- [Dr. Tcheng] Good afternoon everyone. This is James Tcheng. We will be getting started in about two minutes or so. Thanks for joining in, and we'll be on mute until then. Thanks for putting it on the screen.

- [Dr. Tcheng] Hello, this is James Tcheng. I am hearing the beeps slowing down. It's a bit like listening to a microwave popping popcorn. I'm gonna give the beeps just another 30 seconds or so, and then we'll get started. All right, why don't we go ahead and get started. This is James Tcheng, the principal investigator of the project entitled, Improving Healthcare Data Interoperability. On behalf of my colleagues at the Duke Clinical Research Institute, welcome to this afternoon's webinar. I would also like to take this opportunity to acknowledge and thank our sponsor, The Pew Charitable Trusts. The first point is a housekeeping measure. As a courtesy to everyone on the WebEx we're taking the prerogative to mute everyone as you log in. If you see yourself unmuted for any reason, if you wouldn't mind, during the presentation, re-mute yourself. At the end of the presentation, you can unmute yourself by depressing the microphone icon. Just as an announcement also, we will be recording this presentation. We will try to provide approximately 15-20 minutes at the end of our presentation for questions and answers. At that point in time you can unmute yourself to ask a question. You can also use the chat to ask us questions. I do appreciate the interest in this afternoon’s webinar. By our count, there are some 58 different organizations who are represented in the acceptances. And hopefully, the webinar will provide great information for all those who can join. Next slide. The agenda for this afternoon’s... The agenda for this afternoon's presentation is listed here. I am gonna provide about five minutes or so of some background. I hope you find it interesting, if not, amusing in the way I've put it together. Then we will spend most of the time talking about the project itself. We're gonna describe what it was that we're trying, what it was that we set out to try to accomplish by the project, take a look at work that’s been done by two of my colleagues, Anne Heath and Mary Williams, looking at ubiquitous CDEs identified through the review of registry content, as well as a look at national data models. And then we'll talk about, we have a slide to talk about the metadata that we’re at least proposing as a parsimonious approach to describing the CDEs sufficiently so that they can be developed into databases. Again, we will try to provide some time at the end for questions and answers. Next slide. So I thought I'd start off with this. This is a tweet from a fellow by the name of Paul Lomax, who I do not know. But it's that 'The most unbelievable aspect of the Star Trek universe is that every ship they meet has compatible video conferencing facilities.” So if Star Trek can do it, the question is, why can't we? Next slide. So, the real question is, how do we end up where we are in the state that we are? After all, the... If I could ask everybody to mute themselves if you're not on mute please. The journey to interoperability has been one that has been acknowledged and recognized by leadership all the way up to the President. As you recall back in 2004, President Bush established a 10 year goal to develop an Electronic Health Record, a Republican. A Democrat, President Obama pushed EHR development through incentives, specifically the Meaningful Use program, 2009, the HiTECH Act, via the ARRA Act, with the target of being full implementation by 2016. Next slide. If you look back at the enabling legislation that then created the HiTECH approach, one of the committees that was enabled was the HiTECH Standard, HIT Standards Committee chaired by Jamie Ferguson and John Halamka. Next slide. And pulling from their slides, from their original slides, I was struck by the fact that the approach that they were considering, in terms of
creating standards for interoperability, explicitly said that their clinical operations group was recommending that standards for interoperability occur at the borders, at the interfaces between entities not within an entity. And if you go to the very last line what they acknowledged was that there would be a whole number, a whole slew of methods for representing information at the native data level, but that the focus of the HIT Committee was specifically or explicitly not to be at the native data capture level, but instead at the boundary. So in a quick nutshell, the project that we are tacking here is challenging that original notion. And that is that should that be the place where interoperability is accomplished, that is at the boundaries, or should we actually be aiming for interoperability at the native data level? Next slide. I think of the answer is evidenced by where we are today, where we have ended up, where we have landed, especially with respect to the registry data entry, is largely represented by this particular pictorial. And that is what I like to call Swivel Chair Interoperability. I'd like to claim originality to that term. It is not. But I think it captures the concept very nicely. Next slide. The foundational issue is that we still don't use terminologies the same way from one context to the next. Next slide. And so the real question has been, how do we accomplish this role of having a single source, single approach for data capture that then can be reused across multiple different approaches? And the answer I believe is through this concept of native data interoperability, that the project is really focused on accomplishing. Next slide. So this gave rise to the improving healthcare data interoperability project, or as we have code named it The Pew Project, in deference to our sponsors, The Pew Charitable Trusts. And essentially, we took the opportunity that Pew has provided us to ask three simple questions. Using the registry community ways to convene the work, the first question is, what is the current state of registries? How are registries using data standards across, especially ubiquitous common concepts? And we'll get to what I mean by that in a few moments. But the question is, are registries actually a part of the solution or is there opportunity in terms of creating inertia, creating momentum by bringing the registry community together across these common concepts? Second group of questions that we wanted to ask was, what about the national data models? Where are we with, for example, OHDSI, for Sentinel, etc. Are we, or have they converged on using data standards, or are they in the same situation? And then the last question that we pose, the last part of the grant proposal was whether or not we could author the Easy Button. That is, put together an all in one package of best practice recommendations for these ubiquitous common data elements that database developers could then implement more readily, more easily than trying to cross reference data element after data element and figure it out for themselves. Next slide. I do wanna acknowledge that we have now 38 registries that have contributed content to the effort. They are all listed here. Hopefully, everybody who is a contributor is on this call because the purpose of this call is to update you as to where we are today, and then plan for the future. I hope you name is listed up there. If you're interested and you are part of a registry that has not been contributing content, it's not too late. Please consider sending us your content. Next slide. So what are we doing? This is the aims of the grant. I'm not going to really talk a whole lot about the details, but I did wanna include this in the slide deck so that for those of you who are interested in the slide deck that you can download it to see what the specifics were. We actually set out with, I thought, fairly respectable goals. That is, we wanted to just take a look at 20 different registries thinking that they would be representative. And then we realized, you know what, if we can do a few more that would be better. So we were about to know that. The criteria that we described in the grant for defining a ubiquitous clinical data element was that it would appear in more than 50% of a registry case report forms, or the data model thereof. And in fact, we found that that is an overly high bar as well. So we have had to reduce that bar. You'll see some of that work here in just a few moments.
Next slide. So with that, let me turn over, the discussion over to Anne, to Anne Heath, my colleague here at the Duke Clinical Research Institute. She's had the unenviable task, the Yeoman's task of going through all of the source material from the registries, trying to articulate, to identify all of the data elements that would likely be common across at least several of the registry, some of the registry content. And we'll give you a perspective of how far along she is and what she's already discovered. So Anne?

