the national norm of emergency department utilization, every day we can see that, even on week days, 50 percent of healthcare is delivered in the emergency department in Tennessee. So there's a lot of insights that can be gained from this.

(Slide.)

I’ll also mention that we recognize that prospective research is an important model. I think you will be hearing about a consumer approach in the second talk but our stance is that we want to let scientists do the science and provide the enabling technology so that they can get to the science as quickly as possible. There is a company called First Genetic Trust, so we brought in source code or patents

to the technology, and it enables patient controlled disclosure of genetic information through this model—through this web-based model.

(Slide.)

The second to the last slide just is an example that we participated in that pulls many of these topics together. Cerner does the data and project management for CDC for their HIV outpatient study, which is a prospective study. Patients are consented and enrolled and then tracked longitudinally. When the study was launched they had the insight to capture the HIV genotype data, as well as the prescriptions and the lab data.

So one of our questions as we looked at personalized medicine, how--when armed with genetic data, how well are physicians utilizing that information?

(Slide.)

And so we did an analysis of the data and found using just one scenario that--using antiviral resistance that if you mine the data as an analogy of physician behavior, w found that for the patients who

were found to have—the 441 patients with the resistant HIV genotype, 59 had contraindicated genotype therapy initiated six months after that result was determined. So I think that's evidence of the need for decision support.

(Slide.)

So I promised that I would come back to the Farmville comment. The—I think data sharing in the social media world is really completely, maybe it doesn't--I haven't heard it on the Table yet but if any of you are in Facebook and have used a single Facebook application, I actually found one from the NIH, when you sign up for a Facebook application, you are giving data from your profile and your friend's profile to anybody, to the organization launching that application.

So if you are signing up for Farmville and your friends are in the retinoblastoma perineal support group, you are sharing their status, their status with that organizer.

So I think that often technology quickly gets ahead of policy. I think one of the things GINA has

going for it is instead of defining the technology, it defines how we protect the patients from harm. And so I think that should be some consideration as we think through how rapidly evolving these various models are.

(Slide.)

So I think that the aggregate data warehouse has both strengths and challenges in terms of there are some highlights—you can’t—if you haven't pulled the data, you can't go back and add it later so you have to either pull a lot of data or sacrifice on data quality.

Distributed models are much more agile but they involve a much more limited amount of data.

Social media, I think, as yet untouched, but again it’s getting way ahead of things but things are moving faster there than anywhere else.

And in my opinion the role of healthcare information technology is to serve as enabler and to help any of the stakeholders in the process.

So, with that, I will stop and address any questions.

DR. ROYAL: Thank you.

Any questions from anyone? No one?

Okay. Thank you, Dr. Hoffman.

Now we will have our last speaker, who is Mr. Shelton, Robert Shelton, from Private Access, talking about the consumer-controlled--a consumer-controlled model.

CONSUMER-CONTROLLED MODEL

MR. SHELTON: First I want to thank the committee for inviting me to make this presentation and Symma and the staff for making it possible to be here.

(Slide.)

So when I received the topic I thought maybe the best thing to do was to modify the first slide so I would like to submit that instead of thinking what I'm going to talk about as consumer-controlled, I'd like to talk about it as being consumer-empowered. And in the context of this committee, I would like to think about consumer-empowerment not as just empowering the consumer but, also, the consumer empowering the researcher. And so a lot of what you're going to see in this presentation is about ways that the consumer properly empowered can, in fact, empower the researcher to go a lot further than the researcher is otherwise

able to do.

(Slide.)

So I think I decided, as I was sitting in the audience, that I am perhaps the only person who doesn't have an M.D. or Ph.D. behind their name so I thought I'd start with talking about why I'm here, what's Private Access.

And, number one, I start as the parent of a prenatally diagnosed child with a rare genetic condition.

(Slide.)

This is a picture of him when he was about four-and-a-half years old. I selected that for a specific reason which I'll get to in a second. He was diagnosed with 47 XXY, which is a proclivity towards Klinefelter's Syndrome. It's a one in 600 incidence in live birth and roughly 75 percent of the people who have this diagnosis are never diagnosed from birth to death. And so one would assume, as in his case, that it's a pretty mild condition. According to the most recent statistics, a 2007 study done in California, roughly 70 percent of the parents who receive a

prenatal diagnosis of Klinefelter’s Syndrome will terminate that pregnancy in utero.

So when this committee thinks--I added this slide set as I was sitting in the audience because I would really like to bring the individual perception and the perspective of individual patients to this committee, and say that the kinds of subjects that we are talking about in macro in millions, and tens of millions of people, really boil down to individuals and parents make bad decisions based on limited and fear and lots of things that you all know very well.

(Slide.)

So that led me to basically taking off from work for—my day job—for about three years to become first the direct and then chairman of the board of a national disease organization that supports Klinefelter’s Syndrome and also Trisomy X and XYY.

And the reason I selected this particular image is because the organization had been in existence for 15 years. In 15 years we had never had a picture of a person with Kleinfelter's Syndrome on the website because a lot of the people with the condition are

afraid of being recognized as having the condition, and so there's a tremendous privacy concern among that population. It's not one of the protected populations in many state laws but it has very high privacy concern, in part, because it's so mild and, in part, because there's significant stigmas.

(Slide.)

So this picture actually now appears on our website because someone called up our organization and said, you know, "I have decided I am going to terminate the pregnancy because I assume since you go to the Down syndrome site you see pictures of Down Syndrome children, you go to the Klinefelter's Syndrome site and you don't see any pictures of people with the condition." So I decided to go ahead and make this picture available so that people would actually connect with a person who has this condition. So information is power.

(Slide.)

The other thing that has happened is you will see that I am also an entrepreneur and have founded a privacy technology company called Private Access. When

we got started we focused the technology in Private Access on serving some of the needs that we recognized through the disease advocacy area. Today, with roughly $5 million invested, a few hundred--almost 500,000 lines of code, to make possible what I'm going to show you.

So we also think of the world through partnerships. No one would recognize our company unless they have heard us at a conference or met us person to person but we partner with organizations that are already trusted organizations. So we think of trust as being an extension from human being to human being; not based on technology but based on human relationships. So partnerships are really vital to us.

(Slide.)

Our mission is that we focus on creating an environment of trust. And we talked about how we do that. So if I started this slide over in the upper right hand corner, I remember going to a conference at the Health 2.0 conference three years ago, Esther Dyson was speaking and she referred to privacy as the giant hairball that was clogging the drain for data

liquidity, and that we need to blow that hairball apart.

So, privacy can be viewed as a speed bump that is keeping data apart and hurting liquidity. It can also be viewed as something that through technology is an achievable goal that actually will help enable health information sharing. And so I think that what we are really talking about is how to create an environment of trust because inside of an environment of trust you get speed and there's books that are written on this topic and borrowing the fast company quote, the new economy begins with technology, it ends with trust. So we have to build trust in the system.

(Slide.)

So our particular way that we focus on this is we’re looking at creating what we call the perfect balance between privacy and access or accessibility to information. So you see one patient on the side that has got the ability to leverage their words to a consortium of people that are really out to help them to achieve their health goals.

(Slide.)

And the need for speed is something that, as a disease advocate and coming from that perspective, is something that is just, you know, really critical. When you look at the internet, and I presented at the Electronic Patient Record Conference, TPR, earlier this year or mid-last year actually, and I used an example, and actually did it live of searching for a person based upon attributes. It was a public person through Google and in a minute and 27 seconds, without knowing the person's name, just knowing some things about them, we found the person. We located how to get in contact with them and we actually booked them for a speaking engagement. So in a minute and a half in Google we can locate people.

In Match.com, which the previous speaker mentioned, or Monster.com we can do the same thing and so I went through an example, and in under three minutes I was able to locate a person that matched the demographics, the location, the characteristics that I was interested in finding for either a date or for a job.

(Slide.)

So, in healthcare, however, we have got challenges. We take six months to a year to recruit people for clinical trials. We have terrible accrual rates in trials. We also take an average of 15 years to develop drugs for diseases. Time is actually more critical in the healthcare area than it is in those other two domains, and yet we have the worst ability to move things quickly in healthcare. And I would like to submit that part the of the reason is because of the trust factor we need to replace.

(Slide.)

So there was a study done by Case Western Reserve focusing on dried blood spots from newborn screening. The question that was posed was how willing are you to have your child's blood spot sample used for newborn screening for future research studies; and it was done with permission and without permission. And the choices were simple. It was very willing, somewhat wiling, somewhat unwilling and very unwilling.

Over 75 percent of the people that were asked would they give permission for this granted permission to their information being used for research. But when

you change the equation to denying them permission, what happens is all of those positives around granting permission changed dramatically and the opposition to using the information for research increases dramatically. So to me that's what happens if consumers are not asked about sharing their data. And so a lot of the technology we have proposed is set up to focus on that.

(Slide.)

That study is not discrepant with the secondary literature. In fact, it is very consistent with the secondary literature and I am just going to slide through these because you all probably know this data but the one that is the most compelling to me is from the Institute of Medicine study at the end that 57 percent of people would permit their personal health information to be used for research only if various privacy conditions are met and 38 percent of the total, which is the largest share of the 57, want to get information and notice on a case-by-case consent basis.

(Slide.)

So how does that happen? So if we think of

the world as data seekers and data holders, a data seeker can get in contact with a data holder, a data seeker can also put out a query for data around who has got my data, who has got data that I would be interested in, and a data holder could raise their hand and say, “Hey, I’ve got some information you’re interested in.” And if the conditions are right—if the terms are right, I’m willing to share it with you.

The challenge is that for that data holder to act quickly that data holder needs to know do I have the right to share this data with that seeker, that particular seeker? And that entails a determination of what's permissible under federal law, what’s permissible under state laws, what’s permissible under the institution’s policies that the data holder is encumbered by? Are there any special considerations that are entailed for this particular record? What would my patient think? So there are legal and reputational risks that are entailed in that and those answers—particularly where they answer a search query like a Google or a Match.com, particularly for that level and speed, those answers have to be answered

fast. They have to be answered reliably and they have to e answered containing the information about what that record-holder will be compensated for the information, and so even if the compensation is in millicents or pursuant to some sort of a contractual relationship between them.

(Slide.)

And so what are doing in Private Access is replacing those questions with an automated transaction-based system that is programmed with an ontology of privacy that looks at each of those issues, the institutional law, the federal law, the state laws, and the personal privacy preferences expressed by the individuals to give that record holder, that data holder, back that information in under a second or two. So that would allow them to know red light, green light, yellow light, do I have the right to move that data to that particular seeker who is looking for it.

(Slide.)

And then in order to power that--remember the title, consumer empowered--we look at tying the patient in through the ability to dynamically consent or

decline access to the sharing, the proposed sharing, if their voice is permitted under the prevailing law of their state or federal law.

(Slide.)

We also allow them at any time to view the audit trail associated with the data sharing activity. So the data then can be pushed. It doesn't have to be electronically. It could be pushed in the form of a Fed Ex pouch. It could be pushed by U.S. mail. It doesn't have to be electronic but what is conveyed back to the system is an audit trail of the actions taken and any dollars, any amounts of money that are charged between the data holder and the data seeker.

(Slide.)

So that little fundamental architecture is what we have spent all the money building and time developing over the last three years. So to date we have focused our solutions directed to registries and biobanks, and to allow all or selected parts of the confidential personal information to be moved based upon the particular needs or interest of the patient.

So our first focus is to set up a consumer-

centric site, which in most cases we have co-branded with the trusted intermediary, so with the disease organization as being a co-branded indication. So in each case thus far we're working with a trusted intermediary.

The second thing is we use a system of trusted guides to help the individuals set their privacy preferences. If we get down in to the level of granularity that really creates this ability for speed and accessibility, it's incredibly granular. That means someone has got to do a lot of reading and clicking of on and off buttons.

The patients that I am familiar with do not have either the patience and in many cases the proclivity to actually spend that time. And so the way that we focused on it is to set up a spectrum of people's perspectives. We call them Trusted Guides and we select three at least in each case. Those guides reflect a perspective of the spectrum from I am in favor of a lot of sharing of data to I am in favor of very little sharing of data. I am very privacy concerned. I'm very accessibility oriented.

And then we ask each of those guides to pretend that they are talking to a person across the table from them who says, “You know, I’ve got high privacy concerns. What would you tell me about what I should consider doing?”

Or “I have low privacy concerns. What would you tell me.”

So from that we get a broad spectrum of perspectives on the issues of what should the data—what should the subtitles be and we boil those down to—in effect, permitting access, permitting access on a de-identified basis, permitting access on a pseudo-anonymized basis, permitting access with a prior consent, permitting access with a dynamic consent, and so we set a series of stops along the way for each of those perspectives on each of the factors involved.

hose down to in effect, permitting access, in a limited basis, a prior consent, permitting access with a dynamic consent. We set a series of stops along the way for each of those perspectives on each of the factors involved.

And then we, as I said before, have a

comprehensive audit log for each access to the data and when the IHE standards are adopted in HTSB and included the minimally—in the standards required for C-chip certification, hopefully, those will permit the audit trail to touch any EHR, any PHR that is standards compliant so that the patient can go to one place and see the accessibility to their data.

And then the last piece here on this particular element is we--identity verification is vital. We have identity verification up front in the system. We have written this privacy directed language, which is a robust ontology. We have the dynamic consent management, the audit tracking, and then we've integrated the commerce features.

(Slide.)

The initial applications that we have built are focused on clinical research. So we are using these to help people locate--help researchers locate patients for clinical trials who wish to be found. So we call this application that we have built a recruit source and it's based upon a researcher-centric site where a research can go in and enter a natural language

inquiry that is searched based upon either text match or based upon UMLS language for their particular query. So if they're looking for Tylenol they would find acetaminophen hit in the database.

That then results in either—depending upon the privacy preferences, a fully anonymized or a fully personal identified record for them to see and they can search based upon the demographics or the locations of the patient. They can say I went 10-miles within a radius of a specific spot where I've got a research cohort that I'm trying to put together. And if the patient has said I want to be in a de-identified forum, if the HIPAA de-identification rule says there needs to be less than 50,000 people within that radius and we have less than 50,000, according to SMSA data, we can't show them that data. So we use that switch to turn that off in accordance with the federal laws.

(Slide.)

Finally, we have the dynamic consent tools that are built in from privacy layer so that if a researcher says, you know, "I saw you in a de-identified forum. I saw you in an anonymous form. I'm

interested in you and you've indicated that you don't want me to know who you are, you don't want me to know your address until you—until I tell you dress, until I tell you about my research project." Then the researcher can push that information through the switch back to the consumer who can, in turn, decide to push the blue button to permit the data for their contact information to be sent back to the researcher; push the red button to say, "No, I read about it. It's not something I want to do and I talked to my doctor, and we've decided this is not something I want to do"; or push the yellow button in order to snooze and say, "I'm going to wait for this answer for a while."

(Slide.)

The first project replied to was on the organization that I chair, Klinefelter Syndrome and Associates, which is renamed now Support in Action because people didn't want Klinefelter's Syndrome on their return envelopes. And so we have looked at 1,200 patients with five researchers. The persons who have actually completed a survey, 90 percent have indicated that the system is easy to use and they like the

experience; 75 percent have indicated they would recommend it to family and friends, and our experience would be that partnering with that trusted source was overwhelmingly what drove the patients to have an interest.

(Slide.)

And so we have a number of research projects under way. One of them that is in the packet of materials that I believe you have been given is a project that we're doing with the University of Michigan focused on the newborn screening blood spots. It was a challenge grant award for 200 or so of the challenge grants that were awarded. It is presently ongoing and is looking at facilitating a state sponsored population birth cohort to use the information for genetics testing, and looking at the use of consent for that purpose.

(Slide.)

This is the steps of the process. We are presently at the stage of creating systems and environments. And early in the summer we will begin with the pilots and recruitment for that study. So we

are very early in the study but excited about it because all of our prior work has come through working with disease organizations, and this is actually a general population as opposed to a specific disease organization.

(Slide.)

We are pleased that we have strong support from some industry stakeholders. At the end of last year we announced a collaboration with Pfizer and Greg Simon, the senior vice-president of Worldwide Policy was quoted in the release announcing it and saying that patients are the most important stakeholders in medical research. By merging respect for their privacy and access to relevant actionable medical information, we are giving patients more control over their destinies. And this collaboration has the potential to accelerate medical progress by putting patients' needs front and center.

(Slide.)

This was echoed in December by the CEO of Pfizer who in front of 650 people at the Partnering for Cures Conference said when he was answering Mike

Milliken about what he was excited about in terms of accelerating treatments for patients to come through their organization, he said, “Focusing on patient privacy and a technology that would accelerate the ability for us to get in touch with patients who want us to get in touch with them.” So I think that we’re finding support for this.

(Slide.)

And, as I said earlier, we focused on collaborations. Our first round, the one that we’re perhaps most proud of is with Genetic Alliance, who is—we’re in a public-private partnership with and advisory to University of Michigan. We have a number of other projects that have not yet been announced yet with some significant disease organizations working with a couple of government agencies on using this for their informatics grid type computing. And so--and several HIE, Health Information Exchange, Regional Health Information Organizations for their applications, and then we were pleased to be on two sharp proposals for the recently announced IT initiative from the Office of National Coordinator where we—one is with Harvard-MIT

for—in talking about de-identification of data earlier. It’s the Tanya Sweeney’s proposal. And the other is with C-DISC, which is the clinical Data Interchange Standards Coalition, where we’re on their proposal as well.

So, hopefully, we can maybe in a year from now come back and give you a lot more data on how this work.

DR. ROYAL: Thank you very much, Robert.

MR. SHELTON: Thank you.

DR. ROYAL: And thanks for sharing your personal story. It helps remind us why we are here.

Any questions?

Paul?

COMMITTEE DISCUSSION

DR. BILLINGS: So I understand your kind of roll out is limited and maybe you don't have any data on this yet but do you have a sense of when you put your processes in place--do you have more uptake? I mean there’s—I know how you do the comparison but do you have a sense you are going to foster more research participation?

MR. SHELTON: Absolutely. There's no question. There’s a lot of secondary literature on this. The Harris Western Poll has been done for a decade. The most recent was done for the Institute of Medicine in 2009. Alan Weston happens to be an advisory board member of our organization so we’ve seen his polling results and the raw data. And his polling results mirror pretty closely what we are finding in the disease advocacy world.

DR. BILLINGS: He's a character in privacy, I believe.

MR. SHELTON: He’s certainly the dean of it. So I think he wrote the definitive textbook in 1967 on the subject, privacy in a free society.

Thanks, you all.

COMMITTEE DISCUSSION OF NEXT STEPS

DR. ROYAL: Thank you.

We are going to open up for discussion now, general discussion about all the topics.

I am just going to put up a couple of slides.

(Slide.)

Yes, let’s go to the next one.

(Slide.)

So we heard a lot of information today, a lot of interesting information, and just to really move our discussion along I just wanted us to go back to these two questions.

The first one—I think I'd change that to what have we learned, what are the lessons learned?

And then the second question is are there issues that warrant further policy considerations, and issues that SACGHS might consider?

So we may want to talk a bit about what we've learned, what we heard, what stood out for us in terms of all those presentations, those different models that we heard.

CHAIRMAN TEUTSCH: So, Charmaine, let's see if we can flesh this out. So we've heard a lot of presentations. We know that we're trying to aim towards best practices. Have we heard some best practices? Things that we can say, gee, they are already underway, we really don't have a further role. Or are there some gaps here, some issues that could really benefit from what we're—from this committee

actually weighing in. So we've heard several models, right?

DR. ROYAL: Right.

CHAIRMAN TEUTSCH: And I think we saw them—you know, none of them are very old but some of them are—you know, were very—have been thought out but they're in place. So where are we on this trajectory and where are the gaps where we might weigh in? Or do we say these groups are doing a great job, we should move on?

So I think if we can talk about what we heard and, as Charmaine said, what is already there and working well, and what are those needs going forward and, if there are those needs, is there a role for us to weigh in?

DR. : (Not at microphone.)

CHAIRMAN TEUTSCH: Yes, you can. Yes, go ahead.

DR. : No, go ahead.

DR. : She was asking—

CHAIRMAN TEUTSCH: No, she didn't—

DR. : No, you can go ahead.

DR. : But you need a mike.

CHAIRMAN TEUTSCH: You need mike. I'm not sure if it's on.

DR. : Thanks.

DR. : (Not at microphone).

DR. : So can we have all the speakers come up front? Sorry about that.

DR. : I will just finish my question. So I am sort of struck by the fact that people like me, who are trying to develop these large resources, are also very focused--part of what makes them valuable is that they are epidemiologically sophisticated, or epidemiologically known sort of populations. So there are a lot of issues about, possibly, even though obviously everybody involved is a volunteer in one sense, nevertheless, there are issues about being able to conduct research just with people who happen to see the information and volunteer and so forth. And yet—which is sort of the—in some ways the private access—more the private access model and yet the situation that I confront is one where I—it's very expensive for me to update information about what

people in my cohort think about different research projects that we might do or about different new emerging issues regarding privacy and confidentiality that come up.

So I guess what I am trying to get to here is it would be very helpful if within the context of these sort of epidemiologically defined populations we had technology that allowed us to have the kind of communication that in some ways your system envisions would be possible between researchers and participants.

Right now that's prohibitive for us.

DR. HOFFMAN: So when I mentioned a couple of government agencies that we're working with, it's based upon exactly that concern. When you were speaking earlier you talked about your definition of de-identification that you used to establish the policy. Well, one thing I know about politics is politics is 50 percent plus one vote equals a change of policy.

What happens if the change of policy on de-identification is different than you set up the database? Is the database dead? Does the database have to be reconsented? So there's tremendous

challenges that I think cause the need for de-aggregating the switch from the store of data.

I think that the metaphor I would use is in a baseball game you have an umpire. The umpire does not care about the score of the game. The umpire does not care about what the batting average of the batter is, does not care about the pitcher is on a no-hitters. The umpire is calling the balls and strikes, and that’s all their job is.

There's a role for that in this system, I believe, and I don't think it ends up being one company. I think it ends up—we’re in a capital society. I think it ends up being multiple entities that play that role, and the ones that play it best will end up prevailing.

Our ambition is to be one of those parties playing that role. And the reason we focused on this around the federal, state and institutional policies in addition to patient-privacy preference, is there’s a tremendous thicket that exists by virtue of these policies that is mind-numbingly complex. And so in order to address that, we—there’s two things that are

necessary. One is somehow or another develop an ontology for processing it. I think we've got a long way towards that but most of the processing—I had—in discussion last night with a top privacy advocacy. I said, Most of the private—most of the resolution, adjudication would yield amber lights, not green, not red, somewhere in between and then you'd say, you know, it's just ambiguous. The law is ambiguous here.

So the architecture we are putting in place actually pushes the ambiguity back to the policy makers, back to the legislature if that's where the issue exists,; back to the institution if that's where the instrument exists, with the question. We're the umpire. We're not trying to set the policy. What we are trying to do is to say there's a challenge and process in these policies that presently address from a robust ontology.

You as the policy maker need to decide do you believe that the law should be permissive? Do you believe the law should be preclusive? Are there rules associated with that based upon various conditions? We can model all of those but once you make that decision

then adopt it, and if it changes next legislature, tine, then the rules will change the next legislation, and how the system processes. So I think that there is a way to play, whether it's private access or it's Sony that does it, I think there's a role to play for that type of technology inside the architecture that reflects not just patients but also each of the persons that are stakeholders in this in the moving of data.

DR. : So one of the things I heard from many speakers were a number of very elegant solutions to some of these vexing problems.

And the other thing that I was struck by was how under some of these efforts things are going to be put out into the public domain where things could be used for others. I was particularly struck by the idea that you could quantify the level of de-identification in a robust and repeatable way.

I mean, I can think about those of us that deal with IRBs and this issues of the privacy and de-identification. If we could—you know, there's something about a repeatable quantitative way of assessing research projects that are coming over and

over that I think would be highly attractive. I can’t imagine that people wouldn’t be interested in that.

So the point of that is that clearly there are innovative people that are coming up with innovative solutions to these issues, and some of them are being put out into a public space.

Is there a role for DHHS, the Secretary, to aggregate some of these solutions in a space that people could access and use, and try them out, and contribute their data. I mean, in some ways it resembles what's being done on the GWAS side, which is we think the data is important data and we think that people can do interesting things with it, and we’re going to recontribute it and we keep learning new things. That to me was the most important thing that I took out of the discussion.

DR. : I have a question.

People aren't born 55 years old and I know you've thought a lot about the developmental precursors of many of the diseases you are concerned about. How are you going to integrate the developmental aspects of etiologic cascades into your analyses over time given

that you are starting with older, consenting patients? How do you see the developmental processes entering into the database as it grows and matures? Like adding children, reproductive outcomes, maternal histories, things like that that could be attached to the progression or the emergence of disease as people grow and develop.

DR. ROYAL: That's for Catherine, for Dr. Schaefer?

DR. : Primarily Dr. Schaefer.

DR. SCHAEFER: Well, we are actually beginning to build a pregnancy cohort now. I have been very fortunate in working with an established cohort, the Child Health and Development Study Cohort of some close to 20,000 live births that occurred in 1959 to 1967, who were part of an NIH funded study then where they had the foresight to store maternal serum samples from each trimester of pregnancy that are still available for analysis now.

We have used those samples in follow-up studies of schizophrenia, for example, to investigate the role of maternal exposure to influenza and other

viruses and show that maternal exposure to influenza, for example, in the first half of pregnancy is associated with a significantly increased risk of schizophrenia in the offspring.

So just so the--I am definitely a believer in the importance of a better understanding of developmental contributions to adult diseases and have benefited directly--I have seen very directly the importance of this. The prenatal period may be extremely important and unusual, unique in terms of the interaction of genes and environment and playing a role in adult health. And so if we neglect that, we really won't have a full picture of the origins of adult health.

So I don't know that I have a complete answer to your question. We would like to be able to add children. We didn’t originally. In part, there was a resource issue. We had actually a very limited amount of money to initially survey our population. And in part it was the complications of asking parents to consent for children about being part of a very long-term study but we are now initiating a pregnancy cohort

where we will get samples during pregnancy and then intend to follow the children born of those pregnancies, and enroll them, and get samples from them as well.

Sorry, that's about as well as I can do right now.

DR. FERREIRA-GONZALEZ: Initially it caught my attention about the two speakers, the Kaiser Permanente and the Vanderbilt, is the different ways to do the informed consent to recruit patents in these. Your system mails an eight page document and I guess a spit cup for them to send their DNA back. And the Vanderbilt system is an opt out system.

I’m just wondering if at the NIH level or the federal government development or starting to develop best practices of what actually constitute informed consent for these type of studies and where there is all these de-identification and re-identification of data, do the individuals fully understand what they are actually consenting to?

DR. RODRIGUEZ: I think that was for me.

So we are not to the point, I don’t think, in

the evolution of the discussion around this point in terms of what is appropriate informed consent for anything, let alone for genomics where we're talking about de-identified to understand exactly what the right way is to do that because concepts are changing so much around sharing of our data and what we're doing.

So, I mean, we're certainly trying to think about it for genomics. We have a resource on our website which has some suggested language which actually—we have some example language. It's not even suggested language. Where people have IRB approved consent forms for different types of genomic studies and we are just trying to make them available at this for other people to see how it works. And we are hoping that will be a dynamic resource that we can get comments on and that we can build from to have ongoing conversations with the community but at this point I am not sure that we have a best practices, and I know it's something the community wants a lot because they would like to know what the right way is to do it but I don't know that there is one right way to do it because it

will always vary for the patient and subject population that you are talking to.

It's hard to quantify the risks right now so it’s hard to know what to tell them.

DR. FERREIRA-GONZALEZ: But I think that's concerning because we are collecting or recruiting patients that we don’t really know.

DR. : I’d like to put Laura on the spot because you looked at it from a variety of federal agencies and in the beginning Charmaine put up sort of different things that we are concerned about. Informed consent, storage, access and secondary uses, privacy, confidentiality, re-identification, handling sensitive data, the whole ball of wax and then how it all fits into EHRs.

And I guess the question we have is to what extent do you think we are already close to having, you know, good practices, if not best practices, and where do you see the real gaps and where do you think a committee like this could help shape the--what are the key issues that we could actually help in moving the field forward on or if there aren't any that would be

fine, too.

DR. RODRIGUEZ: I mean that’s not fair to do. So part of the first question is you asked if we were close to good practices and my response is on which one of those many issues that you mentioned because I don’t—

DR. : Why don’t you--tell us which ones you think we are close on, which ones you think that are—where we still have a lot of work to do.

DR. RODRIGUEZ: I think that there’s work to do on all of them because I think that the technology is still moving and we’re still trying to understand what participants think about all of this. I mean there’s private access and we don’t have research data to know what is the risk tolerance for people or do they care if we are sharing their data this way. They may not think that it’s a problem. If they’re truly altruistic about wanting to contribute data then we may be worrying a lot about how we are managing the data and how we are sharing it, and it’s not a concern to them. So there are some studies going on right now to try to collect some of that information.

But I think that, you know, informed consent needs—it needs some work. It needs some hard data to understand what people understand in consent and what's the best way to communicate risk to them. But that's at a much more granular level than what this committee can do. So it's hard to say. I'm not--things are moving in such an amorphous direction at this point, I'm not sure if I could really come together and say this is the one thing that this committee should definitely do because I think, again, that there's value in having different approaches go forward and learning from them.

So I'm not sure I have an answer to your question.

DR. : Is there a role then to sort of look at the different approaches to, as Charmaine said earlier, identify what these different elements—what's sort of the different options and which ones look most promising?

DR. RODRIGUEZ: I think there's some of that. One thing that could be useful would be to try and articulate principles. That could be useful to the

group, and trying any of the different solutions. If there is some common principles that should be present in any different—any model that went forward. That would be something that I think is at a high-enough level and that will be common and important in terms of providing leadership to the field as they continue to try and experiment and modulate what is going forward now.

DR. : I think one of the issues, and Robert actually started to bring it up but you didn’t bring up the entire context, is that sharing patient data is very different depending on what you have been diagnosed with. So, you know, I’m a long-term cancer survivor but I had non-Hodgkins lymphoma. That doesn’t really have a stigma. Breast cancer had a stigma. All cancers had a stigma. Your son's disease still had a stigma attached to it. So you can't really say that this is the way patients feel about sharing their data.

I don’t—I mean people know because I am a cancer advocate that I’m a cancer survivor, but it is—the disease is de-stigmatized. So I think that it’s

very difficult to make any kind of blanket statements about privacy and sharing of anything when you're talking about a very large universe of diseases.

MR. SHELTON: I couldn't agree more and I would add another vector to that, which is the stage of the disease. It's not just the disease, it's also the disease state. And you are trying to say something so let you talk.

DR. : Sorry, but I—yes, that's absolutely true. I think that—I mean I work with more in the cancer field but that's absolutely true. And I think the other issue is that people are often less concerned about themselves than they are about their family members and what the impact is of their family members because we already—I mean we assume if you're a survivor of a serious disease that everybody knows you have it but you care about what happens to people that are close to you.

So, yes, it's stage, it's family member. I mean there are so—it's very, very nuanced. It's not something that you can just sort of make a blanket generalization about.

