Transcript

>>Well, good afternoon everyone. My name is Ashley Green and I'm very excited to be here today to talk to you about the All of Us research program. I am with the Data & Research Center at Vanderbilt University Medical Center and I'm a former librarian and I'm really excited to talk to other librarians about our program. I'm also joined by my wonderful colleagues at the Data & Research Center. So if you have any questions, please feel free to drop them in the chat and they will be very happy to answer as we go along.

So today's presentation is a high level overview of the Research Hub and Workbench. During our presentation, we will learn how you can show your students, faculty and staff the data and how you can assist them with their projects or your own projects. You can learn how to leverage your role as a librarian and promoting and utilizing the All of Us research program on your campus. And you can learn to how to examine the steps for registration as well as see how to start a research project on the Workbench.

So first, we always like to take a moment to thank our participants who have generously given their time and biomedical information to the program. None of the science we hope to enable would be possible without them and their amazing contributions. And we can see the National Library of Medicine here on our slide of participants and community partners. So thank you so much for all that you do for all of us. And we are very lucky to have a strong, diverse group of consortium partners who worked tirelessly in promoting and supporting the program. So you can see here on this slide how there are many partners working together on All of Us, including participant centers and the HPO network, both of which work with our participants to gather their data. And it truly is a collaboration with the entire consortium and spreading the word and growing the program.

So what is the All of Us research program? Well, we are working toward advancing precision medicine for everyone. One of our aims is to have at least 1,000,000 participants from across the United States, and we truly value our participants. We view them as partners working with us to help us enable this science. And as we go throughout our presentation, you're going to hear me talk about ways that our participants are vital to our program and ways that we help to ensure their privacy.

So diversity is at the heart of the research program. Our data include a diverse population of participants, especially those who have been underrepresented in biomedical research, and our data are longitudinal, including our surveys, and they follow the participants. So the data combine both biological and social determinants of health on a large scale. And transparency is incredibly important to us because we want our participants to know how their data are being used by the researchers.

So the mission of the program is the acceleration of medical breakthroughs. We have 3 objectives working toward that mission. We do believe in creating and nurturing partnerships with our participants. And we're looking at delivering one of the largest biomedical datasets.

Our dataset is free to access and it is very secure. And we also want to grow and foster our community of researchers.

So let's take a closer look at our participants. Our participants are amazing. Our data set is not only rich, it's also diverse and the program aims to reflect the diversity of the country. 78% of our participants have at least one category of underrepresentation in biomedical research, which is what the acronym UBR stands for. Here, 44% are of non-white race or Hispanic/Latino ethnicity. 24% are over the age of 65. 9% have less than a GED, 26% earn less than $25,000 a year. 10% are sexual and gender minorities. And 7% live in a rural location.

OK, so now that we've had a chance to look at our program and learn more about our participants, we're going to start diving into the data and see what we have available and we're going to learn how we curate them. So we do have data from over 370,000 participants and we offer many types of data, including surveys, electronic health records, physical measurements and wearables. And don't forget, like I mentioned earlier, our data are longitudinal. We do also have genomic data available now. We're still in our beta phase. So that means we currently offer over 98,000 whole genome sequences and over 165,000 genotyping arrays. And the genomics data are only available in our control tier. And we're going to go over the different levels of access to the data in a few slides.

So currently we offer nine surveys for researchers, two of the latest surveys that we've added are the social determinants of health survey and the minute survey on COVID-19 vaccines. So currently, the social determinants of health survey has responses from over 57,000 researchers, and this survey includes questions on health, security, discrimination, neighborhood safety, social support and more.

And our COPE survey is our COVID-19 survey that we sent out to participants for several months from 2020 through 2021. And it looks at the effects of that COVID had on mental health of our participants. So more than 100,000 participants responded to one or more of the six surveys we sent out between May of 2020 and March of 2021. So the minute survey I mentioned a moment ago that was recently released. It asked the participants about their vaccine status, so we have that information available as well. And we will be adding more surveys in the future.

We're going to take a closer look at our curation process and see how we curate our data. So we at the Data & Research Center, we receive the participant data from our partners that we saw at the beginning of the presentation on that slide, and participant level raw data undergo quality control in our curation process to ensure participant privacy. And they are also harmonized. So these data are pushed through a pipeline about once a year on average. And we use the OMA data model and the data are kept on a secure cloud-based platform on their research hub and we use high quality security technology and strict privacy and security protocols to help keep the participant data safe, including what I find so fascinating, hiring outside companies to try to hack into the data.

So we have created multiple tiers of access to the data, so we're going to start from broad access that anybody can see and then we're going to go to a more controlled level of access. So on our left, on the left side of my screen, starting with the Public Tier, this is available to anybody. It is broad and it is a summary of what we have. It is aggregate data and we round it by 20. So anybody can look at this data using our data browser on our public website and we're going to take a closer look at the data browser so you can see how you can use it with your students and with the faculty and all of the tools that it has for you all. So the next level is the Registered Tier and this is where he must be registered with us to gain access to the data. So this was first launched back in 2020 and the Registered Tier is our row level participant level data. Researchers must register and must be approved to access it, and they access it through the Workbench. So in this level we have removed all direct identifiers and all free text data as well as obscure the date of health related events. Also, all the dates are shifted backwards between a random number between one and 365, and the shift is constant for each record, so temporality of events is preserved. And our next tier, excuse me, level of data that we have is the Controlled Tier. And this is where the genomic data are located that we saw earlier. So in addition to genomic data, we have more granular data from the registered tier available, such as the first 3 digits of our participant ZIP code and researchers must complete an additional layer of training in addition to the training that they did in the Registered Tier to gain access to the Controlled Tier.

So you may be wondering what the difference is between the Registered Tier data and the Controlled Tier data. So for both levels you have to register and you have to go through mandatory trainings. But the Controlled Tier has an additional level of training that you must complete before you will have access to that tier of data. And the Registered Tier, the row level data have had many identifiers removed, such as the dates have been obfuscated and only state level geographic data are available now. The Controlled Tier. it has the same data as Registered Tier but the data are more granular such as unshifted dates and expanded demographics. COVID EHR data have been added to this tier as well, and also this is where that genomic data is located. So if you're wondering how you can get access to the data, how your students, how your faculty can get access to the data, in a little bit we're going to see how to register for both tiers.

OK, so now that we have had a chance to explore the data we offer, we're going to look at The Research Hub and this is the website that is home to both the public data and the registered data as well as many of the pre tools we have for researchers. So while I can't go over everything today on the Hub, I am going to go over quite a bit. So I do encourage you to follow along. So I'm going to stop my slide deck and I'm going to move over to our Research Hub. And just a reminder, that is researchallofus.org. And the reason why I wanted to show you live how the how the website works is just in case you know maybe you're teaching a class or you have someone to come up to the desk and they're interested. This is how you can show them in real time all of the tools that we have available for them and for yourself. It's open to anybody. It's

free for everybody and you don't have to register. So this is a great tool for you as a librarian to use on your campus where you are.

So let's learn a little bit about the Research Hub. It is a secure cloud-based Google platform that contains the public data available for everyone to view, as well as the controlled access data that requires registration to view and analyze. And here on the public portion that we're going to explore, researchers can learn about the data and how to use it, as well as learn about the tools available to analyze it. And participants can also use the Hub to see how their data are being used by our researchers, which is one of the ways that we maintain transparency.

So right here on the Research Hub, when you come in, you'll be able to see our Data Snapshots. You can click on the link. Or when you go to the Data & Tools tab, you can see the Data Snapshots. Now our Data & Tools tab is going to be very useful for you when you're teaching this, when you're showing others how to use it, because there is a lot of information in this tab. So we're going to explore the snapshots first.

So what we have here on the snapshots page is it gives an overview of the aggregated public facing data and this is updated daily and hey, we have now reached over 600,000 participants. Yesterday when I came here, we were still in the 500,000. So that's really exciting to see and the snapshots page provides demographics, geographic distribution and more, and it also provides the number of enrollment counts which we just saw. So if you're here, for example, in the United States on our map, you can roll over to see how many participants you have in each state. I'm currently in Tennessee. So I see that I have over 3,000 participants from the state of Tennessee, but I'm originally from Alabama and Alabama has over 23,000 participants. And you can go further down. You can see the breakup and the breakdown of race and ethnicity, gender identity and age. So let me scroll back up to the top.

We're going to take a look at when we come over to Discover we're going to go to the Research Projects Directory and this takes a few moments, a few seconds for it to come up, so don't be too alarmed if it takes a moment. The reason why I wanted to direct you to this is this page is amazing because it is. It lists all the current research projects researchers are conducting. So when a registered researcher begins a new workspace on the Workbench they provide their projects, research, purpose description, populations of interest and more, and this information appears right here for everybody to see. It's another way that we do ensure transparency, and it's updated about weekly. Now here's something that could be really interesting, really useful for you. When you click on it, all the information is going to populate. But let's say you are teaching a class on how to write a scientific question. Well, you have examples all over the project directly from actual researchers on how they construct their scientific question or how they construct anticipated findings. You can also see who is part of the research team, so maybe you see somebody from your own institution. So the research directory alone has a lot of information for students, for faculty, and you can search by the title, the owner, or the scientific questions being studied.

So another part I want to show you is under the Discover tab. We're still here. We have our directory. We're going to look at our Publications page and what's great about the Publications page is that this lists peer reviewed articles that All of Us researchers have published and the page includes links to those articles. And we do require researchers to alert us when their manuscripts are being published, about two weeks before publication. Because we like to list them here and we want to celebrate their amazing work. And it's also yet another way that we have transparency for our participants. So if you're interested in seeing the published manuscripts of researchers, you want to go here. If students are interested in learning about what it looks like to have a published manuscript, to have a published peer reviewed article, you can go here to the publications page. It's free. You can have the links and you can show them examples of actual researchers and how they have gone through the peer reviewed process to be published. OK. So we're going to go back to our Data & Tools tab, but please feel free to explore all these tools.

So going back to our Data & Tools, we have looked at our program. We've looked at our participants and we've seen an overview of the data. So now we're going to look at one of my favorite tools, which is the Data Browser, and this is a way that you can explore the data we have available without having to register. So how you can use this is you can show students, faculty, anybody what we have without them having to register first. This is a great way to see if we offer the types of data that you are interested in. And we do have a lot of students, even high school students who have conducted research using just the data browser. So the data browser is-- it provides an interactive view of the publicly available participant data. So anybody can learn about the data types and the quantity of data that All of Us collects. And this is where the aggregate data are located. So when we were looking at our different tiers, this is part of our open tier that anybody can see. So the data that we have here, they are deidentified, they're rounded up to 20, users can explore it, they can look at electronic health records. They can look at genomic data, physical measurements and all of our survey questions. So this is not updated daily. You can see when the last update is, it's about-- it's updated about once a year. And what's great about this is that you can use the search tool to look at your topic of interest and the results will be automatically generated below.

So for me I am going to look up asthma because my father suffers from it. I'm very interested in conducting research projects on asthma. And when I do that I can see that it automatically populates all the types of data we have related to asthma. I've got two EHR Domains. I've got conditions and procedures and I have two survey questions. The personal medical history as well as the family health history. So let's take a closer look. When I click on conditions, I'm going to see the top ten conditions of asthma that we have in our in that participants have. And you can go further down and you can see all of the other conditions listed with asthma. You can see our participant counts. You can see the percentages when you're here. You can also click on it and it's going to break it down to sex assigned at birth, age, as well as the sources. Now the sources, these are the SNOMED codes and the OMOP codes. Let's see that again. Our SNOMED code and our OMOP code. And when you're here, you're going to see-- you're going to have--

It's very interactive. So there are a lot of things you can click on. If you have questions about things we have that the eye icon you can see next to it and you can get more information about it. So the Data Browser is very, very useful. If someone comes up to you at the at the desk and they want to know if we-- they're interested in working with All of Us, but they're not sure if we have the data and they might come up to you and ask do we-- does All of Us have data on breast cancer? You can take them straight to the data browser and you can show them all of the data we have and the breakdown counts of it.

So next when we go back to the Data & Tools, we're also going to look at our Survey Explorer. And the Survey Explorer is in both English and Spanish, and this is where you can learn about all 9 of our surveys that we offer. So we do have 9 surveys currently. We always keep on adding more surveys and more data in the future. So be sure you come back to this. But if you have students who are interested in understanding how to use surveys, the breakdown, some of the questions. This is a great place to take them to.

OK, so now we have a looked at-- (How do you?) We've looked at the data we've looked at how to explore the row level data and the one other thing I want to show you before we start to learn how to register, if you have someone who comes up to you and they do have a question and what you need to do-- what is a great tool that I love, is when you click on Support it is going to take you to our User Support Hub. You can type in the question here. You can type in any of the topics and it will populate. We have ways on how to get started with the data right here, we have videos how to work with it. Any questions you may have or your students may have or the faculty you work with and that they may have, they could come straight to the Support Hub and ask their questions. Also, we do offer free webinars throughout the week. These are drop in hours and you can ask any questions you have for office hours. And so if you want to know when that is, you can click on our event calendar and you can find out when our next office hours are what the topics are and any other help that you may need. OK, so let's go back to our Research Hub. So the Support tool is really really handy.

Now if you have someone who's ready to register, when you click on the Register button it will take you to our six steps that we have for registration. So we've already looked at Step 1. We've already looked at the data that we offer and we're going to start with Step 2. So Step 2, you must be affiliated with an institution that has a data use and registration agreement with us. Now we are still in our beta phase. So that means currently only US-based academic, nonprofit or healthcare organizations can enter into our data use and registration agreements and only US-based researchers can register. And we'll have a link right here that we're going to explore in just a moment that will show you where all of all the institutions that have an agreement with this in place. So Steps 3 through 4, once you've checked to be sure your institution has an agreement with us, you're going to register by creating an account and you will also verify your identity with login.gov. And for Steps 5 through 6, you will complete the code of conduct training & the data user code of conduct. So steps 2 through 6 can be done at once or as you go along. And it typically takes-- Typically it takes about 2 hours to complete.

So let's go back to Step 2 with our Data Use and Registration Agreement. So when you click on this you will find several pieces of information. First you will see all of this institutions that have an agreement in place with us, and then you will see the level of access that their institution has agreed to have or its researchers, whether it's only the Registered Tier or both the Registered Tier and the Controlled Tier. Now all institutions that are listed have access to the Registered Tier, but some have requested access to the Controlled Tier for their researchers. So you can go through here. You can search for your institution, you can search in the search bar. If you do not see your institution listed here, you can submit a request to get the process started. So this link will take you to a form and you'll fill this out. And once we have the form completed, we can start that process. So if you or another researcher are at an institution that has not signed off on the Controlled Tier access, maybe you're at Alabama A &M. You can see on the right-hand side the contact we have listed at the institution. So if you're needing Controlled Tier access, you can reach out to the contact we have on file and you can ask them if they can get that process started for you. So if you have any questions as you go along, especially if you have any questions with registering or with a login.gov right here on the right-hand side is a help button and you can click for help for login.gov or you can ask any questions at any time. Our support team are amazing. They respond within one to two business days.

So we do have a lot of tools to help you out to help your students out, to help other faculty members to get you started. Well, I hope you'll take a moment to kind of explore all the data that we offer right here and how you can utilize this where you are at your institution.

OK, so now that we've looked at the public facing data and how you can access it, I'm going to take us back to my slide presentation and I am going to show you a little bit more about the Researcher Workbench. So what we just saw is open for anybody. You can use it in the classroom at the desk, in your library, anywhere you are. It's free and it's open for everybody.

Now we're going to look at the Researcher Workbench and this is where you are required to have to register with us to gain access. So I'm going to show you what it looks like to set up a workspace and give you ideas of how you could think about it, where you are at your institution. So the Workbench is where the row level participant level data lives, and all of the analysis that you want to do will be conducted within a workspace and the workspace is your place to store as well as analyze your data for a specific project. So once you've completed your registration that we saw earlier and you're ready to begin, you will receive your login information in an e-mail, once your account is set up. And to get started exploring the workbench, once you have access, you're going to click on the researcher login button and I'll take you to this page and then you will sign in.