- [Anne] Thank you for the introduction, Dr. Tcheng. And also a big, big acknowledgement to Mary Williams who you'll hear from in a minute regarding the national models. What Mary and I have focused on in review of the 38 registry sources and the national models, is these, what are we looking at here? About 12? 15? Concepts of interests. The way that my work began in reviewing and abstracting data elements from the registry sources that were coming in was initially to abstract essentially every single data element and then pivot on those to find commonality. That ended up being a non-scalable, non-sustainable method of going about this. So then we reoriented ourselves and have focused specifically on these concepts of interest. Next slide please. Again, to reiterate Aim 1 that Dr. Tcheng just described, this is the portion of the project that I have been most focused on, which is reviewing the registry sources that we've received. Next slide. So we have received 38, and I have done a cursory review of all 38. I have done this targeted review of the concepts of interest in 24 of the 38. And the examples that follow will focus on three of those concepts, sex, laboratory data, and tobacco use. And what I want to emphasize in terms of sex is that the data element that we are looking for is sex itself, not to be confused with gender. So sex is biological sex as characterized by reproductive functions and secondary sex characteristics. Gender is the cultural concept of the state of being male or female as expressed by social or cultural distinctions. We're focused here on sex. And I'm gonna reiterate this a couple of times probably. In terms of lab data, you'll see here the specific lab values that we have looked for. And then tobacco use, specifically smoking history. Next slide. This overview, I know it's hard to see, and that's intentional. The objective here in this slide is just to show you, my attempt is to show you the degree of variability. Right here there are about 90, 95 data elements represented across, at this point, this is slightly out of date, looking at about 20 sources here. As Dr. Tcheng said a few minutes ago, we had to change our threshold. We had initially assumed that it would not be hard to find concordance in data across at least 50% of the sources. That is obviously not what we found. So I'd like, in the coming weeks, to play with this slide some more and see it mature maybe in a wordle, or a heat map, or something like that. But this is just a quick view of all of the variability. Okay, next slide. Okay, so getting into the details here. For the sex variable, again, not gender, but biological sex. Of the 24 sources that we have reviewed in depth, sex is represented or asked in some fashion in 20% of those. So that's 83%, in terms of exact concordance. So identical CRF label, or identical question on the data collection form, and identical permissible values. You'll see my counts here. So even among what we would think would be a simple concept, or a simple data element of sex, there is a great deal of variability. Next slide please.

- [Dr. Tcheng] Anne, sorry, before you go forward I just wanted to point out that the first four rows of this actually don't list the question as a question about sex. It asks the question about gender, which Anne has already pointed out. And then in the second column, we'll point out that there's one registry
that states, indicate the patient's sex at birth, and the answers are yes or no. So just a curious commentary on how data is captured and how data's collected.

- [Anne] And there's a chance that that was actually an error in the data dictionary. But I did review that a couple times to make sure that I wasn't looking at the wrong row. And I found interesting the last one on this list. And are you? That was the question on the case report form. I'm assuming it was a patient completed form. Okay, next slide. Okay, I know that this is probably a little hard to look at. This itself will demonstrate the variability across smoking questions. So of the 24, 13 of those asked some question about tobacco use. Not specifically cigarettes, but generally cigarettes, but some kind of tobacco use. So that's 54%. And of those 13, every single one of them asked the question differently. So that's why my concordance counts are one, and one, and one, and one, and one, and one. Is there anything you'd like to add about, okay. So next slide. Of the lab data, of those lab values that we were focusing on, interestingly, the case report forms that we reviewed didn't ask, or didn't, yeah, didn't ask any questions about the context of this data collection. So within a certain timeframe, with respect to the encounter or within a certain timeframe with respect to a procedure, that was not specified in the question. So I can't speak to that yet. Hopefully, we'll find some more of that as I continue. But even these fairly standard. I shouldn't use that word. But these fairly basic laboratory data are only represented across 25% of the sources we've looked at. And then you can see the individual percentages. The greatest is 25, and the smallest is 1/24, which is 4%.

- [Dr. Tcheng] So, Anne, a specific question regarding laboratory data. How often were these described in the data models or in any place that you could figure out how they were linked? How were they in terms of LOINC bindings, are these described in terms of LOINC codes or--

- [Anne] In a couple of the data dictionaries.

- [Dr. Tcheng] Okay.

- [Anne] But on the--

- [Dr. Tcheng] Strictly a minority.

- [Anne] Absolutely a minority. I think two. And certainly on the case report forms. When I just looked at the case report forms you didn't see the bindings there, but there were two, I believe. And so if you go to the next slide please, it was not really tenable to list out concordance in this table in the way that I had done in the other two tables because there was so much variability. But if we go concept by concept, or lab by lab, you can see here the way each one of them
[Dr. Tcheng] Keep going.

- [Anne] Okay. So that, we can go to the next slide actually. And I think, yes, I can turn it over to Mary now. So thank you.

- [Mary] Thank you, Anne. Thank you, Dr. Tcheng. Similarly to Anne, I was responsible for identifying data elements. But my task was to identify those same data elements where possible from the national model. Next slide, Becky. So this is just a listing of the national models that I reviewed for the clinical concepts of interest that were shown earlier in the slide deck. Next slide, Becky. And again, this is the overview. While not as exciting or extended as Anne's, there is a lot of variability here. And while there are only five or six models that were used, you can see, if you look at laboratory results, medications, and vital signs, there are a huge number of data elements surrounding those concepts. And that is because, you'll see later on, that there is such a wide variety of information collected about labs, about medications, about procedures even. And the vital signs is a bit of an anomaly in that I included diastolic and systolic blood pressure, pulse rate, pulse oximetry, height, weight, all of those things under the big umbrella of vital signs.

- [Audience member] What's the unit on the x-axis? What's the unit on the x-axis, there?

- [Mary] The x-axis?

- [Dr. Tcheng] Yeah, across the horizontal.

- [Student] The y-axis, sorry. The y-axis.

- [Mary] The y-axis initially started out being the number of sources, but because... Like, for example, with labs there are 30 elements in one source that are collected. It ballooned into something much bigger than the five sources that I was looking at.

- [Dr. Tcheng] So the y-axis is a simple count of the number of times the concept is represented. So the best way to think about this is that Mary is actually charged with looking at six different models. Basically, I think it's six different models. And so the end is, at the end of six would be actually pretty low. So you can see that in the national data models what is happening is that these ubiquitous concepts are appearing, for the most part, in all of the national data models with the exception of over on the very right-hand side of the substance abuse and risk factors. Now, with vital signs there are multiple
components to vital signs, like systolic blood pressure, heart rate, etc. So there are multiple counts there. With laboratory results, there's multiple labs, multiple medications. So, for example, the most vertical column, that's the lab results column, has 60 instances across the six data models. So each one of them has, if you will, 10 that we're looking at. So not the best way to present the data, to be very frank with you. But really, what's more important than this is the findings in terms of the conformance, the use of data standards, etc. So this is just a way to try to quantitify the number of data elements that we're looking at a little bit. Although it's presented in a similar way to Anne's graphic, it means a little bit different in terms of content. The primary reason just to show you this, was to illustrate what the quantity of the information was that we're looking at. So, Mary, if I can turn it back over to you.

