

All of Us Research Program Researcher Workbench Onramp Summary Report June 24, 2021

Welcome to the All of Us Research Workbench Onramp Event

Stephanie Devaney, Ph.D., Chief Operating Officer, All of Us Research Program

Dr. Devaney welcomed the attendees and noted that she has been part of *All of Us* from its inception. She said that the program's mission is to accelerate health research and medical breakthroughs, enabling individualized care for all of us. The program's core values, shared across all partners, help drive its mission in a way that is equitable and inclusive. The program emphasizes diversity from many dimensions and among both participants and researchers.

When the program began, the first data release felt very far in the future. This event, devoted to use of the data that participants have shared, is important to her personally. The high number of trainees attending is important to the program's goal of democratizing the use of the data.

The program intends to ensure that the platform it is building—including the data, tools, functionality, and support—serve the needs of many researchers, not only those who have always had access to research resources. It aims to engage researchers across the country and, eventually, researchers outside of academic institutions, including those at foundations or commercial entities and community and citizen scientists.

The Researcher Workbench launched in beta just over a year ago. The program has received valuable input from the more than 875 researchers who have used the Workbench since then. Today's event is another opportunity for the program to receive feedback from researchers. Dr. Devaney said that she hoped for constructive, interactive dialogue. During the first part of the event, researchers and trainees using the Workbench would provide their perspectives.

All of Us Research Program and Workbench Overviews

The All of Us Research Program: A Program for All of Us

Mitchell Lunn, M.D., M.A.S., FACP, FASN, Stanford University School of Medicine/PRIDEnet

Dr. Lunn said that he is an LGBTQ+ researcher and a primary care provider at Stanford's LGBTQ+ Health Program.

Nurturing relationships, one of the four prongs of the *All of Us* mission, differentiates it from many federal studies. The goal is to get one million people from across the U.S. to participate in a federal longitudinal study for decades. Doing so will feed another prong: creating the largest, richest dataset ever, with access for the public. The emphasis on relationship building and diversity makes *All of Us* unique.

All of Us seeks diversity based not only on race, ethnicity, and sex assigned at birth but also on geography, income, disability, and a variety of other factors. Its goal of having 75% of its cohort come from Underrepresented in Biomedical Research (UBR) communities and 50% from racial and ethnic minority communities is unprecedented. Another focus is diversity of data types. The program includes participants as key stakeholders in its design and conduct, another exciting departure from typical government-run research studies. Participants help design the research, policies, and procedures. Involving participants as partners helps the program respond nimbly to feedback from community members.

Dr. Lunn provided an overview of the current protocol for participants. Participants can enroll online or through HPOs. Currently, they must be 18 or older. Participants are asked to share their EHR data, respond to surveys, have physical measurements taken, and provide biospecimens. The goal is to perform whole genome sequencing for all participants, so the program will produce the largest resource of whole human genomes. The program also collects data through wearable devices.

More than 315,000 participants have completed surveys, more than 263,000 have provided physical measurements, and more than 203,000 have agreed to share their EHRs. More than 63,000 participated in a COVID-19 impact study. The large number of participants enables studies of populations that have been understudied and disaggregation of subpopulations that have been aggregated in the past.

A network of people are working to involve UBR communities. The program has engagement partners, ways for people to volunteer directly, and plans to engage specific populations. Dr. Lunn displayed a slide showing community and provider partners representing a variety of organizations and constituencies, including Stanford. Other program partners include HPOs, the Biobank, the DRC, and genomics partners.

Building a diverse set of researchers is a key priority. The program aims to have a researcher cohort as diverse as its participant cohort and to make program resources accessible to trainees and early-stage investigators. Younger researchers bring new ideas and fresh perspectives. The program also plans to enable nonacademic researchers, including community-based and nonprofit organizations, to use its data.

All of Us Researcher Workbench as an Ideal Environment for Trainees

Romuladus (Romey) Azuine, Dr.P.H., M.P.H., RN, Division of Technology and Platform Development, NIH All of Us Research Program

Dr. Azuine said that he is a product of two historically Black colleges and universities (HBCUs) and of mentorship from three minority physician-researchers. He knows what it means to be a trainee in an institution without many resources and working under mentors without a lot of funding. He was excited to speak to the next generation of emergent investigators and trainees.

Trainees have limited time and resources with which to conduct research. The *All of Us* Research Hub provides an opportunity to produce projects quickly.

The [Research Hub](#) provides information about *All of Us*, privacy rules, data access, and the program's data; direct access to custom tools; answers to frequently asked questions; and a help desk.

Dr. Azuine described the data pipeline. The data collected from participants go into the raw data repository. From there, the data go through a data curation process that includes data harmonization, privacy preservation, and quality assurance. The data proceed through the Curated Data Repository to the Research Hub, where researchers can access data through a secure Google platform.

The Research Hub has two components. The public component contains the Data Browser, Data Snapshots, information on the participant surveys, and a help desk. The Data Browser includes summary statistics from the program's database and information on the types of data available. It is interactive and does not require a login.

In the restricted component of the Research Hub, researchers can access the data dictionary and concept set selector, build cohorts, create Jupyter Notebooks, and have questions answered. The process for gaining access is straightforward.

All of Us uses a data passport model. Once approved, a researcher can conduct multiple studies without seeking additional IRB approval. Researchers compute within a cloud-based system with no restrictions on data size or bandwidth. During the beta testing phase, the program is providing every registered user with \$300 in credit to cover computing and storage costs.

For now, researchers are required to have eRA Commons accounts. The program is asking researchers who use the Workbench to provide feedback on what does and does not work.

Dr. Azuine encouraged attendees to visit researchallofus.org. Anyone who is interested in becoming a participant should visit JoinAllofUs.org.

Promoting Ethical and Responsible Research on the Research Hub

Kate Blizinsky, Ph.D., and Sonya Jooma, M.A., Office of Policy, NIH All of Us Research Program

Dr. Blizinsky said that stigma is a process with multiple potential outcomes, from direct risk to people to systemic, institutionalized, and structural discrimination, as is the case with racism, sexism, homophobia, and xenophobia. Scientific research can contribute to and propel the stigma process. *All of Us* does not want that to happen, especially with the resources that originate with its participants.

The program's participant population is one of the most diverse in the history of science. Some of the groups and communities represented have been omitted or abused by the scientific community in the past, and many bear a significant stigma burden. The program does not want to perpetuate harmful practices, but it does want people to use its resources. Program leaders believe that the path to combating health inequities is through research, but only if the research is done well. The policy team carefully considers what it means to do research well, ethically, respectfully, responsively, and competently and hopes that researchers will consider it as well.

The program places trust in its data users and expects them to take their responsibilities seriously. It also provides tools to help them.

Ms. Jooma said that the program's approach to promoting appropriate data access and use includes education and support for data users as well as mechanisms for accountability.

The primary component for education is the training on responsible conduct of research that all data users must take. The training is specific to *All of Us* data resources and includes information on research ethics, how the program defines diversity, preventing stigmatizing research, terms of access and use, and privacy and security. Users must pass a knowledge test at the end of the training. The program plans to add modules specific to the controlled data tier, which has not yet been released but will include genomic and more granular data.

One goal is to remind users that they are working with data from real people who have contributed their time and information. The participants are trusting the program and researchers to keep their data safe and conduct research that will help their communities.

The program has developed agreements outlining expectations that institutions and individuals must sign. For example, researchers agree not to try to reidentify participants or download data. The program also has several [policies](#) outlining additional expectations.

The Resource Access Board (RAB) oversees compliance with the Data User Code of Conduct. Users can consult the RAB when they begin a project. The RAB also conducts periodic workspace audits and reviews workspaces that have been flagged by the public. All [project descriptions](#) are available for public review, and anyone can flag projects.

Testimonials from Current Workbench Users

Kelsey Mayo, Ph.D., NIH All of Us Data and Research Center–Vanderbilt University Medical Center

Dr. Mayo shared perspectives from researchers who have been using the Workbench and ways the program tries to lower barriers to drive scientific innovation and research.

Dr. Mayo said that Patrick Wu, Ph.D., was working on his dissertation on repurposing drugs using EHR data and gene expression signatures. Dr. Wu found about 69 candidate drugs that might lower cholesterol and used EHR data from Vanderbilt University to study them. He was asked to replicate his work using another dataset. Within one week, he was able to do so using *All of Us* data. Dr. Wu told the program that he appreciated the fast approval process for gaining access to the data. The short timeframe for replicating his study speaks to the program's overall mission of accelerating health research.

The program's access process involves six steps. If a researcher's institution has an agreement with *All of Us*, the researcher can complete steps 3–6 in as little as two hours. This process is one way the program aims to lower barriers to access while protecting participant privacy.

The program also strives to lower the barrier of cost. Dr. Wu's project cost less than \$20. Average user spending is \$11, and it is covered by the credits that the program provides. Access is free. Cost of computing is about \$0.27 per hour, and storage costs less than \$0.01 per gigabyte per month.

The program has heard from researchers that the dataset and data model allow them to replicate study findings from other large datasets. The principles of FAIR (findable, accessible, interoperable, reusable) science are woven into the way the platform is built.

All of Us built the Workbench with collaboration in mind. Within the Workbench, researchers can work on multiple projects and take on different roles within them. Researchers can also collaborate within and across institutions.

The program aims to understand and learn from the challenges that beta users encounter. The dataset is large, complex, and growing, so learning the data model and how to query the data are key challenges for researchers, especially if they are not familiar with programming in code. To support researchers in these areas, the program provides an integrated user support hub and support team, documentation, a community forum, and a help desk within the Workbench. It also holds biweekly office hours and monthly orientations to address questions.

**All of Us Research Program
Researcher Workbench Onramp
Plenary Session, Hour 2
Meeting Summary
June 24, 2021**

Exploring the Data: How Trainees Are Currently Using the *All of Us* Dataset

Ámbar Marrero Sierra, University of Puerto Rico; Kathryn McDougal, Ph.D., Towson University; Nicholas Giangreco, M.A., M.Phil., Columbia University

Dental Care Affordability Among the *All of Us* Study Sample

Ashley Green introduced Ms. Marrero Sierra, an undergraduate student in the Natural Sciences Program at the University of Puerto Rico at Cayey with a concentration in biology. Ms. Marrero Sierra aims to earn a D.M.D. and a Ph.D. and work toward a better health system and world.

Ms. Marrero Sierra presented her research project on dental care affordability. She began by listing the three main financial barriers to oral health: Patients cannot afford care, their insurance does not cover a needed service, or they are unable or unwilling to spend money on oral health because of limited resources. To improve oral health care, research is needed to improve our understanding of barriers and the personal experiences of people who do not have access to care.

Ms. Marrero Sierra and her team began their project by determining their variables of interest within *All of Us* survey data. They collected information on health care affordability from the Data Browser and stratified data by age group and gender. The question of interest was which groups reported that they could not afford dental care. The team compared the data with national averages and existing literature.

Intervention is needed to increase dental care among *All of Us* participants, especially men and the elderly. The team identified ways to facilitate access to dental care, including policy-level health interventions, working for an improved health system, informational campaigns, and providers who use sliding-scale or modified pricing. Interventions should correspond to life stage and age; for example, pediatric and adult populations have different needs.

Ms. Marrero Sierra thanked the *All of Us* team and the Minority Student Research Symposium students who advised her on this project. She encouraged attendees to use the data to investigate and to work for a better world.

Using the *All of Us* Researcher Workbench in a Course-Based Undergraduate Research Experience (CURE)

Ms. Green introduced Dr. McDougal, who is a clinical assistant professor in the Department of Biological Sciences at Towson University. Dr. McDougal focuses on providing opportunities to experience authentic research within undergraduate courses.

Dr. McDougal outlined the parameters of a CURE. Students perform research in a classroom environment. The difference from a traditional classroom lab is that students design the projects with a partner. This way, they gain experience in collaboration. The research is novel and relevant. Iteration is an important component. Students have time to redo their experiments; for example, they may want to try different conditions or tweak their code. CUREs help more students get involved in research, because some students might not have thought about asking to work with a professor or might have felt uncomfortable.

Dr. McDougal planned a CURE on genetic analysis in medicine. Because the *All of Us* dataset did not yet have genetic data available when she first offered the CURE in spring 2021, the course changed to Data Analysis in Medicine. The target audience was students interested in medicine, medical research, or bioinformatics who had taken a genetics course and a college-level statistics course. The class met virtually and had two undergraduate teaching assistants (TAs). Dr. McDougal said that TAs are necessary for this kind of course.

Dr. McDougal chose the *All of Us* dataset because she liked the program's commitment to diversity and knew that it would appeal to students. She also wanted to choose a dataset that would be welcoming for undergraduates, and the \$300 credit meant that there were no lab fees for the students.

Her student population was made up of 12 undergraduates, all juniors and seniors. Less than 30% were non-Hispanic Caucasians, and more than 70% were female. All were majoring in biology or molecular biology, biochemistry, and bioinformatics, although some were double majors. The TAs' majors were biology and computer science.

All of the students' work was done in the Researcher Workbench. Most students used COPE Survey data. Research topics included the effects of COVID-19 on anxiety, symptom severity in relation to age, and the epidemic's impact on physical activity. The physical activity group also used some data from wearables.

Another group compared access to healthcare and quality of life in people with HIV, using data from the Overall Health and Healthcare Access and Utilization surveys.

Two groups used EHR data, which turned out to be a complex undertaking. The two groups encountered some technical issues and were not able to complete as many analyses as they wanted, but they still had some results by the end of the semester.

All of the groups went from not knowing any R to completing a project and presenting a poster. The students presented to the class, but when activities return to campus, Dr. McDougal aims to have her students present at the student research poster session held each semester.

Dr. McDougal asked her students for feedback. The students found it helpful to have some education and practice with coding before they entered the Workbench. The students who watched introductory videos and went to office hours felt more comfortable in the Workbench.

Based on her experiences and student feedback, Dr. McDougal provided 10 tips for undergraduates and young trainees:

- Don't be deterred by feeling intimidated.
- Start with a strong experimental design.
- Learn or refresh coding skills before you begin.
- Watch the Getting Started, User Support, and Office Hours videos.
- Take your time building your cohorts and concept sets.
- Educate yourself on the Observational Medical Outcomes Partnership data model, the common data model that *All of Us* uses.
- Look at your data. If you do not understand the data, you will not know how to analyze the data—and if you lose some rows or have duplicates, you will not know unless you check.
- Use code snippets and tutorial workspaces.
- Keep your code neat and organized. This will save time in the long run.
- Do not be afraid when technical glitches happen. The support team responds quickly, and everyone has glitches.

Pediatric Data from *All of Us*

Ms. Green introduced Mr. Giangreco, a Ph.D. student at Columbia University. Mr. Giangreco's thesis work involves developing biologically inspired statistical and machine learning methods to detect and evaluate medication side effects through child growth and development.

Mr. Giangreco said that he would talk about how he and others in the DRC unlocked pediatric data from *All of Us*, even though no children are enrolled.

Mr. Giangreco is a biochemist, programmer, and computational biologist with a translational, data-driven focus. His thesis combines statistics, pediatric data, and child developmental biology. This work required him to engineer his own data sources to investigate his questions, which took a lot of time and effort. He found that having lots of observations across childhood accelerated his work, and integrating omics or sequencing data with clinical data helped test and generate hypotheses. Unlocking observed retrospective pediatric data became a major interest of his, so he was happy to be involved in unlocking pediatric data from *All of Us*.

All of Us collects health records from participants, and some of the records come from their childhoods. Researchers can derive a participant's age at a clinical event or outcome by using the date for a measurement or observation and the participant's birth date. All of these are standard fields in the common data model, which made using the platform and the data easy.

Mr. Giangreco and colleagues found that nearly 20,000 participants have donated EHRs from their childhoods, and the records span multiple decades. The dataset contains a wealth of information for testing and generating hypotheses.

Mr. Giangreco and colleagues observed that diagnoses of obesity and morbid obesity have increased significantly over time. Also, participants' past obesity diagnosis or lack thereof corresponded with their BMI, which opens up further research questions.

All of the team's data pulling, processing, analysis, and visualization was completed on the *All of Us* platform. No data download was required. Access was easy; it was like logging in to email. The Jupyter Notebooks environment is familiar to many people. Mr. Giangreco said that using the *All of Us* data made his Ph.D. work far more impactful.

Discussion

Louisa Stark, Ph.D., asked Dr. McDougal how her students chose their research questions. Dr. McDougal said that she left the decision up to the students. She asked them to use the Data Browser to see whether the program had data that could help answer their question. She also asked the students about their interests at the beginning of the course and grouped them accordingly.

Wrap-Up

Paul Harris, Ph.D., and Josh Denny, M.D., M.S., expressed their appreciation for the exciting presentations. Ms. Green thanked the attendees and adjourned the meeting at 4:56 p.m.

**All of Us Research Program
Ask Me Anything with All of Us and DRC Leadership
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021**

Panelists: *Paul Harris, NIH All of Us Data and Research Center–Vanderbilt University Medical Center; Andrea Ramirez, NIH All of Us Research Program; Sheri Schully, Division of Medical and Scientific Research, NIH All of Us Research Program*

Facilitator: *Adrienne Roman, Ph.D, NIH All of Us Data and Research Center - Vanderbilt University Medical Center*

Discussion Summary

This hour-long breakout session provided an opportunity for attendees to connect with leadership from *All of Us* and the Data and Research Center (DRC). Panelists described how the Researcher Workbench is currently in its beta phase, with more than 800 registered researchers and over 600 research projects. Opportunity remains to continue building out the 47 workbench and its capabilities. Participants emphasized the importance of working with communities and participants as partners in designing and participating in research on the Workbench. Amid the discussion, participants recommended the following ways to improve the functionality of the Workbench:

- Continue implementing regular updates and additions to the Workbench
- Consider implementing new opportunities for participants to provide data, such as an additional blood test, alcohol biomarker tests, or biomechanics data
- Allow researchers to link AoU data with publicly available datasets
- