PUBLIC HEALTH INSTITUTE
Dialogue4Health Web Forum:
MAKING DATA WORK FOR THE PUBLIC'S HEALTH: DIVING INTO THE DETAILS
THURSDAY, AUGUST 13, 2015
2:30-4:00 P.M. ET

Remote CART Captioning

Communication Access Realtime Translation (CART) captioning is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This transcript is being provided in rough-draft format.

www.hometeamcaptions.com

>> Joanna Hathaway: Hello and welcome to "Making Data Work for the Public's Health: Diving into the Details." My name is Joanna Hathaway and I will be running today's Web Forum along with my colleague, Holly Calhoun.

Closed captioning will be available throughout today's Web Forum. Beth with Home Team Captions will be providing real-time captioning. The closed captioning text will be available in the Media Viewer panel. The Media Viewer panel can be accessed by clicking on an icon that looks like a small circle with film strip running through it. On a PC this is top right corner of your screen and on a Mac it should be located on the bottom right-hand corner of your screen.

On the Media Viewer, you'll see the show/hide header text. Click on this to see more of the closed caption. During the webcast, it may collapse. Don't worry, you can open the window by clicking on the icon that looks like a small circle with film strip running through it.

If you experience technical difficulties during the WebEx session, dial 1-866-229-3239 for assistance.

Write that number down for future reference. The audio portion can be heard through your computer speakers or a headset plugged into your computer.

If at any time you're having technical difficulties regarding audio, please send a question in the Q&A panel and Holly and I will provide teleconference information to you. Once the Web Forum ends today, a survey evaluation will open in a new window. Please take a moment to complete the evaluation, as we need your feedback to improve our Web Forums.

The recording and presentation slides will be posted on our website at Dialogue4Health.org after the end of this event.

We're encouraging you to ask questions throughout today's presentation. To do so, simply click the question mark icon, type your question in and hit "send."

Please send your question to all panelists. We will be asking questions both throughout and at the end of the presentation.

We'll be using the polling feature to get your feedback during the event. The first poll is on screen now.

Please select your answer from the available choices and click the submit button. Are you attending this Web Forum: A, individually
B, in a group of 2-5 people
C, a group of 6-10 people or
D, in a group of more than 10 people.
Click submit when done.

Once done answering the poll, click on the Media Viewer icon to bring back the closed capping if you need it. It's my pleasure to introduce Shell Culp, who has been a complete joy to work with since we started putting together this series many months ago. When Shell isn't talking to us, you can find her talking about it somewhere else. She speaks on performance improvement for government, organizational change management and open data movement across the country. We are so grateful here at D4H to have her expertise and humor on board. Shell, please go ahead.

>> Shell Culp: Welcome and thank you, Joanna for that lovely introduction.

So welcome to our Dialogue4Health web for series "Making Data Work for the Public's Health: Diving into the Details." Dialogue4Health is a community that feeds, builds and shares strategies to improve public health. Hosting Web Forums and share critical resources. This series on "Making Data Work for the Public's Health: Diving into the Details" is sponsored by the California Healthcare Foundation. PHI is dedicated to improving the quality of life for people around the world. Learn more about PHI at www.PHI.org. This Web Forum takes a look at increasing focus on using data to improve health and wellness. The ways in which different sectors are able to select and use new data sources ranges tremendously doing to differences in specific innovations that advance particular initiatives.

The California Healthcare Foundation Dialogue4Health is excited to bring you the series, and in this three-part series we are exploring the current landscape of data and public health, which was our last forum. This forum will choose investigating, exciting and innovating work currently implemented and the following forum which will be imagining and planning for the ways in which improved use of data in the public's health.

The second Web Forum, Stewards of Change, I'm pleased to introduce our panel. We have with us Andy Krackov, associate director for external engagement at the California Healthcare Foundation.
We have Dr. Este Geraghty. We have a recording of Dr. Este Geraghty, chief medical office and health solutions officer. We have Jay Bhatt, Chief Health Officer for Illinois Hospital Association, and tin sent Seaman, the Senior Program Officer at Bill and Melinda Gates.

It's my pleasure to introduce Andy Krackov from the California Healthcare Foundation to help us frame the Web Forum.

At the California health and human services, free data, a catchy name. It took me a year to comprehend the importance of what Andy understood as the definitive route to unlocking data that the public owned to be used by the public, a simple concept. Andy is going to help us frame the second in our series on investigating exciting and innovative work currently being implemented.

Andy.

>> Andy Krackov: Thank you so much, Shell. I appreciate that. And thanks to PHI for running this webinar series and thanks to everyone who is attending.

I thought I would just spend a few minutes to provide, you know, a quick back drop. First of all, it occurs to me and probably all of us, that data as we know, used for multiple purposes in a public health setting, scientific discovery for evaluation purposes, public health surveillance, and of course at the point of care.

But when you think about the overarching title for our series here, "Making Data Work for the Public's Health: Diving into the Details," I think we also need to
recognize the important role that data can play as a tool in communication for achieving some sort of social impact. That could be raising public awareness of issues. It could be changing behaviors. It could be informing policy decisions.

And I want to talk for just a few minutes, and by way of intro, about I think this particular aspect of communicating data to persuade or to achieve some kind of impact. And I think that sort of speaks to why CHHSA was really interested in funding this series.

See, I'm not positive that we've made tremendous headway in terms of our collective capacity to translate and communicate data so that we can really use it to see some social impact. That's not to say that things aren't being done. In fact, one of the reasons for the series is we can lift up the examples of what is being done to raise awareness, to change behaviors, et cetera, through data.

But I think the communication piece regarding data is it's really challenging, and that's partly because there's a lot of steps that need to lead up to that in order to sort of do this work effectively. First you've got to find the data sources and for non-profits or even sometimes public health departments, that often isn't as easy as one would think. And you've got to harvest the data if you want to regularly use it. You have to curate. Finding a source is no in and of itself the end of the battle. Then you have to analyze the data for meaning, which takes time. And all that has to happen before we can actually effectively communicate the data but what I'm noticing, we're getting more efficient with these beginning stages, and that means instead of the communication piece being sort of the 5% of the end of the budget or time that we can devote to this, we now can devote significantly more time and resources to this communication aspect.

And to that I wanted to share some related work that we've been working on with counties, with a number of counties in California through a project that we've been funding.