- [Mary] Okay. Thanks, Jimmy. Thank you. Okay, so across the national models, looking at the sex data elements, again, there's very little concordance. Each of them have represented sex in a different way in terms of data element name, permissible values. So some are similar. They are, for the most part, completely different. Next slide.

- [Dr. Tcheng] So I'll again point out that even in OHDSI and in FHIR, this is referred to as patient gender, not sex. And we're using this as a very specific example just because of the lack of clarity. If you think about how we would translate this, the work product of the project then, the recommendation really is going to be that the data element name should be sex, nothing to do with gender, unless you're asking about gender, but that isn't what we are assuming that these registries, etc., that are being asked about. So gender being a very specific concept, that's not what we are tackling here. So next slide.

- [Mary] Okay, thank you, Dr. Tcheng. And because I do have recommendations from FHIR for these examples, I did provide them as screenshots in my presentation. So for the recommendation from FHIR as for the sex variable, it's required that you have a patient identifier, patient name, and it says gender, and also if it's available, the birth sex. So just informational purposes. This is what the FHIR, U.S. Core FHIR is asking or suggesting as data elements.

- [Dr. Tcheng] Yeah, and the reason to ask Mary specifically to take a look at FHIR is that our assumption, and here's where I would actually open this question up, not to ask the question of everybody on the call right at this moment, but to put it into the parking lot in terms of thinking about how to come up with the representation, the best representation, or if you will, the recommendation that we're going to come out with. We were assuming that FHIR would be one of the better places because FHIR's business is in data transformation. That's, if you will, the universal translator. Going back to the Star Trek analogy, this is the closest thing we have to universal translator right now. So that's the reason for focusing at least with these screenshots on FHIR, because that is the resource that we're going to be referencing over and over again. Mary. Yeah.
And this slide represents how the smoking question or tobacco use, because it does contain both, these questions are asked across the models. I was happy to see that there was some concordance between the CCDS and CCRS. But the others, just looking at them, are collecting this data differently, again, with data element name and the permissible value list. Next slide. And again, the recommendations for smoking status from FHIR, the requirements show that you have a fixed code for smoking observation, patient, of course. Date representing when this was last, and result value code for the smoking status. Next slide, Becky. Thank you. And as Anne said with labs, it was too unwieldy to try and identify any concordance between the sources. So what I've done here is list the source and the data elements that are part of that national model. And the red highlights are what's in common across the models that is part of the recommendation from FHIR for laboratory results. So as you can see, and as the graph kind of showed previously, Sentinel, OHDSI, PCORnet, ask upwards of 15-20 different questions per se about laboratory data, and that is marked here. Anything?

Yeah, so the relevance of this is that the national data models, if your data is represented, if your data, for example, in Electronic Health Records is represented in a national, in one of the national data models, moving the data from one system to the next, one data model to the next actually becomes a little less problematic. But the question that we were asking was, why is that? Is it because the data models themselves are pretty well fixed and you can do a translation, or is it because the data models basically have all conformed to data standards? And I think the answer is really more the former than it is the latter. Without going through this exercise and really trying to figure out the answer, we really would not have that. The second part of this is that, if you look especially at PCORnet, Sentinel, and OHDSI, and having had a fair amount of experience with the American College of Cardiology's registries, only a limited number of information, limited pieces of information, are asked and captured by the registries. And so, I used the word before, and I'll use it again. The question is, how to come up with a parsimonious recommendation where instead of conforming to one of the national common data models with all of the overhead that's asked of the common data models for laboratory data, can we come up with a more limited, and again, parsimonious approach that everybody can use as the common floor where, and I'm gonna pick a number here, but 90% of the information, or 95% of the information is there? The thing that is not included in, by the way, in these national data models that we're struggling with, again, something to put out there in the parking lot, is context. Anne mentioned it before, and I'll mention it again. At the registry level we believe is important, and this is part of her work to survey these registry data capture forms, is that most of the time these questions are being asked in context. That is, within 30 days of a procedure, or within 24 hours after the procedure, or the lowest value, or the highest value. As opposed to the way to see the information here represented, which is typically oriented around date times, as trying to say, this is the time and date that the specimen was obtained. And if you wanna figure this out then you have to understand. In order to understand the context it's a comparison of date and time. Next.

Next slide. Thank you, Dr. Tcheng. And then, of course, the recommendations from FHIR on labs is that you have a status and that it is a lab code, a LOINC code, and a result value. And if you have a value that you have the standard units involved. And even in the observation, which they're saying that
you should have, it doesn't necessarily have that point addressed. Like it's within 30 days of enrollment or context of the procedure.

- [Dr. Tcheng] Context.

- [Mary] So they're asking for a time, basically, and a reference range if there's one available.

- [Dr. Tcheng] Yeah.

- [Mary] So again that whole--

- [Dr. Tcheng] Yeah, Mary used the key word, and that is should have. And the question, and the way we're framing this project is not should have but have to have. And so that gets us back to this concept of parsimony. Next slide. All right, let me finish up with just a couple of comments about where we are today and where we need to go. In the interest of time, I did not ask Anne or Mary to put together any more slides about all the work that they've done. Let me assure you that this is only the tip of an iceberg. I asked them to focus specifically on three different exemplars, again, sex, laboratory data, and tobacco as being representative of the rest of the work. We have much of the rest of the work is underway, and actually tabulated. I think it's safe to say that looking at the other concepts that we suggested via our initial survey of the U.S. CDI, the common clinical data set, and just common sense as to what we would expect to see in the registries. That there are no other surprises that we haven't already represented in the slides, in the presentation, and in the slides that we're showing you here. But we obviously have a lot of work to do to bring it together and to put it together into a reportable format that then can be then used. The last then concept that I do wanna bring up is that one of the tasks that we set upon ourselves in addition to identifying these ubiquitous common data elements to identifying the best way to build them into a database, if you will, the representation of those data elements in the databases themselves, is to understand the other consumer, not just the database developer. But I'll call it the clinician with some reservation because one could argue that even the patient is a potential consumer of this. But to come up with a parsimonious list of metadata that is the minimum set of information that needs to be described in a spreadsheet format, or other format, that allows you to understand each data element in its entirety. And you'll see the list of metadata that's up there. What is it that you're gonna call it? The concept label. That's the human prompt for the case report form, or data entry screen. Line number three is, what's the definition of that concept? Some of these words I think need definitions. Some of them probably don't. If we code them, if we describe them via the label, that is, we code them the right way, last name and family name being the same thing. Maybe perhaps we'd describe it as family name parenthesis last name, or something like that. That doesn't need to be described or otherwise defined. However, sex probably does need to be defined just to make sure that it's absolutely clear what is intended by that term. Allowed values, allowed value definitions. The approach that we're taking is to, again, do the survey and then make a recommendation as to what the
best set would be based upon, again, predicate work. Not our work here. But we're in the business of aggregating the work that's out there. You'll see in line number eight, SDO bindings. This is to acknowledge that in the data standard's world, the standards development organizations, there's been a lot of time and effort. Part of the problem is that sometimes these things have multiple bindings. And again, the question is, what would be the preferred binding that at least we would recommend that everybody use in terms of building databases. Next slide. All right, so here's where we are. I do appreciate everybody being on this call. And I think it expresses an interest or perhaps a frustration is another word that could be used with the current state of data standards, and why we still are not there in terms of using them. Our gap analysis guess, excuse me, our project guess is that the gap analysis that we've just conducted would demonstrate that that the penetrance of data standards into the building of registry databases and national data models was lacking. And I think just in the examples that you seen, we've clearly shown that. Next slide. So how do we go forward from here? Well, we internally here at the Duke Clinical Research Institute, my colleagues, my informatics colleagues, along with several investigators and other contributors from across the United States. I wanna give a shout out to Seth Blumenthal, to John and Tom Windle at the University of Nebraska Seth is with the PCPI, and several others, Julia Skapik, in particular, have worked on this with us. This is not just a Duke effort. This is something that I think acknowledges that there's a large community. Our anticipation is that we will actually push out a work product that everybody can consume. And so, the last concept I wanna bring up to you is that in order to consume this, in order to build it into the databases, one of the things that we're already recognizing is that we need to bring a little behavioral economics into this as well. So for those of you who are looking for something to read this summer, I would encourage you to pick up this book, called Nudge by Richard Thaler and Cass Sunstein. Richard Thaler is the 2017 Nobel Prize Laureate in Economics for his work in behavioral economics where he describes the need for choice architecture. And essentially, that's what we're doing. We're creating an architecture where the choices that you can make in terms of building databases, or perhaps more importantly, refining existing work so that your database's interoperability registry, excuse me, interoperability ready. That your registry is interoperability ready. Needs to then accommodate these things. There really is unrealistic optimism out there. If interoperability were that easy we would be there already. And a decade later after ARRA, and the Meaningful Use Act, etc., I would suggest to you that we still aren't there. We do all suffer as humans from loss aversion. That is, the inertia favors stasis. So one of the things that I'm challenging all of you with is to think about the fact, where we are today is not where we wanna be, but how do we get to where we should be? And that is to have an interoperable environment. You'll have to overcome your status quo bias. We're providing you part of that by creating this "easy button" default option, the list of recommendations. But Cass Sunstein and Richard Thaler talk about the fact that it actually hurts when you have to change from status quo. And then the framing effects are something that, again, you need to think about. And that's how to convince your leadership to spend the resources to actually incorporate the suggestions, the recommendations that we're going to come up with in terms of the common ubiquitous data elements. Their representations at either the physical layer in databases, or if not the physical layer, in a representational layer in the databases of registries. Next slide. So this is where we are. We have a lot to overcome in terms of accomplishing interoperability. For those of you who think this should be kind of an "easy button," just push a button and it happens, I don't think that's where we'll find ourselves. This is going to take initiative, and effort, and resources on the part of everybody. I do appreciate everybody listening in today to understand kind of where we are, the logic of how we're gotten to where we are to understand what the plans are. By the way, the timeline is that we
will finish the work that is involved in doing the assays of the databases from the registries, the resources that are provided, as well as the national data models, sometime in the next month, the next three to four weeks. And then after that we'll be pushing out the implementation guide for everybody to review, comment on, and get back with us about. Next slide. So here is my last slide. I would like to, and I think we did a pretty good job sticking to the promise that we would leave about 15-20 minutes for questions and answers. But here's what I'd like to challenge you with, and that is to ask you, is your registry interoperability ready? After all, that is what we're trying to accomplish. And knowing what you do today about this project, again, thank you for your contributions. But the key work product, that is, the "easy button" artifacts of CDEs and their specifications, what will we need to do to improve that work product? What else should we be incorporating into this project? And what will it take for you personally to incorporate the work of our project here into your registry. So with that, I'm gonna close. And ask anybody who would like to ask a question to take yourself off mute. I believe you can take yourself off mute. Is that right, Becky?

