



**Health IT Standards Committee
Precision Medicine Task Force
Final Transcript
February 12, 2016**

Presentation

Operator

All lines are bridged.

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

Thank you. Good morning everyone, this is Michelle Consolazio with the Office of the National Coordinator. I'm sorry for the public for the late start of the call. This is the second charge for the Health IT Standards Committee's Precision Medicine Task Force. This is a public call and there will be time for public comment at the end of today's call. As a reminder, please state your name before speaking as this meeting is being transcribed and recorded. Also as a reminder, if you are not speaking, if you could please mute your line that would be appreciated. I will now take roll. Leslie Kelly Hall? Andy Wiesenthal? So there are a few new folks and I don't want to butcher your names, so if you could let me know how to pronounce your name if I say it wrong that would be appreciated. Gil, I don't believe he is on. Dixie Baker?

Gil Alterovitz, PhD – Faculty, Biomedical Informatics – Harvard Medical School; SMART/FHIR Genomics Lead

(Indiscernible)

Dixie Baker, MS, PhD – Senior Partner – Martin, Blanck & Associates

I'm here.

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

Oh you are; Gil, can you pronounce your last name for me?

Gil Alterovitz, PhD – Faculty, Biomedical Informatics – Harvard Medical School; SMART/FHIR Genomics Lead

Alterovitz; Gil Alterovitz.

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

Alterovitz. Thank you. And Dixie Baker is on. And Mary Barton? Steven Keating?

Steven Keating – Patient Advocate/Consumer – Doctoral Candidate, Mechanical Engineering, MIT Media Labs

Hi yup, thanks.

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

Hi, Steven. David McCallie?

David McCallie, Jr., MD – Senior Vice President, Medical Informatics – Cerner Corporation

Hi, I'm here.

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

Hi, David.

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

Matthew Might?

Matthew Might, PhD – Visiting Associate Professor - Harvard Medical School; President – NGLY1 Foundation

Here.

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

Hi, Matthew. Andrey Ostrovsky?

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

I am here.

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

Hi, Andrey. Ketan; is Ketan on? Eric Rose?

Eric Rose, MD, FAAFP – Director of Clinical Terminology – Intelligent Medical Objects

Yes, here.

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

Hi, Eric. Joyce Sensmeier? James Breeling? I believe he is on.

James Breeling, MD – Director, BioInformatics, Office of Research & Development – Veterans Health Administration

I'm here. Yeah, I'm here.

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

Hi, James. Josh Denny? Christina Heide? She had let us know she would be a little late. Betsy Humphreys? Mitra Rocca?

Mitra Rocca, PhD – Associate Director, Medical Informatics – Food & Drug Administration

I'm here.

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

Hi, Mitra.

Mitra Rocca, PhD – Associate Director, Medical Informatics – Food & Drug Administration

Hi.

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

And Jon White? Did we get Leslie by any chance? Okay.

Marvin

Yes, she should be dialing in any second; this is Marvin.

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

Okay, thank you. So first of all, let me thank you all for joining us for this second charge of the Precision Medicine Task Force. We do have a pretty packed agenda, so I do want to get started as soon as possible. Our chairs...so Andy is in London today, which is why he is not on the call and something must have come up with Leslie, so I'm sure she will be joining us shortly. But let me just again thank you all for joining. I am going to turn it over to Maya to introduce some of our...to go over the agenda and introduce some of our speakers for today.

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

Sure. Thanks, Michelle; good morning everybody. So we have a few new members, Gil, Dixie, Steven and Ketan. Do you guys want to introduce yourselves really quickly and just say a little bit about your background, very briefly?

Gil Alterovitz, PhD – Faculty, Center for Biomedical Informatics – Harvard Medical School; SMART/FHIR Genomics Lead

Sure. So hi, I'm Gil Alterovitz. I'm an Assistant Professor over at Harvard Medical School and Boston Children's Hospital; my area is in informatics in terms of how do you communicate clinical genomic information.

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

Thank you, Gil. Dixie? Is Dixie on?

Dixie Baker, MS, PhD – Senior Partner – Martin, Blanck & Associates

Yes, I'm on. I'm Dixie Baker, let me see, in the matter, I've been a...I was a member of the HIT Standards Committee for over I guess 6-1/2 years or something like that. At any rate my involvement with PMI, like Gil, we're both involved with the Global Alliance for Genomics in Health and most of my...I work for Martin, Blanck & Associates, which is a healthcare consulting company. Most of my consulting is around the integration and use of genomic data.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Hi, this is Leslie, I'm so sorry; I couldn't get the audio to work.

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

Oh, it's no problem, welcome, Leslie. We were just allowing our new members to introduce themselves really briefly and then I'll turn it over to you.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Super.

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

Steven, do you want to go ahead and say a couple of words?

