



**HIT Standards Committee  
Precision Medicine Task Force  
Final Transcript  
August 19, 2015**

**Presentation**

**Operator**

Thank you all lines are now bridged.

**Michelle Consolazio, MPH – FACA Lead/Policy Analyst – Office of the National Coordinator for Health Information Technology**

Thank you. Good afternoon everyone this is Michelle Consolazio with the Office of the National Coordinator. This is a meeting of the Health IT Standards Committee's Precision Medicine Task Force. This is a public meeting and there will be time for public comment at the end of today's meeting. As a reminder, please state your name before speaking as this meeting is being transcribed and recorded. I'll now take roll Jon White?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Present.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Jon. Leslie Kelly Hall?

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Present.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hey, Leslie. Andy Wiesenthal? Andrey Ostrovsky? Betsy Humphreys?

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

Yes.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Betsy. Christina Heide?

**Christina Heide, JD – Senior Advisor for Health Information Privacy - Office for Civil Rights**

Here.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Christina. Eric Rose? James Breeling? Josh Denny?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Here.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Josh. Lisa Gallagher?

**Lisa Gallagher, BSEE, CISM, CPHIMS – Vice President, Technology Solutions – Healthcare Information & Management Systems Society**

Here.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Lisa. Mary Barton? Mitra Rocca?

**Mitra Rocca, PhD – Medical Informatician – Center for Drug Evaluation & Research (CDER), Office of Translational Sciences – Food & Drug Administration**

Present.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Mitra.

**Mitra Rocca, PhD – Medical Informatician – Center for Drug Evaluation & Research (CDER), Office of Translational Sciences – Food & Drug Administration**

Hi, Michelle.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

And from ONC do we have Maya Uppaluru?

**Maya Uppaluru, JD – Policy Analyst for Health Innovation, Division of Science & Innovation - Office of the National Coordinator for Health Information Technology, Department of Health & Human Services**

Present.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Maya. And Mazen Yacoub?

**Mazen Yacoub, MBA – Healthcare Management Consultant**

Hi, present.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Mazen. Anyone else from ONC or from other federal agencies on the line?

**Debbie Bucci – Office of Standards & Interoperability – Office of the National Coordinator for Health Information Technology**

Debbie Bucci.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Hi, Debbie. Okay, with that I'll turn it back to you Jon and Leslie.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Leslie, since I'm always the one shooting my mouth off why don't you start off this time?

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Well, thanks, Jon. Today I do not want to spend a lot of time because we've got some great presentations coming up and want to make sure we give them time, but I just want to thank you for participating and note how much rich information we're learning about what should be collected and shared in the EMR, what should be able to be retrieved from an EMR to support research.

And I think that we are honing the needs of this organization to make...or our committee to make sure that we can respond to an interoperable environment with recommendations that both are ready or not ready for primetime. So, with that I'll turn it over to you Jon.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

All right, thank you. I was not on the last call but I've been able to review it and I continue to be kind of like stunned and amazed by the wonderful, you know, super bright, super hardworking folks that are like keenly interested in this and I think just a word of philosophy, it just highlights kind of the importance of what we're doing to make this really exciting endeavor work and I'm really grateful for everybody's time and thoughtful input to us.

So, I agree, in terms of moving ahead I don't want to get too much in the way of the presenters because I think that, you know, we've got a lot of good stuff keyed up for today. So, in terms of the slide order, can we...let's see, Mazen I guess you've got the ball, if we could advance beyond the Task Force membership to just take a look real quick at the Task Force charge. Yes, right there, perfect.

So, the charge remains essentially the same and it's pretty consistent with what we've been saying. I just want to stop and just give everybody a real quick chance...any of the Task Force members want to ask any questions about where we are in terms of the charge and how we're moving ahead with trying to fill the gaps with the testimony that we're getting?

All right, silence is consent which leads to the next slide. And also, again, just a reminder, again, all these slides are available, you know, the materials are available on the left side of your screen, but just to, you know, give you a perspective of where we are. We started back in June or actually really in July, we're in the middle of August now and these are a number of items that are kind of moving in parallel to us that are going to be coming out that are probably worth kind of factoring into what we do.

We're still shooting for some initial recommendations later this month to be circulated amongst the workgroup later this month and then, you know, into the beginning of September and we're hoping at the September 22<sup>nd</sup> meeting to be able to put these recommendations forward.

I'll remind you that in parallel to us there is an Advisory Committee to the Director of the National Institute of Health, there has been a series of workshops that have been held. A number of us on the phone have been kind of weaved into that process as well so we're very well coordinated and our recommendations that we're going to work on will ultimately be very complimentary and supportive of each other in terms of what's going to come out of that advisory committee group and ours.

They are also aiming for the middle of Septmberish for their recommendations to be presented to Dr. Collins over at NIH. So, shall we stop there? Any quick questions about that and if there is not now we can certainly get to them off line if you have them, but just Task Force members anything that you want to ask about that? All right a silent but thoughtful bunch today. So, next slide.

All right, here's where we are in our work plan. This is really the last kind of round of presentations and input we're getting. In addition to the on line presentations that we're having, which are really wonderful, there is a series of folks that are probably also going to submit written responses to some of the questions that we have and we will have all that collated for everybody as we get into considering our preliminary recommendations at the end of this month. Okay, next slide.

Again, for the folks who are getting ready to present to us, thank you. The focus of what we're taking a look at is here, the exchange of genomic and phenomic data amongst participants EHRs, researchers, testing labs both for research and clinical care and the questions that, you know, you've all seen before and that you're keyed to for your presentations, but I'll just repeat them here for everybody's focus.

What's the key problem or set of problems your organization is attempting to solve?

What's the minimum interoperable data set of genomic and phenomic data for these data exchanges?

Are there standards that can support this movement today?

And what gaps are there and what's needed for the future?

So, those are the questions we definitely want to hear from you and anything else that you've got in the time that you've got we would love to hear as well. All right, next slide.

Okay, and with that we are onto the racecourse here with our amazing presenters. So, Michelle, thank you for the opportunity and we turn it back to you.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Well, I guess we haven't communicated how we want to do this so let's just turn it over to our first presenter then?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, yeah, yeah, so, sorry, presenters please...thank you so much for your time, please introduce yourselves as I think you all probably do a better job of saying who you are and what you do then we will. So, we're grateful for your time and with that let's go to our colleagues from Duke.

**Lori Orlando, MD – Professor of Medicine & Associate Director of Duke Center for Applied Genomics & Precision Medicine – Duke University School of Medicine**

Good morning, so this is Lori Orlando and presenting with me today is Ricky Bloomfield, and so I'm going to talk first a little bit about our program and what we've developed and why we developed it and then Dr. Bloomfield is going to talk about the technology that we're using and how we're leveraging what currently exists and trying to build something even better on that to support this program. So, do I just say next slide when I want to advance is that how...

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Yeah, that will work.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

That would be great.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

So, we're totally good.

**Lori Orlando, MD – Professor of Medicine & Associate Director of Duke Center for Applied Genomics & Precision Medicine – Duke University School of Medicine**

Okay.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

And Lori, just to mention, so it's 12:11 right now if you can aim for about 12:20, I know that's going to be challenging, but if you can aim for about 12:20 they'll give us about 5 minutes for a little bit of back and forth with you and the Task Force members.