- [Becky] Yes.

- [Dr. Tcheng] And then ask the question. And/or you can use the chat box as well. So do we have any questions in the chat box?

- [Becky] No.

- [Dr. Tcheng] All right, so I do have one question that came through via email before the webinar started. So perhaps, let me just start with that. And that is the question, "Will there be a provision "in the CDE specifications for a lay definition of each "term, giving the increasing focus on inclusion "of affected persons?" So that's a great question. I think that the good news about the work that we're doing here is that these ubiquitous common terms are so self-defining that many of them actually won't need a definition. Those that do need a definition will be both clinically relevant, or clinically appropriate, but also understandable by the, if you will, lay person, the patient themselves. So yes, we do anticipate that this is information that would be useful not just from a registry standpoint, but also for patient reported information, for the direct entry by patients of information, etc. I can't promise any of that right now, 'cause that wasn't, to be honest with you, the target. But just in thinking about this particular project, I think the answer is yes, that it will be understandable. A perhaps more important and relevant question is, what about other domain-specific developed, vocabulary development efforts? That's a little bit more tricky. I think that the work we're doing here will help inform that process. I am fine with adding another line to the metadata that says, lay person definition, something along those lines. We will take that under advisement in our final report back to Pew, and the manuscript, thereof, in terms of what should a parsimonious or minimal data, metadata specification include. All right, any questions out there?
- [Mike] Yeah, so I think this is a fantastic effort. I was hoping that I could actually possibly get these slides. This is Mike Waters from FDA. And I’d like to note that we’re trying to really help incentivize the ubiquitous use of LOINC codes, and making sure that LOINC code are assigned and utilized the same way every single time. And to do that, we try to work with registries to create un-ubiquitous coding manuals that will help that process and actually set registries up to be able to provide a little extra support in the assignment of LOINC codes. ‘Cause as we move forward, we’re trying to. And by we I mean the multi-stakeholder effort that is SHIELD to get harmonization interoperability for laboratory data. We are trying to make sure that as we begin to test and aid in implementation for a lot of the standard content that you’ve actually kind of laid out right here, that we can do this pretty smoothly. So if you could send me these slides that’d be great. And if anybody who actually wants to be involved that’s on this call in helping or collaborating in this effort, or how we can now actually serve this effort in particular, please just let me know or reach out and contact me.

- [Dr. Tcheng] Great, Mike. Thank you very much for that comment. This is Jimmy Tcheng again. And we appreciate. We greatly appreciate the support of the FDA in this. And you and I’ve chatted a bit about shield. We’re quite aware of the registry efforts to create these convening LOINC codes to address the situation where there is perhaps more than just one or two. Sometimes there’s multiple LOINC codes for a specific laboratory type that these convening or universal LOINC codes approach is indeed something that we intend to expose as well. That reminds me to add the comment that looking at lab data specifically, what we were looking for was actually commonalities, or in other words, a framework. How would you best capture the key pieces of information in a way that would allow you to capture any lab data? And the answer indeed is by LOINC codes. So we are definitely embracing that approach. Part of the, and I’m gonna call it, I keep calling it the “easy button implementation guide,” is to articulate how that framework should work as opposed to perhaps unnecessarily burdening the registry database developers with very specific recommendations on a lab by lab basis. For that specific example, it actually makes much more sense to have a unifying frame and to use LOINC in particular from a bindings perspective for those pieces of lab data. So thank you very much for that comment. The other comment I’ll make is that we are going to post these slide on the website that we have for this particular project. I’m seeing Becky’s head nod here. I know it’s hard for you guys to see her head nod on the phone. Just wanted to make sure I wasn’t saying something out of school. But we will send out the link to that website with a follow-up message after the webinar’s been completed here.

- [Becky] We just had a couple of questions come in. So if you guys will bear with me, I'll read the question, and we'll attempt to answer those as we can. So the first question is, can we provide more details about the targeted review of the 24 registries?

- [Anne] So sure, this is Anne. And what I meant by that was having received 38 registries and having done a cursory review of all 38, I had found, initially, I found what Dr. Tcheng and some others didn’t expect that I would find necessarily, which was so much variability. And with that great variability that I was finding and less commonality than perhaps that 50% that was expected, with further conversations and directions from Dr. Tcheng, we honed in on those specific concepts of interest that we listed in the
slides. And I very meticulously then went back registry by registry, and really studied how those concepts of interest were collected to try to better understand if the initial variability that I was seeing was in fact true across these concepts of interest. So that's what I meant when I said a targeted review. I have done that on 24 of the 38. And I'm just keeping a list for myself, and I'm continuing. So I hope that answers the question.

- [Dr. Tcheng] Yes, I think I can add one more dimension to that.

- [Anne] Yes, please.

- [Dr. Tcheng] And that is that my instructions to Anne, in terms of the targeted review, acknowledged the fact that registries are largely comprised, and largely meaning 80%, or perhaps a little bit even more than that, of domain specific concepts. So if it's a registry that's focused on coronary intervention in ST elevation myocardial infarction, it is highly unlikely that more than one more registry might be asking questions that are domain specific. That is, what time did the angioplasty occur, or what size stent was used, etc. So my instructions to Anne, don't worry about the domain specific content. What you need to focus on in the targeted review are concepts that are at least possible to appear in more than one registry, or more than one or two registries. So that's also the concept of the targeted review.

- [Becky] So we have a question from John Donnelly. He asks, "Has your team's research included "any review of the HL7 clinical information "modeling initiative working group? "The working group has been actively working "for about 12 months to harmonize "clinical models around the same concepts, "including a relationship to FHIR."

- [Dr. Tcheng] Yeah, so this is Jimmy Tcheng. Again, I think that, so the answer to the question is yes. I'm quite aware of work within HL7 with the CIMI working group. In fact, Stan Huff and I have chatted a number of times about this particular project. He's, and other folks at HL7, are eagerly anticipating the results. I'm a member of the clinical information interoperability council, which is modeling data concepts via Sem. We are an active participant and intend to, if you will, stand up the terms for formal modeling through the CIMI process. When I talked before about the, kind of the parsimonious or core minimum metadata, that anticipated actually putting it through the CIMI, putting these concepts through the CIMI process, and then doing the formal informatics modeling. So I think the answer to the question is, yes, we have to do this work first before we're gonna however be ready to stand up through the CIMI process.

- [Becky] We have a couple of comments regarding interoperability, more comments than questions. But I'll read them off in case they're not obvious. Agreeing that interoperability must consider native data
forms. Data transformation introduces alterations, errors, and omissions. And then interoperability is end to end or source to use. Agreeing--

- [Dr. Tcheng] And we agree with that.