MR. SHELTON: I was going to go to the same place that you just went. So to build on it I would say before I came here, I always talk about my son because I think of my first role as dad and he is now 11. So before I came out here I asked him whether he minded if I mentioned his having this condition because when he was five it was my decision. He is now 11 and he is still not at the age of consent legally but I'd like to know if he has a sensitivity about me even mentioning he has this condition because it's starting to become his decision, not mine.`

So there's a lot of nuances to all of this and the irony that I see, again from a patient advocate perspective, the irony in our organization is--I talked about this high termination rate. Well, there's not a single person who ever met our son who has terminated—that I know of—that has terminated the pregnancy that they were carrying.

So we have in our organization a hotline for patient or parents that want to meet other parents or children of other parents. One of these people that decided to go forward with their were scheduled for a

termination on Tuesday morning—

we met them on a Sunday and they changed their mind--has become a long-time friend, and they live close to us, and we're invited to their birthday parties and their family events. We are always invited to come there 15 minutes early before anybody else gets there and they use that time to remind us that no one in their family knows that their child has this condition so not a single family member knows. Their primary care physician does not know. They changed primary care physicians because they didn't want the primary care physician to know. They—none of their educators know.

Now compare this to this exact same person comes to every national conference, is tremendously outspoken, is participating in two NIH sponsored clinical trials, and so they are tremendously active in a research context and in a context of advocating for the condition, and at the same time parents, primary care physicians, educators don't know because that's not who they believe the knowledge will help them. They're going to focus on the use of the knowledge in a

way that will accelerate their child's development and they think that the stigma of the condition is enough that—and I don't mean stigma like a bad thing but a stigma in terms of in their case they don't want their son to be treated like he couldn't do something in baseball. They want the baseball coach to throw it just as hard to this child as he would throw it to any child and not say, "Oh, well, he has got a syndrome so I'll throw it to him softer."

So that's their particular concern because they're athletic. They're into athletics.

But every single person that I've met in this role as—you know, as patient advocate and chairman of this organization is just a little bit different, and one person has got this little thing and someone else has got something else. Policy-wise, you know, which is—as I understand this committee—what you're here for, policy-wise I think there are some fundamentally policies that could be developed to empower that tremendous granularity in the society to take place and to take power.

And what I believe—just to the gentleman's

question asked me as I was standing up at the podium, what I believe would be the result of those policies and I think that it is a testable hypothesis if the agency wanted to find out. But what I believe would be the result is that more data would be shared.

The paradox here, I believe, is the Marco Foundation has done research that says the number one privacy protected behavior is failure to disclosure, nondisclosure. Number two privacy protected behavior is distortion of fact. So if those are the two privacy protected behaviors without technology, maybe there’s a way to encourage people to have greater trust in the system so they don’t fail to disclose and they don’t lie about their circumstances but they direct the data to go to the places where they want the data to be to help them or to help a family member or for altruistic purposes as someone said in their remarks.

That’s--the empowerment of that I believe will result in a proliferation of data, not a repression of data but that is a testable hypothesis and, hopefully, that could be something that could be tested and demonstrated as true.

DR. : I think what I am going to do is state the obvious. You know, you were asking about what elements work or any common themes, and it seems the common theme we're hearing is that—the whole notion of community engagement or informed consumers or participation. If people feel that they--if people do, not just feel, really that strongly, if people are involved in the decision making and have some say and have a reason to participate then we know the research about altruism is out there but when it's a feeling of lack of control and things going into the ozone is when—at least the research I've seen is when there is a lack of trust. So if we're going to come in anywhere I would say that that would be a very easy thing that I saw across all the speakers today.

DR. ROYAL: Anyone else?

I think going back--because the question that we have up there about best practices, are there best practices, and I think going back to what Laura said is--and sort of what Barbara is saying, I think, is the need for principles as opposed to best practices per se because these are—these models were quite different. I

think it was probably hard to assimilate all of this information because there are some similarities in It was kind of hard to assimilate all of this information because there are some similarities and there are so many differences but there are principles that seem to kind of—is a common thread in terms of trust and the engagement, and privacy.

I don’t know whether we think as a committee that is something that we may want to tackle in terms of coming up with principles. That probably could be applied across the board or the next question though is should we wait for the Lewin Report? They just started the work in terms of getting some background on what’s going on in data sharing, genomic data sharing, and theirs is a year long process and they are going to do interviews with various stakeholders, and then do a report. Do we want to wait? Do we think it’s best to wait for that report to decide if SACGHS should do something or do we think there are things that we could do now or should do now?

DR. : I think this is probably the worst time to come up with an answer to this question.

If you want to rush the committee you might get an answer today that may be different tomorrow morning at 8:00 when people are fresh.