Steven Keating – Patient Advocate/Consumer – Doctoral Candidate, Mechanical Engineering, MIT Media Labs

Hi, this is Steven Keating. I'm speaking from a patient experience; I've just recently gone through a crazy brain tumor experience actually found through research study data. And I'm going to be speaking about, or at least have background in data access from a patient side and I'm also a Board Member of the Personal Genome Project. So I'm happy to be here and thanks for the opportunity.

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

Thank you. Do we have Matt Might?

Matthew Might, PhD – Visiting Associate Professor – Harvard Medical School; President – NGLY1 Foundation

Yes, I'm here. So I've got three hats; I'm a visiting Associate Professor at the Harvard Med School in Biomedical Informatics. I'm ordinarily a Professor in Computer Science at the University of Utah and I'm also on the patient's side by virtue of the fact that my son has an ultra-rare disease.

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

Thank you, Matt. Ketan? Did we get him? Sorry. Well, perhaps he will join us maybe later in the hour, but Leslie, if you wanted to go ahead and take over and take us through the agenda; I think we're on slide 3.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Super. Thank you very much; I really appreciate all of you joining. Many of you have been involved in the task force earlier and I know as the introductions, you probably heard that we have Gil, Dixie and Ketan joining us to this group and thank you so much; we really appreciate your expertise.

Well we have been given significant challenges. The PMI Initiative is off and running and we expect to have 80,000 people approximately signed up by the end of this year and they'll be participating with data consents in place and eager to have data flowing to support this lifesaving initiative. The Vice President has given us opportunities to shoot for the moon again and he reminds us that innovation and creativity and speed are all necessary to do what seems impossible possible.

The data movement to support this initiative is going to happen and is going to happen either through chaos or some sort of standards that can help facilitate the movement and reach scale; and that's our job and this business case new to us in the standards group. This is really about not...a business problem that's been given to us with a time factor associated with it with a group of patients in need, and it's our job to help facilitate and provide rate recommendations for this data movement. So thank you.

So as part of our initiative in the first effort, we determined that there were three goals and that the ONC had a role in the Precision Medicine Initiative to support these three goals: To accelerate opportunities for innovation and collaboration around pilots and testing of standards that support health IT interoperability for this effort. Adopt policy and standards to support privacy and security for the cohort participant data and advance standards that support the participant-driven approach to patient data contribution. Those are...that's a lot. Next slide, please.

So we made recommendations back on October 6 and we knew that from these presentations, we had a new challenge; that the EHR was likely to capture a lot more than just the genomic data challenge we had, but also phenotypic data, patient's data generated and collected, such as problems and medications and allergies. And the core problem we have to solve is that we don't necessarily have data standard model for EMR to capture and to be as responsive for many types of data and information.

We looked at different standards based on the testimony received and we divided these up into these categories: Readily applicable, which we categorized as green, promising standards as yellow, gaps which are red and then accelerators that could help to promote and advance standards. And this idea of acceleration is one that we felt is very important so that we can identify what pushes need to happen to make these promising standards or available standards or even to help identify gaps and move those forward. So we made a recommendation to the Standards Committee in the form of a grid, which we are now revisiting and fine-tuning to present back to this group. Next slide, please.

So our charge in September was to identify opportunities for innovative collaboration around pilots and testing, recommend existing standards, identify emerging standards, identify gaps; so phase two, which is our charge now, is to identify opportunities for ONC to support our federal partners PMI efforts and the related health IT interoperability challenges including the National Cancer Institute, the FDA, the NIH and the Department of Veteran Affairs.

We also want to identify opportunities for ONC to collaborate with industry and pilot the use of standards to enable data donation and patient access through APIs using standards such as FHIR and OAuth 2. And identify standards for use cases to support interoperability as data types that are critical to PMI-type research and prioritize piloting the exchange of this data type based on phased approach that would incorporate most structured/coded data first and add additional data sets in subsequent pilot

phases. So the crawl, walk run approach to this; what do we need as a minimum and how do we get that going? Next slide, please.

So one of the things we heard is there are a bunch of use cases associated with this and that each of these areas might represent different standards. So we have the lab to the EHR, the patient-generated data to the EHR, the EHR to research and the lab directly to the patient. This is a challenge because each one of these areas might represent completely different vocabulary needs or data structure needs, but it is incumbent upon us to look at each of these areas and make recommendations. And it will definitely be a challenge. We heard a lot already and I think we're all up to it. Next slide.

So our work plan is pretty aggressive. So we start today with some great testimony from NCI and the FDA and ONC. We are also going to, on February 26, get further information from NIH and VA. And you can see through this work plan we've got considerable amounts of information to take in and hopefully to be able to discuss and make final recommendations in a pretty aggressive timeline.

We're allowing for time for discussion in the meetings, so please feel free to ask questions. I know for me this is a very daunting amount of information and detail and it's important that each one of us ask questions and understand the testimony given. Next slide.