We heard from them that communication is truly a challenge that they're also facing in their work, the county health departments. Some have the capacity to find and harvest and analyze data, and some don't, quite honestly. Some of the smaller counties.

But nearly all of them have told us that they really struggle with the communication and the translation piece. Raise awareness of issues, change policies, et cetera.

And I think as we know, the typical way of summarizing data findings for public consumption has been through the traditional PDF data reports, 30 or 40 or maybe shorter, but that data report that is not so much a living document. It's perhaps one that gets put on a shelf.

So we asked -- and this is a long engagement, we asked ourselves, we asked the counties, whether there's anything we can do to move beyond that, to modernize the whole experience of reading and being persuaded by data, to make it more interactive and digital and make it more efficient for those having to create these kinds of reports.

So what we came up with -- I'll show you a few examples. Is this concept of creating engaging data briefs, stories, you can think of it, not so much PDFs, and this is the first one -- one of the ones created. This was created by PHI, one of our grantees, about C-sections in California, and the whole idea is you're bringing in multiple elements. You're bringing in live graphs as you can see here. They're clickable and you can do many things with them. You're building in infographics, contextual information. Even on this particular story we even have videos that are embedded in here, and as the section pulling up right now, you can
sort of see you can pull down -- use the pull-down to get to data for your particular county.

So it functions much differently than a PDF. If we can go to the next one, prescription pain killers, this one was one that Marin County thought they wanted to work on. We know the prescription painkiller abuse is becoming a growing issue. And this is a report that they, on their own, in essence, using the technology we provided to counties from live stories, our technical partner in this, using that technology they put together this story. And it feels hopefully much different than a PDF would. And if we go quickly to the last one, the example, this is a story that was produced by our partners in Lake County, California, a rural county just above Napa, and they wanted to reach the Native American communities there and try to encourage them to, you know, sort of really get themselves checked and make sure they are preventing diabetes effectively as a community.

And so they used this or are using this as one of the elements. Now, I don't suggest this alone can do the job but that this is one of the tools that we could have available to us in order to make a case. And as you can see there's a video also embedded in here along with other kinds of visuals.

So the project is really still very much in progress. We don't yet have results to share. These, in fact, are some of the early stories. And I think it's one example of perhaps we all need to think creatively and really a new innovative ways to see what may stick in terms of getting people to pay attention to data, and I think that's really why I'm excited about our foundation, CHHSA has an opportunity to sponsor this webinar series, because I think it really speaks to Dialogue4Health's -- the capacity for this to be a platform where we can share other ideas for how we inform our world in new ways and how health data can really shape our opinions and actions in ways that we need to do more of.

So I will leave it at that. Thank you, Shell.

>> Shell Culp: Thanks, Andy. I know that, you know, before we move on, I would like to maybe engage you with a little bit of dialogue. I know that California Healthcare Foundation also does a lot with citizen engagement and using the challenge model and sort of proselytizing with this notion of building community and co-creation, and I wondered if you could maybe share your perspective with us a little bit on that. Specific technology models typically -- they give us a lot of really good technology, but sometimes it's very focused on building an app or building something that has maybe limited use when we're thinking about the range of kinds of problems facing public health and these would be really big things like poverty, violence, hunger, aging, housing and transportation, and I was wondering if you could maybe comment for the audience about whether that model scales to the civic engagement model is what I'm referring to there, whether that model scales to these larger issues and what you see the role of civic technology in the future in dealing with these big issues.

>> Andy Krackov: I think that's a really good question and we have, indeed, had the pleasure to work with the California Health and Human Services Agency to help move forward a lot of their open data work, which in essence means them working very closely with civic technologists. I think what I learned from this, yes, apps can be created, and that is very important. But I think what I learned is that it gives us, including government, in this case California's entire health agency, a new methodology that we can integrate into our thinking, and I think that methodology ties into things like human-centered design that often is the name given to it, and in essence think of it as walking in your customer's shoes. And so if
we -- I think we have little hope of solving the kinds of issues that you talked about unless we truly understand how people are interacting with this, what works and what doesn't. I think civic technologists, what they bring to the table, in addition to the apps, is an appreciation we need to understand need and dig deep before we go ahead and build out products that are serving them.

And the best example I can give is the fact that the California Health and Human Services Agency, they rolled forward first with this open data program, and we had been providing them a lot of grant funding to help in essence build a bridge to communities in California that will use those data by building apps, et cetera, but what I didn't expect is that that bridge would be a two-way street and we found a lot of people in turn were giving a lot of good innovative feedback, civic technologists were back to government, and I think partly as a result of that, the agency, California's health agency, has decided they want to launch an innovation office, that they want to bring a lot of this in-house, and that's when -- when you see things like that, that's when you can really start to get some scale.

>> Shell Culp: Really, really good points. Let me just also add for the audience that this is the first birthday of the California Health and Human Services Agency open data portal this month and it wouldn't have been possible without the generosity of the California Healthcare Foundation, so happy birthday and thank you, Andy.

Let's take a minute now to look at our second poll. If you wouldn't mind answering these questions for us. How is your organization currently using public health data sources? A, what is a public health data source? B, my organization does not currently use public health data sources C, my organization uses public health data sources to decide where to concentrate our efforts.

D, my organization uses public health data sources to help with communications and development efforts.

E, my organization uses public health data sources to evaluate our progress.

And F, other.

And please let us know what "other" means to you.

Next we're going to hear from Dr. Este Geraghty and I mentioned she won't be with us, but we do have a fantastic video she put together for us. Dr. Geraghty is a bit of a dynamo and in addition to her travel schedule to evangelize on Esri's tools, she teaches medicine and University California at Davis. I met her when we were embarking on building.

So pushing the limits of technology with this video and we're going to try it anyway.

Some folks may have trouble viewing this video, but depending on your Internet speed, it may or may not work for you. If you do have problems, please know that the video is posted on vimeo and can be watched at any time and the link should be showing on your screen shortly.

So let's get started with that.

>> Joanna Hathaway: This is Joanna and looks like we're having trouble with that. We're going to direct you to the video on vimeo, and Shell, I'm going to recommend we go right ahead with the program. Can you hear me?

>> Shell Culp: Sounds great. Yes, of course. That sounds great. Apologies for that, but we were skeptical that we were going to be able to make it work for more than 300 viewers, which is what we have on the line with us today, and so apologies for that.
So let's see, we have one more poll before we go to our next presenter, and let's see if we can get that up.

There we are. Poll number 3.

So thank you, by the way, for your answers to polls number 1 and 2, so this is our third poll. What data sources does your organization currently use?

A, Community Commons
B, health and human services data portal for the state of California.
C, HealthData.gov.
D, GIS.
E, I don't think we use any public health data sources.
F, I know we use public health data sources but I don't know which ones.
G, other, and please tell us what other is.

So moving along, next up is Dr. Jay Bhatt, and I first heard Jay speak at an open data event in Chicago a couple years ago. He probably doesn't remember I was there, but he spoke from the audience and appeared also appeared to speak from his heart, about the importance of data and open data in particular.

Engaging people to live more healthy lifestyles. We're thrilled to have him with us to share perspectives on health from an early proponent. Dr. Bhatt.

>> Jay Bhatt: It's crazy how time flies, but thank you for all your work in this movement. Thank you tore the Public Health Institute and Joanna and her colleagues for coordination of this, and certainly I was grateful for Andy Krackov and the California Healthcare Foundation for the work they're doing to contribute to social impact with data.

So today I would like to share a perspective that relates to my time at the Chicago Department of Public Health, as well as my short time with the Illinois Hospital Association, where I think there's extraordinary opportunity to use data to improve the lives of our residents.

And I would like to share context around this in population and then offer some examples and questions along the way.

And so when we talk about leveraging data for public health, there's all sorts of different kinds of data available to us. There's big data and unstructured data, there's crowd source data. It's no secret we're living in this time of a lot of data. But the question is, do we have the right data and how will we use it? That's part of what we are embarking on in this world, with taking data from different sources and bringing it together to have actionable insight to make decisions or research allocation.

When we think about population, Kindig and Stoddert proposed a definition as health outcome of a groups of individuals including the distribution of such outcomes within the group.

They emphasize that in the past the concept of population health became confused with its determinant and they urge a clear distinction of one from the other with the inclusion of both population health models. Now, Kindig and Stoddert didn't do it 35 years ago, they talked about it in 2003, but can trace the origin of this still continuing to be vague concept, depending on who you talk to as well.

And so there is a distinction between population health, which is really entrenched and anchored in the community, versus population health management, which is anchored around an attributable population trying to achieve the AAA costs, better patient experience and better population health.

So with that, I'll go to the next slide that looks at the issue of equity and from the institute of healthcare improvement. Prevention and health promotion
and medical care, and they're increasingly are visible number of efforts of integrating this public health and medicine construct or public health and healthcare delivery to improve the lives of people and communities.

And the prevention and health promotion as you might expect includes economic factors, physical and environment, which are upstream factors and some may be familiar with Rishi Machunda and the work he’s done in the health upstream movement.

And then individual factors which include genetics, spirituality and behavioral factors, a lot of resilience. And then intermediate outcomes where we transpose into the medical care arena, health and function, mortality, and then ultimately we really are looking at the best at trying to deliver the best quality of life for people that is possible.

And the interesting thing is really if we look at the data around prevention and health promotion, we don't really have access to all of that data. Now, when I was at the Chicago Department of Public Health I would get this question, what is the prevalence of diabetes? What is the prevalence of blood pressure? What is the -- high blood pressure. What's the prevalence of a number of other conditions? And I would say that I couldn't really tell you because we don't know what the denominator is. And that's where a chronic disease surveillance system can play a big role. Integrating data from the different points of contact but really give you a true sense of prevalence at the neighborhood level. And that's a place that some folks are moving towards and that we all need to get to because it gives you insight into the kind of interventions you can design and deploy.

We know social factors contribute up to 60% of premature death and premature mortality. And so if that's the case, then how do we understand what those factors are and how people are integrating with those factors, and where does the data from that come from?

So the question posed is: Do patient electronic health records provide the information to know how to reduce unexpected readmissions, for example?

So right now the notion is we can explain 60% of readmissions that happen, but then 40% of re-admissions that we can't explain. Could that be related to the conditions people go back to, the conditions that potentially make them sick?

And the IOM has started to -- and other agencies have started to touch on this around metrics, about what are the metrics incorporated into EHRs and how do we access that data? And then how do we access data from the digital footprint that people all across the country are reading?

As part of the work we're doing here at the Hospital Association determinates of health for Illinois, and that includes the data we have access along with public data set and designing insight that can help move the needle on health conditions that continue to play part in Illinois and be models for other states around the country.

So if we look at Illinois, for example, and this can be done certainly across other states, we see that the 30-day re-admission for the state compared to Medicare fee for service and Medicaid fee for service, largely a mix of cardiovascular and behavioral health, as well as now seeing conic obstructive pulmonary disease. If we look at those and go back to the previous conversation around social determinates, we understand there may be social, economic factors that influence these and how do we get that data in front of providers at the point of care to he them ask better questions to deliver smarter evidence-based medicine, but also understand the conditions that people may go back to.

And so part of that work involved a population health stratification. And
so we’re looking at taking our data and segmenting based on risk, which is not a new construct, but then taking that data and segmenting based on social, economic and environmental data as well as community asset data can provide a better picture. And then on top of that the digital footprints that are being left, whether it’s through social media or through wearables or otherwise.

And so the interesting thing about super or high utilizers, you may be familiar with Dr. Brenner, but also others around the country working on this, that the interesting thing between a diabetic visit of a patient I see, there are 2,000 hours they are not in the office. 120,000 minutes where they’re not in the office between that time. And the question is, well, what are they doing? And is what they’re doing affecting their -- and how is it affecting their health? That's part of what I think some of the open data and some of the other nontraditional data helps us get at.

And so this is an example some of you may have seen out of UCLA published in Health Affairs not too long ago. There are a few studies that linked household income to major preventable complications of chronic disease.

So in California, really the statewide facility discharge data identifying the lower extremity amputation, almost 8,000 of them, and about 6800 adults with diabetes, and base on residential zip code and connect that data to the Census Bureau that correspond to maps of poverty rate. And you can see the quartiles, and based on color, and where the amputations sort of fall out at.

So comparisons in the maps show amputation hot spots in low income, rural regions of California. And, again, the important thing, though, is correlation doesn't equal causation, and so that is important to keep in mind, but the method per mapping the hot spots using public data sources may help target interventions.

And so, again, that can be applied across other condition, but you can also look at it as a continuum from the time someone had pre-diabetes to the time that they potentially had to have a lower extremity amputation. Well, what triggered that continued continuum? There were possible opportunities to move that in a different direction. And so that, I think, more and more is something we can do with the data we have. And was there two or three? These are all important considerations as well.

And some of the other work that we’re seeing in sort of this integration between health system and health department, and some of you are continuing to see sort of this happen in other parts of the country, but there is a population health continuum from hospital to healthcare system to health, episodic encounter to healthcare needs across the lifespan and quality of life in community.

This is an illustration from the Department of Public Health where we showed how the Chicago trial, for example, is taking a partnership between the inspectors trained to do home inspection as well as emergency departments across Chicago and triggering community health worker, lead inspector visit to the home with two or more pediatric visits to the emergency room, and that's making a difference. And working with the University of Chicago Center for Diabetes Translation Research, not just having diabetes but creating infrastructure, understanding what are the underlying reasons contributing to diabetes and where are they and where are the opportunities to be able to move that?

And then the other is in partnership with Northwestern and CHITREC taking EHR data and they’re creating an opportunity to now say we can estimate prevalence of diabetes at the block level even though we don't have that data based on validating EHR data from six or seven systems with survey data or other data that might be available.
And then Sinai Health System the using CHW to decrease complications from diabetes, but how do we take the data generated from the conversations between community health workers and others to then feed that back as well as into the data infrastructure.

And this is another example really when I talk about at the point of care, this is putting it visualization interface on a provider's panel, so that they can clearly see trends with their panels and are able to see where particular conditions map to and if there are other social, economic and environmental connections that can be made. So that is really also an opportunity there. So, for example, if I were to have a patient that I would want to ask questions about based on a condition of allergies or headache and I had air quality data in front of me or housing data in front of me, would I ask different questions and would that eventually keep us from ordering inappropriate imaging or other intervention.

So now I would like really to talk about predictive analytics, and predictive analytics is not new, but it is relatively new for healthcare in the way that healthcare is actually using it, and particularly using community clinical data, unstructured data, to try to resource allocate and improve outcomes. This is the arc of predictive analytics and it comes from the raw data all the way through the different points of merging, cleaning data, and then the questions of what happened, why did it happened and predictive modeling answers what will happen, and then optimization, which is what is the best that could happen.

And so at Chicago we've used predictive analytics in a number of ways. We partnered with the center for data and policy at the University of Chicago to develop a model to improve our lead inspection program and exposure to lead can seriously affect a child's health, causing brain and neurological injury as well as hearing and speech difficulties and effects can be seen in educational attainment, learning and behavior problems may cause low IQ attention deficit and other performance.

And so really, we use data to apply for predictive analytics to identify which ones most likely have the greatest risk of causing lead poisoning in children, based on home inspection records, assessor data, environmental data. This slide really shows the return on that data investment is that without a model, clearly many buildings, a lot of dollars, much time.

When you go to the model forecast it really pinpoints and identifies those most likely to be most at risk for pregnant women and children.

And so this is one way to also use the data and integrate with public health. But then you can take this model and deploy it in the electronic health record. And once you do that, that can be used as a clinical support tool at the point of care but also a way to bring people loss to follow-up back into care.

Clearly social media has also taken on its own life. There are many different opportunities to integrate social media data into questions, but also when we think about who is really the public health workforce of the future I would say they're residents of our city, in our community, that can be that. And we see that through our work in Chicago at FoodBorne Chicago which allows residents to report potential food borne illness by Twitter. We have a message we go back and forth with the resident and get that information and investigate the potential restaurant and putting the results of that inspection on the open data portal with a tracker number that is part of it.

Can angry tweets predict heart disease is something that is out recently as well, and this looks at really analyzing tweets from 1300 counties. Researchers found that tweeting would be a better predictor about smoking, stress, and using negative language like "hate" and expressing fatigue was correlating.
Natural language processing, techniques are continuing to play a role. You've seen recent conversations between various retailers and other members that have some of these analytics tools. We're seeing here particularly in Chicago a big insurer share data, claims data with independent practice groups to improve outcomes as well.

And then this -- these are tweets around asthma. Big data, social media and claims data coming together. So collecting and mapping the location of tweets with keywords such as asthma, inhaler and wheezing, mining the hospital in Dallas and comparing air quality reports, they were able to report with 75% accuracy.

It's a tool then, a resource allocation yet to be seen how this really takes hold in other disease conditions and other areas.

Then the other opportunity really before us is taking our community health needs assessment data, plotting it to look at overlapping service areas and then bringing together the community asset data and then also the evidence-based data. We have a lot of different projects happening around particular preventable healthcare conditions or otherwise.

How can we take the information from that and include it to understand which ones are working and which ones are not.

But also to redesign community benefit investment so it's been deployed in evidence-based intervention, driven by data. And driven by what the community needs.

So with that, I would say a few more things. The kind of questions we think about, what happens when a hospital or school or public health department, or community organizations have to do? It depends how they use it.

They have their own data in their silos, how can we bring that data together and what kind of questions would we want to ask and what kind of problems would we want to solve? Data should be accessible to those who need it most.

There's no reason why we should be holding particular data and not sharing it, and an interesting thought is can we de-identify an aggregate claim in EHR data and put it in open data portal so that the community can use it. Can we do that as in government, in other areas as well. There has to be governance in data, so think really carefully about that. So you'll see examples of retailer data being integrated with claims data and otherwise.

And credit card data being integrated to understand chronic disease and predict congestive heart failure, for example.

But in that notion, there has to be, I think, an important understanding of governance and privacy but sharing and analyzing data and increasing the depth of our understanding of public health challenges as well.

So by looking at technology, for example, health in Louisville has mapped inhaler use as a result of putting technology in the top of the inhalers.

The final thing, we know that place matters and that we need to understand what is in that place and bringing and aligning stakeholders together around data as an anchor to improve the lives of our neighborhoods and communities and our nation, I think it's an important thing to really invest time, energy and effort.

And I think the last thing I'll say is there's a lot to learn from other industries and a lot to incorporate as it relates to using data to leverage and improve public health. So with that I'll end and appreciate any comments or questions.

>> Shell Culp: Thanks, Jay. Very interesting nexus that you draw at the end here with a link to other industries and what can be learned.

Just from my uneducated vantage point, it looks like the Illinois Hospital Association work is a perfect fit for you in your perspective and the kind of
interest that you have, although I'm not real sure exactly how things like governance is, but I'm sure they'll figure that out with your help.

A couple of things that your presentation brought up for me -- and let me just take a moment to remind everybody that we're going to take questions at the end of the Web Forum and it looks like we'll have plenty of time, so go ahead and use the Q&A feature over on the right side of your screen here to log your questions.

So we haven't been with the Illinois Hospital Association very long, you're just getting started there, but what do you see as the data focus for the next couple years for the IHA?

>> Jay Bhatt: So I think the focus is, one, using data to address the critical health issues that our state faces and that may be as a result of readmissions. It may be connected to more acute conditions, like sepsis, but also to things that are chronic, like chronic disorders. And finally I would say behaviors in health. And I think our goal is really to take the data we have, which is informatics comp unit and claims database and generate insight from it that helps make better health decisions, allocate research effectively, and connects with non-health data as well as the open data that is out there, so that we can uncover opportunities for improvement, but also use it to design interventions.

>> Shell Culp: Okay. That sounds really very interesting and I look forward to next couple of years seeing what you are able to put together.

And it kind of works into another question that I was formulating around, you know, hospitals and healthcare providers have been part of the system for a long time, you know, hundreds of years even, and the data that has been used during that time has pretty much defined or helped to define those systems of care. But it's not necessarily connected to public health, and I'm hearing you talk a lot about public health.

So what data would hospitals like to share with local and community public health organizations, and do you see any barriers there to that sharing? And if you do, what could be done to overcome those barriers?

>> Jay Bhatt: Shell, I think that's an important question. I think that the delivery and payment environment and the evolving healthcare environment is really I think pushing the healthcare delivery system to think differently about data. And how they use it and how they share it, who are their partners in that work, because they're being asked to take accountability and risk in different ways. And if you don't know what is happening, as alluded to before, the patients outside the four walls of a hospital or your clinical outpatient office, then that's going to have implications for your ability to deliver healthcare and take on that risk. And so I think they are willing to share data that largely is connected to, I think, complex patient and chronic disease, but also this high utilizer group. If we understand the folks using the emergency room most and who are being re-admitted most, can we understand more about them in partnership with other stakeholders, and the other piece is that -- is around, I would say, competency. Some systems from further along in their analytics competencies thinking, but when you look at critical access for some rural and smaller community hospitals, they may not necessarily have the resources or capacity. And so in that case they do have to partner with other folks to do some that work, and in that piece they can share data.

I think you're going to see an increasing amount of retailers sharing data with hospitals, like CVS and Walgreen's. You'll see technology companies and also lab companies sharing data differently with the public.

I think you'll also see, you know, this experimentation about how digital data gets incorporated into the clinical workflow of the healthcare delivery system.
How can that data be used to make different decisions about healthcare for population or to ask different questions or design different systems.

>> Shell Culp: Great. Thank you very much. That was interesting additions and I hadn't even thought about lab companies, so thank you for pointing that out. I don't know, there might be even more.

So thank you, Dr. Bhatt for being a great addition to our Web Forum.

Now we would like to invite you to -- thanks -- now we would like to invite everyone feedback one more time. Our fourth poll is: What do you see as the most important next step in moving public health forward with respect to using data? A, improved training for data resources that already exist

B, more access to specialized data resources for particular causes

C, better integration of data systems already available

D, a wider understanding of the possibilities to spark innovation in the general field

E, a wider understanding of the possibilities to bring more advanced use of data in the workplace

Or F, other. And please tell us what "other" means to you.

While those poll results are being tabulated, it is my distinct pleasure to welcome Dr. Vincent Seaman, who captured the audience at the second open data Fest this past March when he shared experience with polio vaccination in Nigeria and challenges he encountered in that effort. Practicing pharmacist, he's also skilled in geospatial technology. We're happy to welcome Dr. Seaman to our webinar.

>> Vincent Seaman: Thank you. It's been a wonderful, wonderful, wonderful web conference so far and I think that maybe will take care of my heart disease for the rest of the day, hopefully. So I'm going to -- I'd love to comment on a lot of issues that have been mentioned already, but I think I better stick to my slides, otherwise I won't get through them in the time allotted. What I would like to talk about is how we use data and geospatial data in a country that's not accustomed to the resources we have here, and some of the things to be successful with polio there.

I spent three years in Nigeria from 2011 to 2014 and I was working mainly on polio and other vaccine preventable diseases, and there are certain things we need to do to be successful with polio. The counties endemic, Nigeria, Pakistan and Afghanistan do polio campaigns where there's a nationwide campaign to vaccinate every child under five years old, and if you did these right, you could do three of them in the space of probably three to six months and you would be done. But because these countries aren't able to really effectively do these campaigns and reach every child, there's always reservoirs for the virus to live and we can't finish the job. So we've been doing these campaigns about eight or nine of them a year, for the past ten years, and we finally were able to go to Nigeria without a case. But the things on the slide I listed are typical activities that go on, planning, tracking of teams, which we do there using GPS devices and then there's monitoring of coverage so we know how well we did, surveillance for polio cases in between. And then miscellaneous applications that we've found.

And all of these things require an accurate basemap that just has you know, boundaries of all the administrative areas and locations of settlements, and when I got there, these are the kind of maps that I saw people using. This is probably one of the less detailed ones, but it's not unique by any means. And it doesn't really give you much information about where the people are supposed to go, how many people they should expect to vaccinate, et cetera.

These are some better examples, but these are really pretty much the average and there's not -- pretty much the average, and there's not a lot of detail,
and the detail you have is not spatially correct. If they list a settlement in a particular place, it doesn't necessarily mean it's there. They kind of just write down the places, you know, and fit them inside a shape of some sort and the shape typically isn't all that correct either. So this is what we had to work with and we realized we needed something better.

So we used satellite imagery, which nowadays is very advanced. You can get high resolution imagery and see individual houses and smaller detail, and we used that to identify all the structural features in the northern part of the country where we had virus, and we put points on small settlements and what we call hamlets, the small clusters of ten or less houses, and then we actually drew polygons around the larger urban areas. So we created a layer of all these features and then we sent teams out to go to all of those places and find out what the name was and collect points of interest and then we put them on a map. And so the map that you see at the bottom is one of our finished maps for award, which is kind of the lowest administrative level in Nigeria. And the maps across the top are that same ward if you tried to search on any of the common reference map sites for geospatial information you can see there's really nothing available, maybe a main road and maybe one or two settlement names, and in actuality there's a lot of stuff there. If you're going to do anything on the ground where you need to know where the people are, you can't -- you know, you can't make it with the maps on the top. So we really visited over 100,000 settlements to collect the data, and we have a rich map layer. We collected points of interest where vaccinations sometimes take place or we put a fixed post where we can cover lots of kids at once. These would be like marketplaces, mosques, schools, and water points.

And if you are thinking about things, you can probably imagine a lot of other uses for this kind of data besides polio vaccination, and I think that's the whole idea of the talk today, is branching out beyond specific use cases.

This is a picture of a map that we used for vaccinator tracking. Now we know where the settlements are and the names, and the settlements on this map, I didn't put the names because it cluttered things, but there are dashed outline circles that are small hamlet areas or settlements, and then the dots are the movements of the vaccinator carrying the phones. It's date and time stamp so we know when a person was at a particular place at a particular time, and you can see in this ward, really every settlement was visited. The green dots are what we call mop-up, which means if we see they missed something during the campaign, which is four days, on the fifth day they go back and visit places that, you know, we noticed they were missed.

Using this kind of technology we were able to ensure that every single settlement was visited and that every child hopefully was vaccinated. We don't know really what happened on the ground, but we work on that, too, in other ways.

So this is just a summary of the different innovations that we were able to do in Nigeria with this kind of information besides planning and tracking teams, we found some settlements were never visited, even though we kept directing people to go to them. And so we actually put all this information on a tablet and showed the local people how to use a tablet to go find a particular place.

So it's like a compass and map on the tablet and points you in the right direction and how far you have to go. We used digital elevation models for surveillance sites where we sample sewage or wastewater to see if there's virus in there. This is all gravity flow in Nigeria, so if you know that you're downhill from a place, then theoretically a specimen, you know, will come to you from that area but not others.
We did a lot of things I won't go into, but I do want to talk about the population, because one of the issues we had in Nigeria -- and this is also in a lot of other countries -- is that the census data is either old, out of date or inaccurate or all of those. Nigeria, the last census was 2006, and people don't feel that it was very good or the quality was very good at the time. And if you can imagine, you know, we're introducing a lot of vaccines and a lot of medicines in Nigeria, some that cost over a dollar a dose. So if you're off by a few million people, that's a few million dollars that, you know, we are wasting or maybe coming up short.

So we thought we really wanted to get better idea of population, and the denominators are important for many reasons, and we already had this rich satellite imagery data for the north, and there are tools now that you can actually train a computer to identify certain objects and then select those out preferentially. So our plan was train the computer to identify residential structures and we would do counts on the ground to tell it how many people on the average live in each kind of structure and then we can use this kind of a model to estimate population independent of the census data.

And here is just a shot of the African continent showing the latest date for a census or other information that has reliable population estimates. And you can see most of them are a few years, five to ten years old, and so this is a huge problem. And if you look over in west Africa where we had the Ebola outbreak you'll see some of them are more than ten years old. So the responders there had little idea about how many people they were dealing with in that response.

There are some resources on the web for population. Most of them are -- use population density modeling and they basically take the census data that is available and distribute it using some different tools. Imagery is one, but they then assign people to places, but it's the same census data that we feel is unreliable to start with, so this is not really a solution for everything. It's good for modeling and spatial, looking at things in the big picture, but it doesn't help us.

So we are doing this what we call a bottom-up model, where we find the locations. We determine how many people live in each location type and then put a number in the cell.

And we have partners -- we're partnered with Oak Ridge National Laboratories and a group at the University of Southampton to do demographics, and quickly I'll run through these so I don't spend too much time, but the computer looks at different shapes, and it has -- there's an electronic pattern associated with every shape, and if those patterns are different enough, then you can differentiate those shapes from one another automatically, and interestingly enough, they trained the computer to pick out all the trailer parks in the United States and it took less than two minutes once they had trained the model to find all of those places.

So it's a lot of interesting things you can do.

The way the computer does this is it looks at shadows. It looks at edges, and it basically creates these patterns based on the different even patterns of structures and that gives them a lot of information not only about the structures but how big they are, how close together they are, the pattern of the structures and you can even do poverty mapping this way, so I was thinking for the East LA study it might be interesting to see the socioeconomic status of the communities, which they probably did.

But anyway, what we did was we classified the types of neighborhoods that we found in different communities. If you do it at the building level it's too
granular and there's too much air, but you can classify a neighborhood, and then we sent people out to do microcensus data or collect microcensus data in each neighborhood. So I'll just kind of skip -- this is the requirements for the project, but basically it's imagery. We do the microcensus data and we get an output we validate and then recollect additional data if needed. In Kano State in the north of Nigeria, which is where we started, we found six settlement patterns, and every place in the urban areas in Kano State falls into one of these categories. We sent out to ten different ones of these places and when they came back the computer was trained so it knows now in area A there's a certain number of people per square meter, and then it can apply that to the estimation.

And the same token for rural areas, three different types of rural settlement types in that state, and so we sent teams there to collect data on those types of places.

And so this is what a microcensus data set looks like. We define an area with a polygon and we tell the team to count every single house inside that polygon and find out how many people live there and that's what the numbers are that you see right there, the total number of occupants. And then they also collect demographic data so we know the different groups.

And we also did urban -- a huge set of urban polygons because urban areas are a little more complex. We took pictures of the buildings and made sure we understood if they're mixed use, single use, et cetera.

So we did a lot of collection on the ground.

We also collected five validation sets per state. We did 12 states across the north. It's the same except we put it aside and reserve to test the model to see if the output matches what our validation count says.

So here is what the output looks like, an urban area in Kano State in capital city, and here the population output is basically a 90-meter grid that overlays on top, and you see a number inside each box, and that number is the number of people that is estimated to live in that box. And we wouldn't really try and estimate this on a one-box level because there's just too much air. But if you take maybe 10 or 15 of these boxes you would get a fairly reliable estimate. I think we're at about 85% now for something that small, and the more the larger area that you select, the more accurate the estimate is.

The validation set that I showed you before, that particular sample we were almost right on the model predicted 304 people and the actual count was 300. They weren't all that close but they were all within 15% of the right number. There's another one in an urban area 19.3% variance. So we have evidence this works and we can prove that and we've established some statistical boundaries.

And we compared that to the census data, which I mentioned was collected in 2006, and every year The World Bank and other international partners decide how much growth to apply and it's about 3% per year. So since 2006 the population in all these -- these are LGAs, which is the level right below the state in Nigeria.

And they all got a 3% growth increase for the last 9 years. So it's kind of across the board, they were all increased the same.

And you'll see our model predictions. The brownish-orange highlight is areas where we predicted less than what the census says.

The blue is where we predict more.

And the white is where it's fairly close.

And what we noticed is all the ones under-predicted, where we predicted less, were rural areas. And the ones where the model predicts more is urban areas.

And we know from a lot of other information that in developing countries
especially, that most of the growth is in urban areas and people are moving from the rural areas to the urban areas. And while you might have an overall 3% growth, it's not equal in different types of areas.

So if you use this kind of data from the 2006 census with a 3% across the board increase, you -- by this time it's going to be way off. And you'll notice some of these, like if you look at Ungogo, third from the bottom on the right, we believe there's twice as many people there you would get if you used the census data.

Here is a picture of that place, picture on the left is imagery from 2006 and the imagery on the right is 2015, and you can see it's a huge amount of growth and we have pictures of rural areas in there. There's basically no change.

So the imagery actually backs this up.

So I just have a few more slides, but this is the important part. We wanted to create a user interface that everybody can use this kind of data and it's available.

So we have a website that you can log on to now. It's public. And you can go and get population and demographics estimates at the state level, at the local government area level, the LGA. You can go down to the ward level. And you can even go to the settlement level and it will give you a population and demographic information for every single settlement.

So this is a great tool

You can also self-select an area.

So we have a tool where you can either select a point and then a buffer area and it will give you all the settlements inside that buffered area. It will give you their target population. The total population and the distance of that settlement from the point that you selected.

So we use this for -- to plan outreach activities and vaccine requirements for healthcare facilities in the field.

There's lots of other ways you could use this.

You can also self-select a polygon, so you can create your own shape and it will give you the same type of output for that.

These are simple tools anybody can use and now all of this population data is really at their fingertips.

Here is an example of one of the maps we made for our measles campaign, which is coming up. Maybe we need one of those in northern California. But what we found was that people won't travel more than one kilometer to a post to get a vaccination because it's just too far.

So we use the maps and the population to cluster all of the settlements in each ward in groups where there was no settlement further than one kilometer away from a central point. Then we used the population tool to estimate the target population inside each of those clusters and that tells the people how many days they need to have a post there and how much vaccine they need and gives them an idea of what the -- you know, what the overall target should be for the activity.

And this population work is part of a larger project we have called core data layers where we collect administrative boundaries, settlement point, transportation network and population demographics and the gates Foundation is going to try to do in all of Sub-Saharan Africa and south Asia and we're starting that project later this year.

I think that is it.

Thank you.

>> Shell Culp: Thanks, Dr. Seaman. I was certainly intrigued when I saw that your skill set was so diverse, but I can see now from the presentation how that
all comes together and probably fuels your interest in a lot of different subjects.

Towards the end of Dr. Bhatt's presentation, he mentioned that place matters, and I think that that concept is pretty relevant for what you have shared with us. What sorts of work with data analysis would you like to encourage our audience to do that you add to the collective vision

>> Vincent Seaman: That's a good question. Place does matter. There's two types of geospatial mapping, and one is like the modeling, global perspective where you have a heat map, like some of the maps we've seen where it kind of shows you general trends and those kind of maps and analyses are important, and they tell you where you have a problem and don't, kind of in general. If you want to effect a solution or implement an intervention, you need to know names of names of places and how many people you're dealing with. I think both are important and I think sometimes we do a lot of the former but not as much of the latter.

I guess the other thing is, I used to work on domestic projects and I know that, like, for example, census data is available to everyone. I think you can go on Google Earth and get census tract data. I haven't seen a lot of tools like the one I just showed you that the average person could use just if you were doing some quick analysis. Gee, how many people live here? Or you needed to match a population denominator with another data set you had and you couldn't find, you know, the right numbers. Or you just want to fact check something.

I think that developing some basic tools like that would be really useful.

So I guess my work, you know, I think that what was maybe different about it is that there's a lot of data out there and we're getting more and more data all the time, and the ability to take data from the different programs or specialty areas, you know, data was produced for a specific purpose, but that data could be useful for somebody else. So that's a good reason to make it available, but then you also need people who can recognize that different data sets brought together might have more meaning.

So I don't know how to encourage people to do that except to just maybe try and be aware of and not afraid to go out and look and see what you can find that is relevant to what your work is.

>> Shell Culp: Yes, thanks. Do you think that there are themes that transcend cultures when it comes to using data, since you've done a lot of work abroad, Africa in particular, are there lessons learned from your work in other countries that can be applied here in the U.S?

>> Vincent Seaman: I think, you know, we hope the lessons learned in the U.S. will guide these countries to do a better job, but, yeah, there are things. I mean, number one, one mentioned already is that there's a mindset of data protection and that you have to guard data closely and it's not the NSA type data, but it's like data that could be publicly available, but for some reason the people who own it or manage it don't want to really make it available, and, you know, it's worse in these developing countries.

We have to fight with the governments to get them to release anything. And so it's a challenge, and we need to change that mindset, so I think that's on the negative side, but it's something we all can work on.

I think the other side is that people -- because of all the data that is being produced, we have a lot of different kinds of data, and I think establishing standards for the types to have data that is -- that are used within certain, you know, fields, is very important, because the ability to compare two different data sets -- I mean, even if you listen to the presidential debates, there's no way you can fact check a lot of things because everybody is using a different set of data or
different inputs or different denominators, and I think if we could all agree on use this particular source for denominators and these particular sources for inputs, then everybody's data would kind of agree and if we both do an analysis, then we can compare them, you know, effectively.

So I think having standards and getting people to share are probably the two things I am most concerned about.