- [Becky] Agreeing that interoperability must consider native data forms, data transformation. Yes. And then there's a suggestion at GitHub for some information about LVLs. So recent. LVL versus LVL can be addressed by annotations and post-coordinations. Mentioning SNOMED's compositional grammar as being one way to annotate the data elements to semantically relate them to one another. A number of requests for slides, and we will make those available. And then as stated data requires not just cataloging, an assortment of data elements but the specific context of use including providence, who, what, where, when, why, and for what purpose.

- [Dr. Tcheng] Yeah, so let me just make a comment about providence. This effort, it does not focus on providence purposely. So I would actually state that the providence question really is up to the machinery of the original data capture, whether it's Electronic Health Record or a dedicated system, whether it's by a swivel chair interoperability. Those processes that are used really reflect the providence of the content of the data. What we're really focused on is the data itself, and the representation, the capture and representation of the data at the database level. So yes, we completely agree of the importance for audit trails, for the understanding of providence, etc., but that is not within the scope of the particular project.

- [Becky] Okay, so we have another couple of questions here. Besides Pew, what are some of the other potential sources to work on interoperability? In my experience funding is scarce with taking on a broad interoperability project. They've been working on cancer check lists pathology reporting data and getting those into registries, but asking about funding mechanisms to take on projects like that.

- [Dr. Tcheng] Yeah, so that's a great question. I know we have several participants on the call who are from Pew. They've been asking the question of, where should we go from here? So that's one funding source. However, that also depends upon request for proposals, the typical granting requests. We were just fortunate enough to have asked this particular set of questions at a time where it seemed like it was the right set of questions, and then being able to convene the registry community around coming up with answers. I agree with the individual who asked the question. Finding money to do this type of work has been very difficult. We again, do wanna express our appreciation to Pew for supporting this work. However, you might find it curious, or might find it interesting that our original proposal, in terms of the amount of monies that we thought it would take to do the work, was much more than what we were actually finally contracted with by Pew. So we actually had to cut out a few things in order to be able to deliver on what the core work product is. So even having said that, money is limited. That is acknowledged. I don't have a magic funding source. I wish I had won the lottery, or done something like
that, create a foundation just dedicated to this. But if I can ask folks to buy lottery tickets, and then create a foundation from those of you who win, that might be a good way to go.

- [Becky] Okay, so we have a question here.

- [Mike] I have a note. Like so 4:30 today, I asked three grants that are going in to help support this. I'm doing my best to actually get another lottery ticket to help support these processes. But I'm writing grants like crazy. If you want to actually collaborate in writing some more grants, just let me know also.

- [Male] Thanks, thank you.

- [Mike] Alright.

- [Becky] That's excellent, thank you. So we have a question here, what will be included in the HL7 implementation guide? Is the focus commonness for Aim 1 or Aim 2 data models, or both?

- [Dr. Tcheng] I'm not sure I understood that question. Read it again please. Yeah.

- [Becky] What will be included in the HL7 implementation guide? Is the focus on commonness for Aim 1, like registry information, or for Aim 2 data models, or both?

- [Dr. Tcheng] Oh, so first of all, this is not an HL7 product. This is a work product of this particular grant. Having said that, I am a member of the HL7 common clinical registry framework working group. So the intent is to consume this work product in the CCRF work so that at the end of the day it gets put through the standard HL7 processes for public exposure, for comment, for updating, etc. So there is a mechanism for accomplishing this. We are focusing again, on the ubiquitous common data elements. The target audience is still the registries not the national data models. It is the registries. The reason for that is that we believe that if we can convene the registry community around this particular set of work, that we will actually create a lever that then puts pressure on Electronic Health Record vendors in particular, but electronic health information systems in general to comply, to lower the barriers, and to increase data liquidity by having to spend less time, and effort, and resources on data information.

- [Becky] Okay, I think we've got time for one last question. Do any of the registries that were reviewed have their datasets already in--
- [Anne] So the answer to that is yes. I don't have the specifics in front of me.

- [Dr. Tcheng] So how many--


- [Dr. Tcheng] Okay.

- [Anne] Not very many.

- [Dr. Tcheng] So I think the answer is a handful, and that's one handful not many, all right? So, this is Jimmy Tcheng. I do appreciate everybody's attendance today. We have recorded this, and we will make the recording available to you as well as the slides. Please feel free to contact myself at James.Tcheng@Duke.edu, or Becky Wilgus at Rebecca.Wilgus@Duke.edu with any other thoughts or questions. Thank you very much.