So the questions we're asking the presenters are: What is the agency's specific mandate and contribution to the PMI Initiative? What data types are the highest priority? We want them to share their challenges or questions related to health IT standards, interoperability and data exchange and here are just some of the examples. What existing health IT standards are you using in your projects? Are there challenges? Are there existing standards you might consider working with? Are there areas where you're exchanging data, but robust standards don't exist? And what pilots might be launched to enhance technical capabilities related to the data standardization, data donation or transfer?

So we've asked some pretty good questions of this group...these groups and I hope that we will have a wonderful amount of input to use for our recommendations. Next slide.

So today we will be hearing from again NCI and hearing more about the status of the current initiative and then from the FDA we'll get the update on computable privacy, which has impacts I think for beyond this project. And then Q&A, public comment and adjourn. So with that, I think, next slide, please. We'll turn it over to Tanja Davidsen, from the National Cancer Institute. Welcome, Tanja.

Tanja Davidsen, PhD - Biomedical Informatics Project Manager – National Cancer Institute

Thank you, thank you very much. Great. So today I'm going to be talking a little bit about The Cancer Genome Atlas, TCGA, the Genomic Data Commons, the GDC and then the Cancer Genomic Cloud Pilot. Next slide, please.

So here are some brief answers to the questions that you guys have asked of us. I also have a more detailed slide at the end of my part of this presentation for your reference. But just briefly, since our time is short, the NCI is responsible for the PMI Oncology component. We have four main areas; first creating new adaptive clinical trials, which will incorporate genomic and molecular diagnostics. Second, creating novel biologically and clinically relevant pre-clinical models for cancer; that would be things like xenographs and third, understanding the methods of resistance of tumors to therapy. The final component is building a cancer knowledge base and the NCI Cloud Pilots and the Cancer Genomic Data

Commons are two specific activities that will build this framework of the knowledge base; and that's what I am going to spend most of my time talking about today.

Which data types are our highest priority? We have a lot of priorities. We have imaging data, lab data, molecular characterization methods; so that would be DNA and RNA sequencing, gene expression, copy number analysis, methylation, etcetera. We're also interested in pathology and radiology labs as well as clinical labs, detailed therapeutic data, diagnostic information and outcome results from cancer treatments.

And the challenges that we have; so EHR interoperability is not a current focal point for PMI Oncology, but we do have barriers to effective data exchange, especially data access, which I will be addressing somewhat in this talk today, and especially to controlled access data. So controlled access data is data that might contain personally identifiable information, PII, something like germline variants from DNA sequencing would be a good example. That data needs to be controlled and it needs to be appropriately protected, but we also don't want to put up undue barriers to the researchers.

So some of the things that we feel we must address include investigator training, possible improvements to the approval process and for a longer term solution, more consistent and understandable patient consent forms. Next slide, please.

So before I talk about the two projects, the Cancer Cloud Pilots and the GDC, I first need to introduce the data that's going to be in these resources. So the Cancer Genome Atlas is what I'll be talking about. Next slide.

So this was launched in 2006, a joint effort by NCI and another NIH institution, NHGRI. It's a complete characterization of approximately 33 adult cancers; that includes 20 common cancers at 500 patients or cases each and 13 more rare cancers at a lower number because of accrual issues at 50-150 cases or patients each. We do copy number, gene expression, methylation, DNA sequencing which includes both whole genome and whole XM and we collect clinical data for all of these cases, and it's a total of about 11,000 cases.

The project is ending this year, so we do have most of the data at this point, but future projects are going to use the TCGA infrastructure that was built. Some of the projects that will be using the TCGA infrastructure include Exceptional Responders, ALChEMIST, the Clinical Trial Sequencing Program and the Cancer Driver Discovery Program. And to learn more about TCGA, you can go to cancergenome.nih.gov. Next slide, please.

Just briefly this gives you an idea of how much data TCGA has generated. We have 11,000 patients both with a tumor and a match normal sample, tissue sample. We have 12 data types, 27 different technology platforms; as the technology has improved, the platforms have changed. We have different data levels from the raw data all the way up to the fully analyzed data and finally, we have to do data updating when we have different versions of, for example, the human genome coming out. Next slide, please.

So first I'm going to talk about the Genomic Data Commons. Next slide. The rationale for the Genomic Data Commons is that TCGA and many other NCI-funded cancer genomics projects that are similar to TCGA each currently have their own data coordinating center or DCC. What this means is it results in inconsistent data across the NCI. Because different DCCs are collecting the data in different ways,

there's a duplication of effort and a waste of money. In addition, our BAM files, which are the raw sequencing files that we collect, are stored in a different repository than all of the other data. It's very confusing to users; the users have to go to two different websites to get all of the data across, for example, the TCGA Project.

In addition, again this is a barrier to access data, right? In addition, the two different websites have two different user name and password combinations that you need to use, for the same projects of TCGA. So this is again, this is something we want to fix.

So the GDC is the solution. It will be a single repository for all NCI cancer genomics data. It's starting off by including TCGA and two other genomics projects from the NCI, but it will eventually expand to the new, upcoming NCI cancer genomics effort that I mentioned in a previous slide and users will be able to upload their own data, their own cancer data to the GDC as well.

The GDC will harmonize and standardize the data when at all possible; for example, they're realigning to the newest human genome standard currently, and they will do so again when another version comes out. And they're also recalling all variants using a standard calling method, because many of these projects have used different variant calling methods. Finally, this will be the authoritative reference data set for these NCI projects. Next slide, please.

We're also employing metadata standards from the Cancer Data Standards Registry and Repository or caDSR, which is here at CBIIT, which is where I work. This is our first step toward development of a knowledge system for cancer. The project was started in the spring of 2014. The University of Chicago with the PI, Dr. Robert Grossman has the contract and our go live date for the GDC is late Spring of 2016, so we're anticipating late May, early June timeframe right now.

Just note that this data is not collected onto a commercial cloud, there is a private cloud at the University of Chicago where all the data is hosted and all of the data here will be freely available for download. And I alluded to it on a previous slide, but I just want to reinforce that we won't have two places to go for people to get the same data...or, the full data from a single project; everything will be in one place, there will be one login. Next slide, please.

And next slide. So next I'm going to be talking about our cancer pilot. So again, I'm going to give the rationale for the project first. The standard model of computational analysis is that the user, who is represented in the middle of the screen, is having to download all of the public data to their local area. They need to download publically available software to their local area and then locally they will also, of course, have the data that they've generated or the software that they've generated, but everything needs to come together in one place in order for a researcher to do their analysis. Next slide, please.

This has been working fairly well, up until recently; however, now the data is getting so large that it just can't be done this way anymore. If we assume that the TCGA completed data set will be about 2.5 petabytes, the storage and data protection costs can be as much as two million dollars a year for that data size. In addition, downloading all of the TCGA data at 10 gigabits per second would take 23 days, and I can tell you that we never actually get 10 gigabits per second. Right now it takes about two to three months to download all of the data.

Only large institutions have the ability to utilize this data. Smaller research labs simply cannot get access to it. And finally, this data is going to continue to grow at an increasing rate as we have new projects and more funding. Next slide, please.

So our solution is to co-locate both the compute and the data in a commercial cloud. We have the computational capacity in a commercial cloud; we have the space for all of the data. We are putting all of the core data for TCGA as our pilot project into these cloud pilots. The users can then also upload their data or their software into the cloud and process everything there. What this avoids is now people won't have to download the TCGA data locally. They'll have to upload their own data, but that's likely to be much smaller than the full TCGA data set. Next slide, please.

This slide shows the GDC and the cloud pilots and how they work together, so we have data generation on the left-hand side from the different projects. Then you have QA and QC validation and aggregation to come up with the final authoritative NCI data reference set. And this is hosted at the Genomic Data Commons; this will be hosted and available to the public later this year at the GDC.

This data is also being copied into the NCI Cancer Cloud Pilots represented by the yellow box on the right-hand side. The user is represented in the lower right-hand corner and the user would use these two resources for different things. If they wanted to search, retrieve or download the data they would go the GDC represented in the blue box in the middle of the screen.

However, if they wanted to do analysis on this data, high compu...high performance computation, instead of going to the GDC where they would just be able to download the data locally, they would go to the cloud pilot. At the cloud pilots they'd be able to bring their own software and data to put it together with the TCGA data set and run everything on the cloud pilot. So we have slightly different goals for the two projects. Next slide, please.

So our goal again is to democratize the access to the NCI genomic data and the other associated data because right now it's difficult for small labs to get access to this data and actually run their pipelines on it. The project is managed through CBIIT, where I work, in partnership with the Center for Cancer Genomics, the CCG which generated the TCGA data and is managing the Genomic Data Commons. We are working together closely with the GDC so that our goals are aligned.

As far as standards, we are using the Global Alliance for Genomics and Health, the GA4GH standards. They have four groups of standards; clinical, data, regulatory and ethics, and security. I've got a link to their site. We in particular on the cloud pilots are...two of our three cloud pilots are using the Common Workflow Language for workflow generation and all three pilots are using Docker containers. The period of performance for these cloud pilots is September 2014 through September of this year. Next slide, please.

We have awarded three cancer genomic cloud pilots; the first to Broad Institute; Gaddy Getz is the PI. It's working on the Google Cloud and the goal of this is to put a Broad pipeline called Firehose into the cloud. The second is at the Institute for Systems Biology; Ilya Shmulevich is the PI; also sit on the Google Cloud and their focus is on the interactive visualization and analysis of the data. Third is Seven Bridges Genomics, and I apologize, there appears to be a small typo there, no "K" on the end of Genomics. The PI is Deniz Kural; they are working on Amazon Web Services and they have over 30 public pipelines already in place. Next slide, please.

This is our project schedule. You can see that we had a Phase One and Phase Two for design and build, and we are currently where the green arrow shows in the Evaluation Phase. Next slide, please. And next slide.

And so here are some of the NCI sponsored evaluation activities that we're doing. We are doing independent testing and evaluation of all three cloud pilots; highly structured tests of functionality, security and load capability. We're trying to assess the key strengths, weaknesses, issues and risks associated with each of the three pilots. And I should mention that we chose three pilots because we wanted three different approaches to this problem. So we're hoping to find the best combination of solutions.

We're also funding eight supplemental...administrative supplements; eight proposals are being funded on active NCI grants to utilize the cloud pilot for a purpose. We also have a DREAM Challenge that we're launching in 2016. We're using the TCGA RNA-Seq data to identify somatic mutations. We're asking the public to create solutions to this, again using our cloud pilots.

We have two hands-on workshops, one on May 24 and one on May 25 at the NCI Shady Grove Campus where you can learn more about the cloud pilots and really do some hands-on work. And finally we have an NCI Intramural Evaluation where we've dedicated resources to support intramural investigators and their adoption of the cloud pilot.

Finally I do want to mention if anyone happens to be going to the AACR which is in New Orleans this year, April 16 through 20, we do have two NCI-sponsored sessions, one of which where the GDC will be discussed and the second where the Cloud Pilots will be discussed. That will not be a hands-on workshop, however. And finally, on April 15, before the meeting, GA4GH is having a pre-meeting, a full day pre-meeting; so if you're interested in that, I would be happy to provide some details. Next slide, please.

And then this is just a summary of the project team and again, I have put the URLs for the three cloud pilots, so if you're interested in learning more, you can go there. All three of our cloud pilots will be launching in the next month or two; so, you can go to these URLs to get more information.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Super. Thank you so much Tanja, we really appreciate that. We'll be having our questions at the end of the presentation so, keep notes folks and let's move on. So our next presentation we have is through the FDA.

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

Thank you so much. Dr. Litwack, do you want to start?

David Litwack, PhD – Policy Advisor, Center for Devices and Radiological Health – Food and Drug Administration

Yeah, sure; thanks. I'll just give a brief intro and then let Dr. Kass-Hout speak. So just to set the stage for what you'll hear about today; so the work that FDA has been doing under the PMI is focused on the regulation of genomics because we're FDA and we regulate medical products and in vitro diagnostics, we've had that focus. Because we have an interest, you know, our mission is to protect public health while at the same time allowing innovation, in this case really advancing technology that is driving

advances in precision medicine. So we want to make sure those tests are safe and effective and that they are...that that is assured in a way that permits innovation.

The, you know and our challenge here is that the generation of big data by next generation sequencing and genomic tests, this is critical for precision medicine, but it also poses challenges when people who develop genomic tests try to validate them; in other words, just try to make sure that their tests give accurate, reliable and meaningful results. This is requiring not just new policies, say new regulatory policies, but also new technical and scientific approaches that will make it easier to develop, regulate and use these tests.

So for the PMI, we at FDA have been developing new regulatory approaches that will rely on the development and use of standards. And while there are many pieces to this work, we want to focus today on one of FDA's solutions, a platform that we've developed and opened for public use that we are calling precisionFDA. So I'll now turn it over to FDA's Chief Health Informatics Officer, Taha Kass-Hout to talk about what precisionFDA is how it works and how we believe it will advance precision medicine. All right, so it's all yours.

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

Thank you, David. Well first, I appreciate the opportunity for us to present on our work. So as David mentioned, today I'll be talking about the precisionFDA platform which is supporting FDA's role in the PMI Initiative. And I'll also be joined by Ms. Elaine Johanson, who is directly managing the project. Next.

So we set up the platform on December 15, so it's about two months ago, with the goal that we want to help researchers evaluate next generation sequencing software or technologies that ultimately will be used in diagnostic tests. And the way we're doing this is through a crowd source approach. Next slide, please.

So to follow on what David was trying to convey about our role, throughout the presentation we'll focus on these initial sort of use cases; for example, how can you assess the consistency of these tests. For example, let's say I have a bio specimen like a blood sample and to run a genome test, if you go to a lab A, you get your results, can you repeat the results on another instrument or to another lab and get comparable results?

The other use case is to assess the accuracy. So as you heard from the prior presenter, there are...genome, you know, material available to the public and to these researchers and test developers, we want to make sure that if you introduce a new software, run the same specimen, can you compare to a reference genome and also show where you match or you have disagreement and further enhance the argument there.

And then also introduced through the precisionFDA platform this comparison feature, which I'll go through to also assess the agreement on how you can evaluate that agreement. For example, if there are things that you can further QA or...other areas where you know you need to go back and update your software, your methods or different technologies in a full and transparent way?

And finally, is also assess the efficacy of the software that you have...has on the market. We have various simulation tools that we do benchmark everyone using the same sort of input and then as that input goes through the software, then you can compare the output of that against a prior sort of well-known

standard. And this is particularly interesting in case that there is no sort of gold standard or there's no reference material that you have to go against. Next slide, please.

So these are the areas of the features currently we offer on the platform, and this is a platform which we will constantly be adding new features or new modules. But primarily is working in this environment, we want to first assure the entire community that they have private areas where they can work independently of one another, and we do meet various standards anywhere from data standards or exchange standards; many of those that apply to the Global Alliance for Genomics and Health or various different security or CLIA for labs or HIPAA regulations that we have to ensure that that environment doesn't violate any of their certification.

But there is also common area; this is really where encouraging sort of the use of the proxy model of sharing with the community. But then we have four core components; so it's about the data, Apps, comparisons and notes, which I'll go through each one of those next. Next slide, please.

First of all we have a, you know, you have to register to come on to precisionFDA. The registration is fairly simple, but we do stress on two-factor authentication using the combination not just username and password, but also some kind of unique combination. And for our instance here, we use a 6-digit code that is unique to each individual that changes every 30 seconds, and I'm sure you've seen this before. Next slide, please.

All the data is encrypted, by the way, on the platform or during exchange, so I'm going to take securities really, really seriously here so...and we do exist in a public cloud environment that has been authorized to meet the federal standards as well as FDA's standards. When the user logs in, we do present them with...guidelines and understand the terms of use of this information and environment as well as the intended use for this environment, primarily at this point as a sandbox, just a sandbox but ultimately this could...can be used as part of regulation downstream, but that's something we're working together with the community on achieving.

And the whole neat idea here is how can we achieve common standards for evaluating software that grows out of the community rather than just like here's one set of regulations, the minute you put it out there goes out of date, given that software is constantly changing. So we're using standards-based or analytic benchmarking. Next slide, please.

So the first area is the data or the file; this is primarily, and you can see here various groups whether it's the precisionFDA team, the Broad Institute, NIST, various academic institutions, PhRMA has started to add reference genome material or reference software to the site, and those are what we call generally a file. And each one of those files have various asso...metadata associated with them that the user will add those.

We also indicate on the right, for example right now I'm seeing just what's being shared with everyone, so that's why I see all public. But if I'm logged into my own instance, I also see files that just...that only me or my group can see. You can also see here that we work in a sort of ecosystem so we allow a lot of entrepreneurs and folks that have the computation power to be able to upload the data, so this is their private areas or the common area, so also allow this sort of, you know, we're in one cloud environment, but we can talk to other cloud environments or other data centers...this is sort of the future way to go and fetch...go to files or call our APIs and be able to sort of interact with precisionFDA. Next slide, please.

So I mentioned the second feature which is comparison. So the basic idea here is you could select, you as a user can select what is it that you're testing and then you can define also the benchmark. And the benchmark can be, in this use case, a prior known, well factorized human genome that was...in this case by NIST or by other institutions that have worked on sort of analyzing or assembling the gene.

In this case I'm also, I have run the same human genome in my own software and then produced results, so what I'm doing right now is comparing my results against a standard that I selected to be the benchmark. The benchmark can be I'm comparing my results to you, the benchmark can be, I'm comparing my results to a simulation data set or it can be, you and I are testing something that we don't know really what's the absolute truth to that, but we just want to try to sort of benchmark all software against each other.

So here you have the option about selecting the result set for the test and the benchmark; you can also select which regions of the genome that you're trying to target. If, for example, you say you're trying to compare an exome against a benchmark that has a whole genome but you're only trying to focus on high confidence calls that you made in your software, and that's the only area that you're interested in, you're not interested in the whole genome. We also give you that option as well. Next slide, please.

So after you run the analysis, then we present you with the results of the comparison; this looks like a really awesome comparison; you have a precision of 99.3%, recall or sensitivity of about really high 99.1%, also an F-score about the comparison. You can look at the total of sort of a true positive; this is where the two samples have agreed to.

They also see what there is false positive and false negative and even though you have such high confidence now in...or high precision and high recall, you still have tens of thousands of areas that the two samples didn't agree to, and this is really what makes this interesting and also challenging. You know if you want to assure the accuracy of the test, for example comparing it to a reference material or...the reproducibility where you run the same sample multiple times, you want to make sure that you can decrease the false positive and the false negative to something that the community can agree to. But also we give the option now at this point to the user to further drill into, you know, the results.

So the next slide, if you will. Here we allow the user to go down all the way to the base level of the...and be able to see what, you know, why they had such a discordance. For example, in this particular instance, I'm looking at a false positive in this area and the reference or the benchmark material had...however, which is an AT, however my stuff was only able to get an A; and again, you can go back and say my results are accurate or you can say, this is where I need to go back and further update my software and rerun the comparison and the cycle...next slide, please.

This is, so this is now an App; sort of consider that almost like an App market. A lot of these materials exist in the public space, but...thinking about it, it now exists on one platform, they can just easily grab and run. But also what we've seen from the community over the last several months, new Apps that start to come out basically folks have a different software or different material and that will still be scanned, you know, different ways about...or components of the pipeline anywhere from sequencing to reading the or read mapping or the variation calling, etcetera, including one that was given to us by Intel to do high performance computing in a cloud environment, how we can...process.

So you can see how the community starts adding components and packaging them on this platform as Apps. And this is very similar to an App market on your smartphone where you can see how these Apps

can be developed by various developers and then offered to you to use. And we also respect licensing issues so if one of these Apps have licensing behind it, is that the terms of use of those are also respected and provisioned...next slide, please.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Doctor, I'm sorry to interrupt, this is Leslie and we have to be mindful of our public comment time, and I'm so sorry to cut you short. I need to give you about three minutes so we can get in some questions and also public comment. Thank you.

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

Yeah, no problem. This is...thank you. So this is a simulation software that was developed at FDA and now is provided in the cloud environment for the first time for the community to use. This is...the goal of this, how we can simulate a human genome so that we can run it in the pipeline and also help with the benchmarking. Next slide, please.

One thing that's really important here is we can track the provenance of all the work that's been done, so every user for every experiment to give them a precise sort of recipe for all of the components that went into designing an App or running an experiment. And you can see over here, not only all the tests but also we would highlight which ones of these have been shared publically in green, which ones are private and then here the options are for the user to modify that. Next slide.

This is another file that was added by a PhRMA institution to help simulate a human genome by looking at NIH's ClinVar mutation database. This is another flavor about how this simulation can happen. Next slide.

And then the notes section of this is really where...allow the users to document the experiments and share with the rest of the community, so that way others can replicate this, full transparency and also for reproducibility of results. So you can...you'll be able to go back and track every step with each note. For example, if you go to the next slide, I'm going to highlight one note.

For example, this one note was shared by an academic institution about looking at the coverage of human genome compared to...by using a reference out there and not only has she documented the steps but also she attached all the files that were associated with the experiment, whether an input or an output. Next slide.

This is one of the outputs that looks at the human reference genomes against which bases were not covered by the pipeline, and you can see a few outliers over there. Next slide.

Here is the entire App. So you can see...now, let's say someone else is looking at this App and they can be able to fork it, almost like we do on GitHub. So I'll be able to borrow this code, further enhance on it; I can either contribute back to the community or I can incorporate it further within my software and everything is logged to show provenance. Next slide.

Along the way we have very rich documentation on the site, whether you follow step by step doing this, we reference the documentation or only actually here you can go back to the documentation and have very fully interactive documentation along with examples in the "How" section of precisionFDA. Next slide.

Finally, last month we offered the entire code on Open Source on GitHub, so if you go [GitHub.com/fda](https://github.com/fda), you'll see the code there. Next.

And this is just a sample of participants that we have available currently on the platform. You can see from government agencies to institutions to societies to even individuals and industry from entrepreneurs to software developers to instrument providers and even PhRMA and the like. Next.

Here's just one quick shot of where we are right now, so as of today, we have over 900 members joined worldwide, representing 430 organizations. And you can see about one-third of those are from outside the United States. In the United States about half of the users are either in Silicon Valley or in the Biotech Valley. Next slide.

So we look forward to you joining us. It's really easy, go to precision.fda.gov and you can request access, whether you want to contribute or just browse, and we'll be able to add you to the members. And with that, thank you so much; that's all we have today.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Thank you very much. We've heard some really rich testimony today and we have limited time for questions. Michelle, do we...can we go over at all today?

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

We can't go over but we could have a couple of questions now and, you know, go to public comment.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Okay. So I had one question for Dr. Davidsen, what role do you see your cloud platform operating in care with regard to interacting with EMR or providers or participating in patient-generated health data and phenotypic data that might come back to your cloud?

Tanja Davidsen, PhD – Biomedical Informatics Project Manager – National Cancer Institute

So we do hope that in the long run that we do have patient-generated data coming into our cloud pilot, so that's kind of a long term goal. Initially, like I said, we're just going to be starting out with some of the NCI-funded projects, but I think ultimately the goal will be to have the patient data there as well.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Super, thank you. Do we have other questions?

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

I have a quick question or comment, Andrey Ostrovsky here from Care at Hand. This is all really exciting, thank you guys for sharing this. One caution I want to put out to our group as we're synthesizing observations or recommendations is that there has been, outside of the Precision Medicine Initiative, a lot of hype around open data and in fact, that data is not very open at all. And I have found that in order to be able to access any meaningful data that there are incredible loopholes that innovators, researchers, entrepreneurs, etcetera have to go through in order to actually access that data and do anything creative or meaningful with it.

So I think as we're moving forward with the precision medicine work, particularly around IT and interoperability considerations, I think we need to place a huge emphasis on the design considerations

for user experience so that all of the work we're trying to do in aggregating this data is actually accessible and folks who want to contribute in a meaningful way, including patients, can contribute and do so without having to hack the system, so to speak. That goes especially for patient and/or innovators who may not...may be predisposed to disparities in access to resources in the first place. So I just wanted to make sure we have that word of caution, but this is super-exciting; thank you guys for sharing.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Thank you. I think that that's an important design consideration tasked to both of our presenters, how will we make this data more usable, available to not just the research community, but the patient community at large. Are there other questions?

David McCallie, Jr., MD – Senior Vice President, Medical Informatics – Cerner Corporation

Yeah, Leslie...

Steven Keating – Patient Advocate/Consumer – Doctoral Candidate, Mechanical Engineering, MIT Media Labs

Yeah, Steve Keating here; I had a quick question for the precisionFDA fellow. I was curious when it's going to be leaving beta mode and when average people can join without having to send a special request e-mail as a contributor?

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

Yeah, so, you know, I mean the beta to access really meaning that we're continually enhancing the platform and adding new modules. The subscription is really just as simple as just letting us know who you are so you can sign up; immediately you'll get a browsing access the minute you join the platform, so that way you can see ever...what's being shared in the public area of precisionFDA without needing any additional steps.

The only thing that we try to provision beyond that if someone is trying to now actively add Apps and whatnot; we have to validate that, so that way it's not like some, you know, any group will be able just to join this. You know, there are some bad actors out there and we want to be able to protect the privacy of the ecosystem. However, anyone who has asked for access has received the browsing access immediately.

As far as getting out of the beta, I mean, we really want to try and see where the community is right now. We're constantly getting new response every single day; we're trying to prioritize those. We want to make sure that when we exit the beta is we are...at that point and we want to just extend that as much as possible and Elaine and David can talk a little bit more.

Steven Keating – Patient Advocate/Consumer – Doctoral Candidate, Mechanical Engineering, MIT Media Labs

Okay, thanks.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Thank you. I think I heard David McCallie?

David McCallie, Jr., MD – Senior Vice President, Medical Informatics – Cerner Corporation

Yeah, for Dr. Kass-Hout, does the precision medicine include hardware testing or is it all just software evaluation? In other words, do you look at actual performance of the sequencers or just the pipeline?

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

Right now it's the pipeline put it in like a benchmarking. So if you look at this as three separate steps, there are the instruments, the software and then the clinical interpretation. Right now we're focusing on the analytic standards and the software benchmark.

David McCallie, Jr., MD – Senior Vice President, Medical Informatics – Cerner Corporation

Okay, thanks. Is there a plan to expand that or is that not...

Taha A. Kass-Hout, MD, MS – Chief Health Informatics Officer and Director of the Office of Health Informatics – Food and Drug Administration

We are going to...absolutely, I mean, we're looking at various way...I mean, there are a lot of regulations around the instruments already out there and people follow those. Right now it's, we're trying to kind of focus on the next sort of challenging thing with this environment. Also respecting the privacy and intellectual property of a lot of these users because a lot of folks use their own environment and also a lot of...this is why we offered the private areas to start with, because we want to be able to help people sort of achieve that next step.

David McCallie, Jr., MD – Senior Vice President, Medical Informatics – Cerner Corporation

Thank you.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Thank you, I think that's it for questions. We will ask though people to continue to ask questions via e-mail. We will also publish those answers to those questions. We've learned today that we have a lot more information that we can get in an hour and appreciate the flexibility of Lucia Savage, who will present in the next meeting and I'll do a better job next time. So with that, I think Michelle we need to turn it to public comment.

Public Comment

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

Thanks, Leslie. Operator, can you please open the lines?

Lonnie Moore – Virtual Meetings Specialist – 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 telephone 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

It looks like we have no public comment. So thank you all for your patience with us today and thank you Lucia for your flexibility and we'll have you present during the next meeting; and our next meeting is Friday, February 26 at 1:00 Eastern. So thank you all and have a wonderful weekend.

Leslie Kelly Hall – Senior Vice President of Policy – Healthwise

Thank you Michelle. Look forward to e-mails, folks. Thank you, bye.