So let's learn a little bit more about the Workbench. It is where you will create your research by creating a workspace and a workspace is where you will build and analyze your data set. You can access our point and click data analysis tools here and in a workspace you can select participants based on specific criteria using the cohort builder tool. We have another tool, the dataset builder. It allows you to select the variables and the values of interest that you are

interested in researching in your cohort and then you can save it and export your data set to a notebook to begin analysis. Now a really great feature of the Workbench is that it does enable and facilitates team science. So our workspaces are very collaborative, so you can share your workspace with your research team and what's great about it is that they could be at your institution or they can be at a different institution. However, all team members must be registered with us and what's really great about this is if you have a research idea in mind or a scientific question in mind, but maybe you're not familiar with the coding to run the analysis you can add a team member to your workspace and the other Member can run the analysis. There are so many ways that you can think of for how your faculty can use this in their classrooms. Maybe they want to collaborate between one course and another and learn how you can use biology in one class and how you can use informatics in another class. And so there are a lot of ways that you can be very collaborative in working with anybody in your institution.

And one of the questions we get asked a lot is, is there an IRB approval needed? So we use a data passport model that means we approve the user. You don't have to have a research project in mind when you register and you do not have to gain IRB approval from us. So we have had many researchers who create a workspace to test the data first and to explore it. And now, as part of the data passport model we do ask for a detailed description of your research each time we create a workspace. We saw that earlier on our Research Project Directory and even if you're looking, you're just using the Workbench to kind of explore the data and play around with it you'll still have to provide that information for us.

Also, another big question we have gotten asked. Can more than one researcher explore the same topic? Yes, it is not a first come, first serve. We do not limit how many researchers can explore the same topic, so there may be other researchers conducting the same type of research as you or research-- the same type of research as your other-- as your students.

So let's go ahead and look at what at the Workbench and see what it looks like. So when you log in, you will see the landing page and this will show you the workspaces you've created as well as those that have been shared with you. Now you can see my workspaces here on this slide. Also I have more workspaces than what you see here, so they do like to hide under the See All Workspaces. So when you click on that you can see the full amount that you have. So you can see all the workspaces I have. I do have a lot of workspaces on asthma and you can also see your recently accessed items that you've built to be analyzed and also at the bottom you will see the videos and tutorials we have created. These are short videos. This is great to show if you're in a classroom or if somebody needs a small little snippet to learn about the Workbench or anything. Directing them to the videos would be very helpful for them.

So when you're ready to begin starting a new workspace, you will click on the plus sign (+) and so after you've clicked on that plus sign (+) you'll be taken to this page you see right here, and each researcher must fill out all fields. This is where you will select if you are using the Registered Tier or the Controlled Tier, if you have both tiers. And the banners next to each will denote if it'll be publicly displayed on the directory that we saw earlier. So we do ask you to be

as detailed as possible. You can edit your description as your work progresses at any time. And on the right-hand side. Right here. You'll see how it looks on our directory that we saw earlier. So you can see how the fills go from one site and have a look on the directory.

Now an important characteristic of the All of Us participant cohort is that it does include groups that have been historically underrepresented in biomedical research. So we do ask you to indicate if your study will focus on a specific population, or if you intend to study a phenotype disease or condition with a focus on a comparative analysis of a specific group, for example a group based on race and ethnicity, gender or age.

OK, so when you click on the About tab, this is where you can share your workspace with others. This is where you can go in and you can update your purpose if you need to, add your work progresses and just like all that Data Browser, we do have a lot of helpful tips and hints for you scattered throughout. So when you see the eye icons we have a lot of information there so be sure you play around and see all the stuff we have available. Now when you click on the Data tab, this is where the fun begins and this is where you can access our point and click tools and begin the research project. So these tools include the cohort builder for selecting participants to include in your dataset. It also includes the concept set selector, which is a data set tool for selecting the variables and values of interest that you want to research in your cohort to be used in the dataset now, as you can see on the right-hand side we do have a lot of tips that I've mentioned earlier, so be sure you do look at that because there are a lot of tools to help you out. They were very helpful and handy for me when I began my first project.

So taking a look back at my research project, Asthma Developing in Adults. For my cohort builder, I included participants over the age of 25. I had all genders and ethnicities and for my concept set I included asthma, EHR data, physical measurements and surveys.

Now, once you build your cohort and select the data you wish to analyze, you're going to save the dataset and you will export it to our Jupyter Notebooks to begin analysis. So the Jupyter Notebook is an analysis tool where you can use human readable language as well as code to narrate your analysis. Now you can think of this like your lab notebook. Currently Jupyter Notebook only use R and Python coding languages. Helen Plinker also available for genomic data analysis. We do have plans to add more analysis programs in the future and one of the things I love about working with All of Us is that we take the feedback we receive from researchers and we like to incorporate that and we have heard that a lot of people would like to have other coding languages besides R and Python. So that's what we're doing. We're working on adding that. Now in the meantime, we do have a workspace created with code snippets that you can use as well as other help tools for using R and Python. Now please note that you're not able to download the data and take it off the cloud and the reason why is that this is done to ensure participant privacy. So what you see on the screen here is one of our colleague's notebooks. And it is located in the Featured Workspaces section. We're going to look at that in just a moment, the workbench, and it's open for any registered user to use. And this notebook

was created specifically to be a tutorial on how to get started. And it also includes how to install packages and do common queries with the data.

So I have given you a very high level overview of the platform and the tools and now we're going to look at the support tools we offer. We did look earlier at our Support Hub. We're going to look at that again. But before we do that, we're going to look at the Featured Workspaces as was just mentioning to you. So these are located in the workbench and this is where you can access our tutorial workspaces. So these workspaces offer example notebooks on how to work with the data. And what's really great about them is that you are able to duplicate them and edit them on your own workbench and they do also provide help for your coding. And our support team offers New User Orientation each month as well as office hours to help answer any questions you may have. And you can use the calendar that we saw earlier on the Support Hub to see when the next New User Orientation is or when those office hours are. So our drop-in office hours happen weekly on Tuesdays at 1:00 PM Central Time, and you can come in or anybody can come in at any point during the hour and-- any of our registered users, they can come in and they can ask any questions they may have, whether it's a basic questions or very detailed questions about their code, we always have our team of experts on hand and Fridays also have office hours, but these are biweekly. And they're also at 1:00 PM Central Time and these Friday office hours have themes on how to use the data and tools. And just like the Tuesday office hours, they're also led by our experts and our analyst. The Friday office hours are recorded and they're posted on our YouTube channel, also on our support hub. So if you're not able to make it to one of the Friday office hours, you can still see it on the on our YouTube channel. Now when you register and you've gotten access, you're going to get that e-mail I mentioned earlier and the e-mail will have Zoom links to the office hours and user orientation. But if you misplace it, you can always reach out to the support team and they can always send you those links.

OK, so diving back into our User Support Hub, I do have a QR code in case you need it. And like I was mentioning earlier to you, the Support Hub, it offers a comprehensive set of resources to help you with your projects. It's open to anybody to access on the Research Hub we saw earlier and you're going to find a ton of articles and resources to help guide you through your data analysis journey. This is great if someone if a student or a faculty member or anybody else on campus, or even yourself. If anybody has questions, you can look on the Support Hub, you're going to find a lot of documentation to help you out. Very user-friendly. I used them a lot when I was first getting started and I found it very helpful from me and just like the Data Browser, we do have a search tool at the top so you can search by your topics that way quickly. And you can also contact our help desk with any questions you may have. They are very responsive. They will typically answer you within one to two business days and you can get you there. Their information here on the Support Hub, you can click on the help icon on the help button. You will see on the entire Research Hub and you can reach out to them that way. And we do offer a bimonthly newsletter, so if you'd like to know all of the information that's going, that's the newest updates coming out or new announcements we have, or if anybody you know might be

interested in it, you can use our QR to sign up. Or you can use our link to sign up. You will also find this on the Research Hub toward the bottom of the page. But this is a great way to stay up to date on all of the news and see some of the new spotlights we have featuring our researchers and some great tips to help you out. So we do hope that you join with us. You're welcome to sign up for that newsletter I mentioned.

You can find us at any of our social media outlets also right here. This e-mail address, this is our support team's e-mail and you can reach out to them with any questions you may have. And they're very responsive and always happy to answer any of your questions. So that is it from me.