**Lori Orlando, MD – Professor of Medicine & Associate Director of Duke Center for Applied Genomics & Precision Medicine – Duke University School of Medicine**

Okay. So, just background on myself, I am an Assistant Professor of Medicine, I'm a Practicing Internist, I'm actually in the middle of clinic right now, so I won't be able to stay on the phone too long, but, and then I am the Director of the Precision Medicine Program at Duke within the Center for Applied Genomics and Precision Medicine.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

And just briefly, I'm Ricky Bloomfield, I'm a Med Ped's Hospitalist and Director of Mobile Technology Strategy and I lead a lot of our innovation technology and interoperability work here. So, I'll turn it back to Lori.

**Lori Orlando, MD – Professor of Medicine & Associate Director of Duke Center for Applied Genomics & Precision Medicine – Duke University School of Medicine**

Okay. So, our goals for what we developed was to be able to facilitate precision medicine in a clinical practice, particularly in primary care that was where our focus is. And so we decided to start with something that enabled risk stratification within primary care populations because it was a good demonstration of both precision medicine and for improving population health.

But, you know, the problems in the primary care environment, I think you guys are probably all aware of, the very short time face-to-face time that doctors have with their patients really allows very limited uptake of these evidence-based guidelines that recommend risk stratification because they can't collect the data, they don't have the time, they don't even know what to do with it when they do collect it.

So a really good example of this would be breast MRI recommendations for patients with a lifetime risk of breast cancer that's greater than 20 or 25% and calculating those risk scores is not such a simple thing to do in clinical practice or in getting a family history that might tell you that a patient warrants further evaluation for Lynch syndrome.

So, these are the kinds of things that we were looking at trying to do as well as serve as a sounding board for developing learning and shared decision making between patients and providers. So, we really wanted patients to receive information as well as providers to receive information about risk and what to do about that risk. Next slide, please.

So, these are the barriers that we were encountering, patients have limited awareness about their family history, very limited education about what family history is and what it means and how to collect it so at the point of care they really don't provide the information that's needed and the depth that it's needed in.

Providers don't really have the time and don't really appreciate the prevalence of these hereditary conditions in their clinical practice and then the health system really doesn't capture that data for them in a way that they need it to be done. Next slide.

So, we developed a program that we felt needed to have four key factors to it, it needed to be patient-facing, it needed to be guideline-driven, it needed to have embedded clinical decision support and it needed to be a web service so exist outside of the EMR and communicate with the EMR in real-time.

So, for the patient we developed embedded education about how to collect family history, what it meant and then how to talk to their relatives and how to gather...what information they needed to gather. We allowed them to update it at any time and provide them specific clinical decision support documents that are tailored to their needs. We also provided English and Spanish versions.

And then for the provider we gave them time to actually just discuss what risk management needed to be done rather than collecting data and creating all of that gathering at the point of care. We then integrated clinical risk calculators, performed the risk assessment and tied it to actions. So, instead of just saying “hey, this patient is at high risk” we say “this patient meets these specific action oriented criteria such as breast MRI or genetic counseling.” Next slide, please.

So, we tested this program when we initially built in a community-based primary care practice in Greensboro, North Carolina and we found, you know, a tremendous amount of positive patient and physician support, a lot of providers initially felt like they were doing a good job but after using it for a little while started recommending it to their friends, saying “this is, you know, a really great opportunity to improve our practice.”

It significantly improved the amount and the quality of the data that was being gathered around family history from patients. And we found that there were 40% of patients that were in the routine practice that were at high risk for one of the clinical decision support conditions and 90% of those were not getting the risk management that they met criteria for.

So, there are a huge number of people out there in the general population who need this kind of information collected so that they can get the right care but they aren’t currently doing it in the system that we have today. So, we need a systematic process for doing this. And then it helps physicians direct the right resources to the right patients. So, next slide.

So, moving forward, so, you know, its family history-driven right now because that’s the information that we know physicians accept that they should be collecting and we don’t do a good job of it and the EMRs don’t do a good job of it, but going forward we really need to start to be able to build in additional pieces of information to this for example genetic and genomic information like when a breast tumor comes back positive, you know, or if they are triple negative breast cancer status we would need to know that because that information can be used to guide whether the patient gets testing for hereditary cancer syndrome or what’s PALB2 status things like that can begin to be incorporated in with the family history and the other available risk algorithms in a way that can be very actionable.

So, we know there are pieces of information that are, you know, all genetic sequencing obviously may not need to be in discrete data elements but there is certainly some data that needs to be put into a discrete format that can be transmitted and used in clinical decision support rules. So, next slide.

So, these are just some screen shots and because Dr. Bloomfield is going to talk about the technology a little bit I’m going to go through these pretty quickly, but this just gives you an idea of what patients see when they go through so they get this information upfront when they first enter the tool. Next slide.

And then these are the screens that they see and you can see here if you hold the cursor over one of the diseases they will get information from Medline Plus Connect that immediately pops up within the App to give them more information which is written at a 5<sup>th</sup> grade reading level. Next slide.

And then they’re required to enter three generation family history. Next slide. And they can drag and drop additional relatives onto these bars using a touch screen which quickly creates a structured family history for them. Next slide. And then we create a full pedigree as well as a grid view. Next slide.

The providers in primary care practice, at least, much prefer the tabular family history view then the pedigree. So, we provide both. Next slide.

And then this is an example of risk information as the patient sees it they get a talk to your doctor about this. This is the piece of information that you entered so this is why you're getting this recommendation and then some additional information that they might need about that recommendation. Next slide. So, they're prepared to have a discussion with their doctor.

The doctor's reports are very different. They have been tested repeatedly with the physicians to give very specific action oriented recommendations upfront with references that they can link to. They just click on the reference it links them out to the guidelines and then again the specific pieces of information that are creating the trigger for that, that way they can verify just the important pieces of information and the family history and they don't have to verify the entire thing. So, next slide.

So, the tool is built on the AHIC criteria for high quality family health history and we support HL7, SNOMED codes, ICD-9 and now SMART on FHIR which Dr. Bloomfield is going to talk about.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Next slide. So, I'll be brief here, I imagine most members of the Task Force are familiar with these technologies as well as Argonaut, but at a very high-level it's an open source, a standards-based platform that allows us to integrate Apps like MeTree into the EHR both from a technical perspective as well as a workflow perspective which was really important to us so that we could get this in front of our providers as well as in front of our patients through their personal health record. And Argonaut is sort of a collaborative to help accelerate the development of these standards which are still not finalized but they're robust enough that we can actually start using it. Next slide.

So, what we've done here is about a little over a year ago we felt like we had a need to incorporate some of our innovations here internally into our EHR in a standards-based way so that we could scale those in a meaningful way not only to other institutions who might have our same EHR but also those who may have a different EHR. And we felt like that the SMART FHIR platform was what could do that.

And so we internally, our development team built support for the standard on top of our EPIC-based EHR and we had this functional earlier this year in January presented it at HIMSS and were planning on having an App live in production by the end of this month or the first week in September which will be a pediatric growth chart's App and then the next step will be to put MeTree into production not only for physician's use but also for patients and that includes appropriate authentication and authorization according to the Argonaut's specifications.

And MeTree will be the first patient-facing application that we're planning on going live with and that will be available through the MyChart, the personal health record and we think it will be a great way to collect the data and then to integrate it into the patient workflow. So, I'll stop there I know we're two minutes over our 12:20 but perhaps that will leave us a little bit of time for questions.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

All right the two of you are sprinters, oh, my gosh, that was really cool, thank you. So, I've got a ton of questions but you know what I'm going to defer to my colleagues. Folks on the Task Force are there any burning questions? We've probably got time for a quick response for about 2 or 3 at least. And if you don't I'm going to ask them so now is your chance.

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

This is Betsy Humphreys.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Hey, Betsy.

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

Thanks for the great presentation. I'm glad to see that Medline Plus Connect is being helpful to you. I'd like to just hear a little bit more from you about your feeling that the standards in terms of Argonaut or whatever are, although not finalized, where obviously far enough along for you to implement them. Do you...

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Oh, Betsy, you may have just been cut off. If there is somebody out there who is not on mute, mute yourself. Betsy, could you just real quick repeat?

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

Oh, okay, I was just going to say that you commented that although the standards, the Argonaut thing is not final you've found that it was far enough along that you were able to implement it. Is this because your team is, you know, able to leap tall buildings with a single bound or do you really feel it's pretty much at the point where, you know, we just need to pat this thing on the head and then the standard would be usable by a broader audience?

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

So, great question. I think that right now one of the goals of Argonaut is to have implementers that actually test this stuff out so that we know what gaps may still be within the specification. Having said, that, yes, our team does wear capes and they can leap tall buildings.

The reality is we don't have a big team here and we probably had two FTEs working on this for about six months to get the initial implementation done and now about three FTEs over the past few months so not a massive effort by EHR standards and I think that's a testament to how easy these tools are to work with that while not final they are well documented and easily accessible especially for those who maybe familiar with web development paradigms already and web technologies so it has made it very easy on the team.

I would note though that our team prior to Duke going live on EPIC several years ago they actually built the EHR that we used prior to EPIC. So, they have a tremendous amount of experience that probably not every healthcare system has in that they had built an EHR and they knew what they needed to do to make that happen so we were fortunate in that respect.

And I wouldn't necessarily recommend everyone build support themselves because through Argonaut most of the major EHR vendors have committed to adding support themselves and once this official support is available to us from EPIC we plan on using that and not continuing to develop something on the side.

Our goal is to be standards-based and we do have the ability to create some additional infrastructure should we need it, if there's a resource that may not be provided by our vendor we could add that, but overall we would like to use a supported product rather than continuing to create a product that we have to support.

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

Thank you very much I think that's very helpful given the task of our group to know where you think this is and I'll just comment that it's been our experience at NLM that adding more programmers to a team does not necessarily help you get the job done right faster.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Agreed 100%.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

All right so that was an excellent question, thank you, Betsy for that. With that I'm going to thank Dr. Orlando and Dr. Bloomfield very much for your time especially in the middle of clinic, oh, my God, your running off to your next one I'm sure, but I really...

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

And Lori has to jump off but I'll be here for the rest of the call, this is Ricky. If there are any questions...

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, fantastic.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Later.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, okay, Lori, get your charts done, thank you and Ricky, just, yeah, I appreciate if you can hang around a little bit and if there are more questions or discussion at the end I really appreciate it. Okay, so with that the whirlwind continues on. Let us go to our most excellent colleagues and now my federal colleagues, partially at least, Dr. Denny and Masys.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Hello. Let me go ahead and jump to the next slide. My name is Josh Denny I'm at...you can see that I have dual e-mail addresses here, I'm at Vanderbilt and direct some of the precision medicine efforts there around informatics and reuse of EHR data, and am a PI of our eMERGE network site which I'll be talking about here, and then, but currently I'm an IPA with NIH. So, next slide.

So, we all know this slide very well, but the key component of this that I'll talk about is just that I think as we said, EHR data will play a big part in what we're thinking about for precision medicine as a method for providing passive follow-up on individuals and a key component of what President Obama has mentioned and our desire through the program really is that this be something that could be open to everyone.

And when you think about what is open for everyone that means that patients as they become participants in PMI would need to be able to provide data that becomes useful for secondary use for research and so that's the key problem I'm trying to answer here is, you know, what types of data get used for research uses when you think about genomic or precision medicine studies and leveraging the work that's been done in eMERGE and other networks to guide that. So, next slide.

So, this is just a graphic of some of the networks out there that currently have electronic health record data linked to biorepositories, the Million Veterans Program and Kaiser being two notable examples and then all the rest of the stars back there in the background would represent the eMERGE sites and if you see the total there we have over a million people with existing EHR linked data that we can use as examples with eMERGE doing a lot of work so far and MVP and Kaiser doing other work with the PMI, you know, intending to be another set of a million or more. Next slide.

So, this is just an example from our DNA Biobank which its sole source of data for phenotyping really is from the electronic health record. We have about 206,000 individuals collected so far with a mean follow-up of 6 years. We have about 22 million ICD-9 codes with 7 million distinct patient codes, patient ICD-9 code payers. You can see the number of labs, drugs, notes, radiology tests there even the small numbers there like the radiology test represents, you know, about 10 radiology tests on an average per person. So, it just shows there is a lot of dense information and not all of it may be treated equally. Next slide.

So, the basic process to determining a phenotype in the EHR is something like the...once we identify what we want to study its and iterative process whereby usually informaticians working with clinical experts propose an algorithm and refine it with subset review and so this is in contrast to, you know, maybe a lot of studies where you might purely use an electronic resource to do a study and not really validate it and just acknowledge it's weaknesses or where you mainly adjudicate and review, you know, all your files, what we would do is say, you know, for many of these phenotypes, you know, we can't review 206,000 charts and neither could we review 10,000 cases of diabetes.

So, we review small samples of 50 or 100 cases maybe randomly intermixed with another 50 or 100 controls with algorithms to find them until a positive predictive value feeds 95% or 90% whatever you're threshold is and then in a network like eMERGE we would deploy it at one site and then we would get a few other sites to validate it as well and tweak the algorithm based on different perspectives in the EHRs, etcetera, and then perform a genetic association test and replicate that algorithm across many different sites.

And a lot of times you can take a set of genotypes and reuse them. So, they may be genotyped for one reason but they can be used for many studies and this kind of efficiency of study is what is imagined for PMI. Next slide.

So, what we found, in this very simple graphic, is that finding a phenotype with high precision usually involves multiple different elements and combining these different elements in different ways often times with Boolean logic but also could incorporate machine learning techniques or scoring systems, or regression algorithms, many different things have been used that have combined these kinds of elements the most common of which are probably some combination of sort of and/or not types of logic with thresholds of maybe having certain types of elements and temporal reasoning as well. Next slide.

And when we look at any given population of patients they can kind of break down into four groups and this is an example of a study we did when we only had 10,000 people in the Biobank and tried to find patients with rheumatoid arthritis. We picked this because we thought it would be easy. The two of us who were picking what we were going to do both have internal medicine training and we decided that, you know, we all know what RA looks like so we went after it.

And what we found is that, you know, it's actually fairly tricky with a number of patients who are being ruled out for rheumatoid arthritis, patients that may have similar diseases they get treated with similar medications like psoriatic arthritis, people could get billed for RA but then later end up having, you know, some other kind of inflammatory arthritis or even a variant of osteoarthritis as kind of a driving factor.

And so, when you look at the individual populations there are...you can find algorithms that accurately define controls and cases that exceed those thresholds and then a group that may have possible cases that you can manually review or refine if you need to for sample size and then a bunch of patients who just may not have enough information in their chart.

We were pretty strict with this case, but maybe for rare diseases that, you know, most people are a control and you may not need to be so exacting with that but for, you know, things like diabetes or hypertension that are very common finding controls is actually one of the hard parts of generating an algorithm. Next slide.

And this represents a slide from one of the first studies we did in this where we looked across five diseases with 21 SNIPs that were known to be associated with those diseases. The red boxes represent the published odds ratio for these 21 disease SNIP pairs, the blue diamond's represent what we found in our Biobank.

Notably, we were underpowered for most all of these analyses, but you can see that the errors bars for our blue associations overlap the known associations for almost all of them. In many cases our odds ratios are similar, in some cases they're lower, some cases they're a little bit bigger. Overall, we replicate 8-9 out of the 21 associations depending on which type of statistical test you do.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Hey?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Yes?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

It's Jon, just for the record as my inner science geek comes out with all the discussion we've been having recently about non-replicability of studies that's cool.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Thank you. And we have actually done a number of other studies. We've found and I don't have a slide here on this, but we replicated...we look to replicate 750 different disease SNP associations and we're able to replicate a number of them including about 60% of those that we were adequately powered for, but if they'd only been published once we replicated less than half of those. Those that were published more than once, as you would not be surprised, we replicate 75% of those.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

You hold the power of meta-analysis.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Yes, exactly. So, next slide. Next slide, please.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Oh.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

There we go. So, this summarizes what we've investigated in eMERGE so far. Each of these is a different GWAS we've done. Each of these has a phenotype algorithm that's been developed for it and validated at more than one eMERGE site. The bold ones represent ones that have had geno-wide significant results. Many of these are replications but a number of them also had new discoveries, probably about half of these associations actually, you know, found something, a new genetic signal that we were able to replicate in another population.

And I'll also point out the group on the right there where it talks about contributing consortia, many of these consortia are typically mostly composed of non-EHR related samples so it's provided a nice opportunity for us to compare what we find in the EHR versus not and what we find in our data look very much like non-EHR data that can be observed from clinical trials in terms of finding our other cohorts, in terms of finding signals and we can also do pharmacogenetics phenotypes and I list five there that would be classified as that and we have two significant signals for two of those that have been recently published. Next slide.

And we created a site called PheKB where we've posted these algorithms, all the eMERGE algorithms are up there in some stage of development. There is a total of 92 that are in some stage of development and 53 of those 92 phenotypes have been evaluated 149 times and it's a way we can share across a number of different networks. Next slide. This is a...

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Josh you've got...

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Yeah?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

You've got three minutes.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Okay, this is a schematic of one of these algorithms. Next slide. And here is a schematic of using...doing a pharmacogenetic algorithm where you can see that we look at labs and temporal distances between them as well as codes and medications and things like that. Next slide.

And what we did is we evaluated those 92 phenotypes in PheKB and showed what is used for what and you can see the most commonly used element are ICD-9 codes then medications, then some sort of text processing or natural language processing, CPT codes and laboratory and test results. And you can see there is not any one element that's used in everything but it gives you some idea of the types of information we would want to share if...in doing this, individually sharing records. Next slide.

And this is overall performance, they do pretty well. Red would be the median performance both over 90%. Next slide.

And this just shows you the size of different elements and what I want to point out here the color represents different things. So, if you look at atrial fibrillation there is a really big circle of people that have highly specific medications for atrial fibrillation. If you look at the green circles many of them have big green circles representing text mentioned for those diseases but there are people that have ICD-9 codes that don't have medications, ICD-9 being yellow, and vice versa. So you want to think about different elements and different components when you want to find a given diagnosis. Next slide.

And this shows relative performance. No method is perfect combining methods across 10 diseases did better than any one component does. Next slide. And I'll just go to the next slide. We'll switch the...let's get to the...so summaries, ICD-9, CPT codes, medications, labs, problem list elements were used a lot and narrative notes become important especially to confirm things.

We typically don't use family history information and actually try to remove that in identifying who has what disease. It gets important to know what is an inpatient, outpatient, laboratory or medication given and we want detailed data information so it is often how often it happens and temporal reasoning requires that information. So, these are elements that we'd want to have shared for research use. Next slide.

And if we look at the data dictionary elements that gets shared for these covariates everybody uses age, sex, race, ethnicity, date of birth, height, weight, body mass index are commonly used and very common looking at smoking and tobacco use, diabetes and hypertension often pop up as comorbidities, and then you'll see a given disease that gets used, specific labs may be reported. Less common were other vital signs things like metadata around enrollment in the healthcare system and number of visits per year or something like that, and alcohol use. Next slide.

So, we've proposed this idea of Sync for Science which would build on things like Blue Button or related technologies to share data but something broader for Sync for Science would support...would be the idea that you would link or sync a healthcare data system to continuously update another entity and this could be something like the PMI. So, a user could decide, a patient, that they want their data from hospital X to continuously accrue to another site it could be another hospital, it could be a research use and the idea that it would not just be a record of care from a given visit but hopefully it would be more of an entire health record and we would prioritize that based on the data we showed before billing codes, then some degree of medication data and then narrative data or problem list data, then to labs seems to be a decreasing importance. And, you know, many technologies could be used to think about how we structure this but FHIR of course is one potential that has a lot of traction now. Next slide.

And the idea would be that we would have tools or a process for patients to review, validate and contextualize their data, it could be from some central research resource. This would not have to be part of the standard.

But where we need new standards is how you would develop this link and how you would give patients the ability to control the link and disconnect it which would be especially true if you were looking at a given hospital system to another system that...and given the number of connections someone may have you may need to think about if there needs to be centralized control over this and we might need to think about incentivizing providers and EHR systems to adopt such a technology and PMI being a prototypical use case, but the impact could obviously be much broader for healthcare in general in setting a standard for clinical care. I think this is my last slide.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

All right, so, yet another dense, rich and satisfying presentation and really nice contrast to our colleagues from Duke that talked about how they're using their EHR to deliver research results into care, you provide a really nice demonstration of how you use the data that are available in our EHR systems to do the science and more importantly share with us some very specific recommendations on what data elements you need. So, given the time, Josh, I hope you're able to stay around and participate in questions at the end, is that all right?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

That's completely fine I'll be here.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, fantastic, thank you, and I want to keep going with our schedule. Next up we have...

**Andrew M. Wiesenthal, MD, SM – Director – Deloitte Consulting, LLP – International Health Terminology Standards Development (SNOMED)**

Jon, this is Andy Wiesenthal, can I make a very brief comment or would you like to reserve...

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

You may go.

**Andrew M. Wiesenthal, MD, SM – Director – Deloitte Consulting, LLP – International Health Terminology Standards Development (SNOMED)**

It for the end?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

No, no.

**Andrew M. Wiesenthal, MD, SM – Director – Deloitte Consulting, LLP – International Health Terminology Standards Development (SNOMED)**

Yeah, so, yeah, just on the last presentation to note that Kaiser Permanente's data is not only ICD encoded but also SNOMED CT encoded and it might be useful to compare the performance of algorithms using one or the other or both that would be novel given the size and scope of the data that they've already collected.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, that's a great flag, let's hold that for the discussion section to potentially talk about here and if not we can certainly talk about it off line, but that's a great point Andy, thank you.

**Andrew M. Wiesenthal, MD, SM – Director – Deloitte Consulting, LLP – International Health Terminology Standards Development (SNOMED)**

Thanks, Jon.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Yeah, no worries. Okay, so let us go to my friend John Wilbanks from Sage Bionetworks. John you can go, since we're starting you five minutes late, you can go five minutes long up until 1 o'clock if you so wish and with that the floor is yours sir.

**John Wilbanks – Chief Commons Officer – Sage Bionetworks**

Okay. Thanks and I actually don't have slides I've sort of done something unusual for me and I've written a document, which I'll mail out Jon to you so you can circulate to the group, trying to sort of speak directly to the questions for presenters, so I'll run through those, if we want to ask questions I'll be here until the end of the call.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

No problem, actually just introduce yourself real quick just in case there's people who don't...

**John Wilbanks – Chief Commons Officer – Sage Bionetworks**

Sure.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Know the legend of John Wilbanks.

**John Wilbanks – Chief Commons Officer – Sage Bionetworks**

I would envy you. So, I'm John Wilbanks, I'm the Chief Commons Officer at Sage Bionetworks, we're a non-profit biomedical research foundation based in Seattle, I'm here in DC, so I work from home and you may hear my dog in the background.

I've been working on open systems for science and health for about 15 years with a focus on the ways that we can essentially design aspects of emerging technology systems so that they are more beneficial to individuals, entrepreneurs, small companies, ad hoc teams so that things look more like software more or less. And so I work on policy, I work on informed consent in particular, data governance and other sorts of things at Sage and I'll walk through a little bit about what we do while I answer the questions.

So, starting with, what's the key problem or set of problems that we're trying to solve, I would say that there are three key problems that we're working on. The first is what we see as a total failure of existing methods for clinical research to generate and analyze data in ways that are capable of real stratification, personalization and precision.

The way that we are set up structurally to do clinical research studies is based on lots of human touch and involvement and on really, really small scales and so the data is going to explode in scale, the sample sizes are going to explode in scale and our methods aren't ready to keep up and so that's probably the first thing that we started to work on.

The second is what I would call the centering of massive organizations. So in some cases corporations but certainly government, large scale non-profit systems like health systems and payers those organizations are at the center of our healthcare system over individual citizens, small companies, entrepreneurs, ad hoc teams, networks all of those things that we take for granted in software are uncensored in our health system and we want to work on solving that as well.

And then third is the weakness of describing disease in human terms as opposed to data informed terms. And so I'll unpack each of these a little bit real fast.

So, the failure of the methods we've worked on by trying to create federations, communities to bring things like version control and provenance to the analysis of data and we've got observable productivity gains again and again, and again from these kinds of methods working on things like colorectal cancer subtyping, so taking 12 groups that have different ideas on how to do genetic subtyping for colorectal cancer bringing them together to improve each of their methods and then to build a consensus method out of the best features of all 12. The consensus always performs all the individual groups over and over, and over again.

We see this with our work with the cancer genome atlas, the accelerated medicine's partnership for Alzheimer's disease and we've got about 8000 data scientists who log into our various systems for this on a monthly basis, but that really serves the experts so we started working with a platform called "dream" to bring in outside experts through competition collaboration cycles so the two most recent ones from this summer we got high impact publications on tumor genome simulation with crowd sourcing and predicting tox response to unknown compounds replicating this concept of a contest for the individual methods and then a consensus tool that comes out that takes the best pieces of all the methods. Again, we see here, in the sort of contest space, again that the consensus tool always out performs all of the individual tools.

And then last we wanted to work on actually running clinical studies that go straight to the individual through our bridge platform. So, this is a mobile centered clinical study platform that forms the backbone of the first wave of Apples Research Kit Apps and that's really looking into digital phenotyping as a way to do...as a set of methods for generating data that comes in that's more individual.

And so those platform approaches tie to the methods failure. To the massive organization centering failure we really try to preach an open ecosystem, you know, platforms are not enough we've got three between...dream and bridge but the magic happens when they come together and get network effects from each other as an ecosystem that's how sort of the world we live in works.

And, you know, we have Google ecosystems and Apple ecosystems. We're trying to build out an open ecosystem for clinical studies that enables all sorts of unexpected uses and reuses because we want to move clinical research from something that is controlled exclusively by the elites sort of like web programming was in the late 90's to something that is more democratized like web programming is today. It still takes a lot of work to do web programming but a skilled and motivated individual...15 years ago and that's what we think needs to happen in clinical care.

And then last is the idea that the way we describe disease is a fundamentally ancient artifact of humans. We have these collections of symptoms that emerge in human observation that drive the vast majority of our descriptions of disease but what we're finding is as we collect data straight from devices the symptoms that are human readable don't match up to the patterns that come from the data. So, Parkinson's to us is not something described by a tremor recognized by an expert but by 25 core sensor data features and there's a real transformative capacity to connect that kind of data straight to machine learning and deep learning that's actually hampered by routing through human language and that's going to be a transition that's a little tough for folks I think. So, this is a problem.

To the question on minimum interoperable dataset, we don't really know yet, I mean, what we've seen at Sage is that when we think we've got 90% of what's needed you discover that you need more almost every time and it might be more like the early years of telescopes where each new kind of data generation method unveils new complexities that makes that first version of the data less useful than we expected.

Now sequence of an individual is certainly useful especially when it's tied to phenotype for the individual. Omics is certainly useful, again, when it's tied to the phenotype but the digital phenotyping from pervasive computing may well unlock a massive amount of what we need to do for precision medicine.

Phenotype is set to explode and almost all of it is going to explode outside the EMR, outside the medical system. And even when you can get the EMR data from our perspective it's lost a lot of information compared to what we can get going straight to the participant.

So, in our Parkinson's work we do a tapping test that gives us dexterity and dyskinesia on the touch screen. Now in an EMR we would just get the number of taps and that's useful but, you know, on the actual touch screen we get 25 kinds of data about the taps, we get the raw number but we also get the median and the mean, and we get a sense of where they're tapping which gives us a real indication of their tremor. So, we see two individuals who claim to have benefited from L-dopa, one of them would be measurable through the EMR because that would give us taps or from the research reports, giving us raw taps, but if we have a different person who is getting a benefit in accuracy but not raw tapping we wouldn't even be able to get that out of the clinical research data because it wouldn't get recorded.

And the opportunity cost for us as a non-profit has been so high to get phenotypes out of EMRs that so far we haven't even tried to do it once. We're looking for studies that want to tie to data warehouse queries because that appears to be a feasible route is to not say give me all the data about John Wilbanks but to say, give me the following nine data elements about John Wilbanks and we are committed to supporting relevant standards and we have studied...but for us it comes down to sort of minimum interoperable data is contextual to the question being asked in the study and so probably would just sort of give a generic answer to that.

Then standards that can support the movement today, there are some to a certain extent, the caveat that we would make is that the tech is changing so fast that standards developed in a traditional way through working groups that publish hypothetical descriptions of standards are probably going to be out of date before they're drafted much less released.

So we would endorse frameworks that encode de facto best practices or standards, technical frameworks, science frameworks, legal/ethical over document creation to describe standards.

For mobile the research kit stuff from Apple is a good set of functionalities to standardize on. So, this is not a recommendation to standardize on the research kit itself but to think about the set of functions. It is coming for Android and it will be functionally equivalent so that's a good place to sort of create traction is to say, we're only going to fund studies that either use these standard functionalities through these open source products that are available or that add to those frameworks so that you can create sort of accretive standards and functionalities.

For EHRs, you know, we looked at FHIR and Blue Button, and Direct we love them but it would be really helpful for us if we had reusable architecture components that we could use to add Blue Button to a mobile study application or that would make it really easy to add FHIR and Direct to cloud and platform frameworks such that an average Amazon web services developer could easily drop in modular code that did that in a useful way because that's something that my tech keeps coming back and saying to me doesn't exist in a satisfactory form.

And then last and sort of gaps and what's needed in the future, I may sound like a broken record, I mean, for us it's all about reusable architecture that's built in modular ways. So, we want to have something akin to WordPress but for building a clinical study at a technical level so that you don't have to write all of the code to make WordPress you have stylesheets and plugins and all sorts of other things and we'd like to have that so that we could build studies and that non-expert people who understood themselves could build studies.

And then something on the back end like Bioconductor which is a set of protocols and best practices for high-throughput genomic data we need to have something similar to that but for analyzing the clinical study data, because there's going to be a lot of variation in analysis and the more we can encode best practices as products and protocols the better.

We also would like to have some legal and ethical frameworks for what's acceptable for different levels of risks and benefits like when is it necessary to verify identity of enrollees, when do you need an in person date to make sure that people understand and to provide consent. And just generally for the scientists, you know, on all sides of this we need protocols. You know we have methods in molecular biology, it's been around for 30 years and it's arguably helped accrete the field. We don't have methods in precision medicine and so a lot of what I think when people say they want standards and where there are gaps, the gaps to us are methods gaps and rather than sort of declaiming that this is the new standard method I think simply collating and publishing, and annotating the protocols that are out there and the methods that are out there could be enormously powerful especially with the stamp of the initiative.

As I said, I'm going to send this as a Word document to the group because I know I went really fast and I'll hang around to answer questions afterwards and thanks Jon and to the working group for having us here today.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Well, John it is absolutely our pleasure. I feel like I've put on the ring from The Lord of the Rings and I'm seeing the world through different eyes. You know and to just throw a few more...in there just because I can, you know, consistency maybe the hobgoblin of little minds but focus on what you put a dent in the world so I don't mind you making that same recommendation a couple of times in different ways.

Let me stop and say, look, you know, we're a little bit over we're not hugely over, we've got three minutes or two minutes now, folks from the Task Force any further points you want to draw out perhaps not have deep discussion about but pointed questions that you want to ask?

**Andrey Ostrovsky, MD – Chief Executive Officer – Care at Hand**

This is Andrey Ostrovsky, I have a quick question and I also don't want to sound like a broken record. I think the last speaker summed up an incredible overview of the challenges and some paths forward, and also to call upon the speaker prior to that, and this may be out of scope of a precision medicine conversation but within scope of precision health conversation, if family history is not relevant like two speakers ago, I'm wondering is it relevant to try to capture information like housing and security or whether someone has access to transportation to go to a doctor? And are there frameworks to call upon that will incorporate that non-EMR or not typically EMR data into this whole body of work? I'm not aware of any other than the eLTSS work going on, but if the previous speaker is available to comment, is there any way to create a framework for introducing new frameworks more efficiently?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

And Andrey, did you mean previous speaker John Wilbanks or did you mean previous speaker Josh Denny?

**Andrey Ostrovsky, MD – Chief Executive Officer – Care at Hand**

Oh, for John Wilbanks.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Yes, go ahead, John.

**John Wilbanks – Chief Commons Officer – Sage Bionetworks**

So, we've not done anything with family history other than use surveys to sort of tease it out. One place that would be tremendously useful for standardization to exist would be to have a standard baseline set of surveys. So, we've got some stuff through NIH PROMIS there are some copyright issues that were a little difficult getting them useful for Apps, but to have sort of again decomposable elements, because like having a 70 question survey isn't very useful if you don't have a person asking the 70 questions because people get bored and stop.

So, if we could have decomposable surveys like here's 10 questions on family background, here's 10 questions on transportation, here's 10 questions on medications, right, if we had something like that and I think transportation and access is tremendously important, right, a lot of the value that we're seeing comes from, no surprise, it comes from some of the same places that re-identification comes from which is a diversity of different points about the same person.

And so if you can get, you know, good solid standardized questions about their transportation options then that would be tremendously valuable but the real value comes if every App that asks transportation asks the exact same question and everyone who scores the answer scores them the same way.

**Yaniv Erlich, PhD, B.Sc – Professor of Computer Science – Columbia University**

So, this is Yaniv Erlich from the New York Genome Center I'm going to speak next but I just want to comment on that. So, one source of information we are incorporating right now for family history are death certificates. So, in most states, the United States, death certificates are considered as public information and right now we have all the data from Vermont since 1986. You can get quite a lot of phenotypic information about the individual, you have the ICD code of the cause of death and also about their profession, their environment, demographic status of these individuals and you can incorporate that into family trees to get more about the family history.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

So, actually, this is Jon White, that is a great segue over to Dr. Erlich. So, Andrey I appreciate your question and I appreciate your response. But let's move to the next presentation and then after that we can have a longer question and answer session or longish, you know, relatively speaking. So, Dr. Erlich the floor is yours until 1:15.

**Yaniv Erlich, PhD, B.Sc – Professor of Computer Science – Columbia University**

Yeah, thank you, very much I appreciate giving me the opportunity to speak for this group. Can we please go to the next slide? I don't see my slides on the...oh, great, okay, terrific that's great.

So, my group works on the intersection of social media and genetic information. In the past we had a study where we showed that we can take whole genome sequencing data that is allegedly de-identified and by querying social media websites for genetic genealogy we can infer the surname of individuals by matching the "y" chromosome.

And interesting thing about this study is that we don't need the same person to be a participant in the genomic study and also in the social media websites because the surnames reserved this connectivity with the "y" chromosome also with far relatives.

So, today I'm not going to speak about privacy at all I'm going to speak about opportunities how we can build very large family trees using a social media website. Now I would like to apologize that some of the slides didn't convert well to PDF so I also posted a link with my PowerPoint version of this presentation so if you want to see the full version it's available. Can we please go to the next slide?

Okay, so we have now this brave new world we can sequence many genomes and I want to emphasize that large family trees are a multiplier of genomic information. Next slide.

So, we had a study about a year ago where we sequenced a family who had Goldenhar syndrome, this is a facial malformation, we sequenced a small family, we sequenced the grandmother and her two grandchildren that suffered from this disease and we saw that we had eight variants that could explain using exome sequencing this malformation, but since we are skeptical geneticists we worked hard with the family to identify some far relatives that suffered from the same disease.

We identified a distant cousin of the grandmother that suffered from the same disease, we collected DNA and it actually turned out that none of the mutations that we thought in the nuclear family are the cause of mutations because they were not segregated with this distant cousin. So, having this large pedigree we could identify copy number variation on chromosome 14 to identify the cause of gene.

So, I want to take a message from this story, of my own group, is that if you have large pedigrees you can really get much more information from your genomic datasets. Next slide.

So, another way to look at that is this study in Iceland by Decode Genetics. So, in Decode they were able to sequence or to infer the genomic information for most of the Icelandic population by having a large family tree of this island and also sequence only 3000 individuals. So you have 3000 individuals, you take this information, you place it on the tree and then you can infer quite a lot of the genomic information of all the individuals that you didn't sequence. And this is just to highlight this opportunity. So, next slide.

The question is how can we do that for the Precision Medicine Initiative? How we can construct very large family trees because this process is not very easy. Next slide.

So in my group we...to address this issue we are looking now at a social media website called Geni.com. On this website genealogists can upload their family information and their family trees and if there is a match between two genealogists, between the two trees, the website will offer them to merge the two trees together and to create a much larger family tree.

And what you see in this slide is my own family tree with approval of our IRB and from Geni.com we downloaded all of the public information from the website everything that you can see if you have...internal information from the website just what you can see when you search the website from the outside. We obtained 80 million profiles by this process. Now the question is, can we get to these very large trees and the answer is "yes." Next slide, please.

Once we have this data from the website we use the GEDCOM standard that is out of the file that each individual is given and we just take all these files and then we can construct this very large family tree and we need of course to clean also the data to make it usable. Next slide.

And here is an example of a very large family tree. We have here 60,000 individuals in this tree but in fact this tree is part of a much larger family tree of 13 million people that are all connected, they're all together in that one big tree, so I'm just showing you 0.5% of the data from this very big tree.

We have some famous people in this pedigree, we have Barack Obama, we have Sewell Wright he is the father of human genetics, we have Kevin Bacon of course you have to have Kevin Bacon and we've found that the number of steps between Kevin Bacon and Sewell Wright are 24 steps based on genealogy. So, we can really get to these very large family trees using this website. Next slide.

We also did some analysis using "y" chromosome in mitochondrial information to know if the data is accurate and what we found, you cannot see it on the slide but I will tell you the numbers, for the mitochondrial data that measures known maternity event on the website we found 0.3% of our edges on this graph on this family tree are wrong. For the known paternity rate using the "y" chromosome we found that 2% of the edges in this family tree are wrong and we're very pleased to see these numbers because we know that in Europeans the known paternity rate is about 2% and that is what you see in this slide.

So, although this data was collected from social media from collaborative work of millions of genealogists we see comparable numbers to what we expect to traditional studies about known paternity, known maternity rate. Next slide, please.

We also used natural language processing to convert the place of birth of these millions of profiles into longitude and latitude and what you see in this slide every pixel over here is a profile in our data and we have such a dense sampling that you can see a map of the western world and we can use this data to analyze migration patterns and maybe transfer onto the next slide.

So, we can do things like quantitative anthropology for instance on the left we can ask where is the love of your life, we can look at distances between the birth location of husband and wife, a function of the year of birth, you can see that in 1700 most of our couples were only 1 kilometer away from each other and for couples that were born in 1950 they had to travel, in most cases, more than 100 kilometers to find their spouse.

And we can also ask who is the love of your life basically how these two individuals, our couples, how they are connected based on the genetic similarity and we see that up to 1850 most of the couples are about 4<sup>th</sup> cousins and then people realized that cousin marriages are...and we see a reduction in our data so right now most of the couples now in our collection are not related to each other.

Now this is just one example of what you can do with this data but in fact it tells you quite a lot about how mutations segregate geographically because mutations pass from parents to their offspring, we can look at migration patterns and right now we're trying to see how much rare variation passed in the population, how many kilometers we need to go to sample such variation. So, out of this batch of distribution of mutations in the population. Next slide.

Another thing that we can do since we have so many different types of relationships on the website we can take a trait, we focused on longevity because we had over 1 million profiles the exact date of birth and the exact date of death so we can measure longevity very accurately for this profile. And we can look at the concordance of longevity between different types of relatives on the website all the way from 5<sup>th</sup> cousins to monozygotic twins.

And then we can look at how basically the slope, the slope of the correlation of longevity, a function of the genetic similarity tells us something about the...of longevity, about the genetics of longevity. So, here from this analysis we can infer that genes determine 20% of lifespan by analyzing this big dataset. So, let's move to the next slide.

So, what we want to do now, what we're working on right now in the New York Genome Center is on methods that we can take individuals, the samples that arrive to the New York Genome Center, and if they are consented for this process we will be able to take this genomic information and place it on this very large pedigree that we created. So, basically if we have this we can integrate the genomic information with the genealogic information and this will create this multiplier. We can now take individuals that were not part of our data let's say deceased individuals we can infer their genome, we have these death certificates that I spoke about and we create an association study between the cause of death and the genomic information of individuals that have no other way that they can get their DNA.

We can also use this data to recruit new individuals based on phenotypes or genotypes. We can use this data to enhance the quality of the sequencing data because we can phase and impute and we can also study parent of origin effects so we are quite excited about this project.

We have standards, we have the way to present a graph this is a GEDCOM file. What we are working on right now are APIs basically codes that allow us to query this pedigree, place the genomic information and infer new information by integrating the genomic and the genealogic information. Next slide, please. So, this is just my acknowledgements for the people that work on this project and our funding. So, thank you very much.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Jon?

**John Wilbanks – Chief Commons Officer – Sage Bionetworks**

Yes?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Yes, sorry, my mute button was stuck, couldn't believe it, sorry. Thank you that was an outstanding set of information and a little different from the other things that we've been hearing today. So, if you don't mind, so Dr. Erlich let me lead with the first question then ask if any of our other colleagues have other questions.

So, to the extent that we're trying to focus on Health IT standards and implementation specifications, one thing that you mentioned that caught my attention was an application program interface to genealogic information. Are there standards or types of data that might be found in electronic health records or other kinds of IT that would be helpful to you in terms of the work that you're doing or ways that you might be able to extend the work that you're doing, you know, for the benefit of all Americans? You know are there recommendations specifically that you make to us either for things that are out there or that you need that aren't out there right now?

**Yaniv Erlich, PhD, B.Sc – Professor of Computer Science – Columbia University**

Yeah, so I think the things that we need basically...so right now most of the effort is to integrate phenomes and genomes. We can propose to have another layer, a genealogic layer. So, in order to integrate the genomes and the phenomes with this layer we need to have identifiers that appear in both datasets. So, the full name of the person, the name before marriage if it is for females, the exact date of birth all of this will help us to created identifiers to merge the two datasets.

So, I don't think that we need new types of information in general but we just need these pieces of information to be present in the other datasets that we try to integrate.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Perfect, present and standardized, sorry, just a bias of where I'm coming from, so that's perfect, okay.

**Yaniv Erlich, PhD, B.Sc – Professor of Computer Science – Columbia University**

Yes.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Thank you for that clarification. Other Task Force members, questions for Dr. Erlich? Okay, all right, so thank you so much for the outstanding presentation. So, that concludes the last of the specific presentations that we've had.

So, now what I'd like to do for Task Force members, understanding that this is the last of our kind of testimony gathering sessions for this time. Are there questions that you want to ask of the presenters or to kind of throw out to us as a group for discussion in terms of what...questions that they raised for you that you didn't get a chance to hear during the presentations or as they apply to the overall broad kind of charge of where we're getting to? Now is the time to kind of, you know, start laying those things out on the table for us for discussion or consideration. So, with that the floor is open.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

This is Leslie, I have a question, we heard from Duke about the American Health Information Community standards for family health history and I wondered if the others presenting today are also familiar with or using those standards?

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay, so Josh, John or Dr. Erlich any thoughts on AHIC standards or other kinds of standards that are being used that you find particularly useful or not maybe?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

I think for implementation of genomic variants we've been trying to use the HL7 standard there to, you know, drive our decision support though we've predated that in implementation and the active form is still our locally developed standard.

**Betsy L. Humphreys, MLS – Deputy Director – National Library of Medicine**

The...if I'm not mistaken, this is Betsy, if I'm not mistaken what the American Health Information Community did was try to promote advancing of the HL7 standards in that area through the Clinical Genomics Workgroup so I think that if you're focused on the HL7 standards in that area then that is the...that was influenced by the AHIC.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

And that is the very standard, yeah, that we are working with and there are members of the team that are...I'm not them, but that have been engaged with that group.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Thank you.

**Yaniv Erlich, PhD, B.Sc – Professor of Computer Science – Columbia University**

Yeah, I think, this is Yaniv Erlich, so we are using right now VCF files for genomes and for genome variants. One thing that we are exploring is a new method of storing a large number of genomes and compressing them and to do very fast searches, this is the GQT toolkit developed by Aaron Quinlan and it allows basically if you have tens of thousands or hundreds of thousands of genomes to store them very efficiently and to do searches to find specific variants where right now we are exploring this toolkit and we are quite...it looks pretty good so we think that we will adopt this standard.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Great, other questions or comments, or for that matter, I opened it up to Task Force members, do our presenters have anything also that they might want to add either on top of their presentations or in response to the other presentations?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

I had a question for Ricky, this is Josh.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Go for it.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Ricky, in terms of the communication from the EHR and then back at MeTree, and then back into the EMR, what exactly...how much is being exchanged in the data that actually gets back into the EHR? Is it kind of a summary? Does it look like a note that would just be in a normal note or does it also include all those graphical representations of the pedigree and tree, and table views?

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Great question. So, we haven't done that piece of the implementation yet. Right now our initial implementation plan is for read only since that's what DSTU 1 of FHIR supports DSTU 2 will support reading or writing back into the EHR, but obviously there are a lot of factors there when you're writing back in, in terms of the data standardization so we need to be careful.

But we can generate reports and we can save those reports to our on-base file system basically so those can be accessed by the provider and we can do that in PDF so we can capture some of that if we want.

The plan right now is simply to allow the physician to see the trees right within their desktop environment so that data is pulled live and they can view the information, the same thing for the patients they have a live view that they can edit and add information to and then the provider will see that same view.

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

The data that you're getting out of the EHR doesn't include...the family history data is all independent correct? So, it's all just within MeTree, so is it just kind of patient registration kind of information that you get out of the EHR?

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Yes, so the first phase, exactly, so the first phase MeTree has a separate database that stores the family history data because it does it in a bit more detail and we're using the FHIR patient resource basically to pull the demographics and to match the patients but the plan is once we add write support to actually take the information that's been contributed by the patient and include that using the FHIR family history resource and include that within the EHR in a structured way that would be the next phase. So, we would like for that to happen and we would like for the EHR to be that single source of truth so we're not maintaining multiple sets of data.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Are you using the InfoButton call? It looked like that in the examples you cited for the patient education.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

That would be a question for Lori. I'm not sure if that was initially designed to use the InfoButton it looked like it could have been but I'm not sure if they did that or if was just a one-off to Medline.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Thank you.

**Richard A. Bloomfield, Jr., MD – Assistant Professor of Pediatrics & Medicine – Duke University School of Medicine**

Sure.

**Andrey Ostrovsky, MD – Chief Executive Officer – Care at Hand**

Can I ask a quick question? This is Andrey, actually two quick questions and one is around, have any of the speakers been able to take the data that you all are collecting, processing and gaining insights from and tie it in any way to measuring quality or existing quality measures? I recognize that quality measures are emerging.

And the second question is, have there been any either challenges or best practices you've identified in trying to scale the use of these insights to clinicians in your institutions and in particular I'm coming from the perspective of a recent trainee, there is so much already being forced down our throats to learn that quality improvement is already a hard thing to learn let alone like how to actually meaningful use this type of information. Any thoughts on those two questions?

**Joshua C. Denny, MD, MS, FACMI – Associate Professor, Departments of Biomedical Informatics and Medicine – Vanderbilt; Office of the National Coordinator for Health Information Technology**

Yeah, I can speak to both of those points. In terms of quality metrics we did implement one of the quality measures, CMS 30, which is statins after discharge, use of statins after discharge from myocardial infarction in an ETL language called NIEM and its constant information modeler and then we shared that across a couple of different institutions and were able to show that we could share that and execute across different sets of logic. We actually had a transformation module that took the quality data model format and then transformed it into NIEM as an executable engine.

The second point is some of these phenotype algorithms we have used for quality improvement locally at Vanderbilt, the notable example, and maybe the only one today, is identifying diabetics which, you know, don't always have the problem list entry noted or it may not always get billed as having diabetes depending on which...if they're not going to primary care clinic at Vanderbilt, what we've done is used our diabetic algorithm to identify those who have diabetes and then fed them into our chronic disease management program so that they could sort of...so we can look at quality management and connect through that system sometimes then partnering back with primary doctors and things like that too.

**Andrey Ostrovsky, MD – Chief Executive Officer – Care at Hand**

Thank you.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

So, this is Michelle, we have five minutes left, I'm not sure, I think maybe we should probably wrap up. I'm not sure...Jon's line had dropped I'm not sure if he got back on yet, if Jon or Leslie want to make any closing remarks before we close up for the day.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Can you hear me?

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Yeah, there you are.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Excellent, thank you, sorry about that. You were reading my mind because what I was going to say is that it's 1:24, but now it's 1:25 so we're actually on schedule. This is actually the public comment time. Michelle do you want to do public comment before we make wrap up comments or after?

**Public Comment**

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Yeah, that would be great actually. So, Lonnie or Caitlin, can you please open the lines?

**Lonnie Moore – Meetings Coordinator – Altarum Institute**

If you are listening via your computer speakers you may dial 1-877-705-2976 and press \*1 to be placed in the comment queue. If you are on the phone and would like to make a public comment, please press \*1 at this time. Thank you.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

So, while we wait for public comment now we can make any wrap up comments.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Excellent, Leslie, the floor is yours, please.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Just I guess I would like to encourage the group if you haven't seen any of the attachments, the JAMIA article, that were provided to us I think it provides a good background on SMART on FHIR for genomics and I would also reference anyone listening on the call to read that as well. This has been a great day of presentations from various points of view and I appreciate all that we've heard. Thank you.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

And the only thing that I'd like to add is that we are going into now the initial recommendation writing period so Task Force members please be thinking about recommendations that you think we might want to make back to the full Standards Committee. If you'd like to share those by e-mail with Leslie and myself, Mazen, and Michelle we can start gathering that information, we'll certainly work on it as well but it would be great to start getting your thoughts and contributions either in written form or if you want to have a phone call to discuss them we're happy to do that too.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Jon, this is Michelle, since we have a couple of more minutes until the end of the hour or the half hour, I just wanted to check in, we don't have another call until August 31<sup>st</sup> there maybe vacations and other things that we need to schedule around but it might be good to provide an update to the Standards Committee next week about where we are and how things have gone. So, I'm just wondering if it would be possible to schedule a quick call before the Standards Committee to kind of put together some type of general briefing for the committee and then we'll work on finalizing recommendations after that.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

I'm certainly fine with that. Leslie are you on board with that?

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Absolutely.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Okay.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Okay, so we'll work on that.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Yeah, I'd be excited to report back to them actually.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Okay, thank you.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Do we have any public comments?

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

We don't.

**P. Jonathan White, MD – Acting Deputy National Coordinator – Office of the National Coordinator for Health Information Technology**

Alas we have stunned them into silence. All right, well, very good. Well, in that case, again, thank you everybody so much for your attention and we look forward to our future iterations of this coming up soon. Thank you.

**Leslie Kelly Hall – Senior Vice President of Policy – Healthwise**

Thank you, bye-bye.

**Michelle Consolazio, MPA – Federal Advisory Committee Program Lead – Office of the National Coordinator for Health Information Technology**

Thank you, everyone.

**M**

Thanks, everybody.

**M**

Thank you.

**W**

Bye.