>> Shell Culp: Audience questions. Sorry, I think I was talking without being unmuted.

Now it's time for audience questions. And thank you, Dr. Seaman for the follow-up questions and answers.

There is one question here that we have got that sort of pivots off of the answer that you just gave on -- regarding how governments are very protective of their data, and it's somewhat related, and here I would invite all of our panelists to provide feedback on these questions.

So let me pull that question up.

It's the notion of garbage in, garbage out. So the question is: What technologies are being used to collect data and make sure it is accurate? And Dr. Seaman, you just went over a lot of technology around geographic information systems and geospatial techniques and validations, but outside of that kind of technology, when we're looking at building models to be predictive, how much can we trust those models and how can we -- how are they being validated?

>> Vincent Seaman: If you're asking me to go first, I will, but I think the other panelists might even have a better insight.

It is true, garbage in, garbage out. And unfortunately the complexity of the modeling these days is beyond the capacity of most of us to really understand, like, the mathematics and statistics and all that.

So we're at the mercy of -- you kind of have to believe what people are telling us. And I think it's very important to know the source of the data and as much as we can about how it was done, but I think most of the time for me, if I know what the data -- what data was used for this analysis and specifically where did it come from and what the date was and how they -- which parts of that data were used, then people can at least check with their own experts if they want to. Too many times we don't have that metadata what went into this.

So, yeah, I will stop there.

>> Shell Culp: Sort of do your homework kind of, you know, know what sources you're using, what your models look like and how they were created.

Okay.

Andy, do you have anything you're doing or that you have seen in the civic technology area that would help you know, maybe validate models or cleanse data or sort of do something to eradicate the garbage before it goes?

>> Andy Krackov: It strikes me potentially these days there's -- I guess I could frame it as multiple sources of garbage, and that's something we haven't had before. The idea is that if we have all the different kinds of data -- and Jay was speaking to a lot of various sources of data, retail data, sensor data, and the list goes on, that maybe those can help verify or more traditional sources, survey data or the public health surveillance data that we typically use.

So I think all of that helps. And there also are these days a growing number of do-good data scientists who want to help in this arena. I can think off the top of my head, two data science firms that are around not necessarily to help commercial organizations solve their Internet advertising problems but to help
improve social conditions. Impact and data kind are two organizations that come to mind. They are trying to help with these challenges and we need to reach out and work with these organizations all the more.

>> Shell Culp: Jay, do you have anything to add?

>> Jay Bhatt: I echo what has been said, but I would also say that there is, you know, digital data from wearable technology. There is the data that Andy was talking about. There is electronic health record data. There is the algorithms and programming language that helps comb through the data, and I would also just again go back to something I said earlier, that correlation doesn't equal causation.

So you don't always need causation to make decisions based on data.

But the correlation can be enough. And, you know, we've done instances in Chicago, actually run a test based on the predictive model. So you can actually see if it plays out like it should or differently.

And then incorporate it into business practice.

So I think those are opportunities as well.

And, you know, I would say that using clinical data to figure out intervention, it's certainly happening, and recently one of the systems discovered that patients had a hard time accessing clinicians because of transportation and that's why they were missing appointments. And I think the extent that grocers are willing to sell and share their data, that you can try to combine purchasing data related to clinical data. And I know one of the questions was did I talk about Medicaid or not? When I talked about claims, I incorporated Medicaid. Medicaid is certainly powerful data as well.

The other thing is when we refer individuals to community and social services, some of you may be familiar with an organization called Purple Binder, which is digitizing services to community social services in Chicago and elsewhere, but it's important for the care team to know that they actually went and engaged in that service or not.

And so that closing the loop data that comes back can be helpful too.

And then the other data is around behavioral health. So if -- can we trend the data to understand what behaviors lead to particular outcomes or what behavior leads to potential behavior changes, what data leads to potential behavior changes. So that's helpful as well.

>> Shell Culp: So you mentioned in your presentation the idea of the social determinate of health. If there was one thing that you could pick out, what would you say is the most important change that can accelerate the availability and use of social or human services within the healthcare environment?

>> Jay Bhatt: I would say simply, yes, I agree, and one of the places is
certainly the electronic health records. The other places are involved with payers, and other consumer organizations as well.

Shell

>> Just to add to that, I found in my work with various communities across California, sometimes the issues we face is getting new data you know, for example, a current measure of poverty in the community. Sometimes what we face is getting more localized data, especially in a city environment, like Los Angeles or, say, San Diego, trying to really pinpoint what is going on. And that, I think, has really created difficulties for us and it's the same kind of difficulties that exist in rural communities. Sometimes they don't have a large enough cell to pinpoint what issues. We probably do need to get better with metrics and I'm hoping some of the sources we soon or already have available to us will help in that regard.

>> I'll just make a quick comment because I know we're running out of time. I agree with both of those. I think one thing that is going to impact the whole data revolution, I guess we could say is social media and crowd-sourcing type data. And if you think about it, that is data that people are using -- people are using that data on a daily basis to manage their lives and do things, and I think that we're just at the tip of the iceberg as far as figuring out how to use that and leverage that to inform, you know, us about, you know, health issues and really anything.

>> Shell Culp: Thank you, everyone. I think we're at the point of wrapping it all up here. And so I would like to take a minute to thank our presenters, Dr. Jay Bhatt. Hopefully everybody is going to get a chance to go out on vimeo and view Dr. Este Geraghty's video. It was very engaging and informative and I highly recommend it. Thanks to Andy Krackov, for the generous support from the California healthcare association and for a fantastic framing our forum for today and thanks to Dr. Vincent Seaman of the Bill and Melinda Gates Foundation for an interesting look at how GIS technology is helping to eradicate diseases in some of the nuances that go with that.

I would also like to thank our behind-the-scenes people: Holly Calhoun, who has been feeding me questions and Joanna Hathaway, who has been trying to keep me on track and make sure that our technology issues with the video did not overcome us, and I think she's done a fantastic job. Thanks both to Holly and Joanna.

And to our sponsor, the California Healthcare Foundation. Without their generous support we couldn't do that, and that would be terrible.

Next time we have our next forum in the series Forecasting the Future on Tuesday, September 9th at 11:30 a.m. Pacific and 2:30 p.m. Eastern, and we hope you find time to join us there.

And that is it for us.

Thank you for joining us today.

Have a great day.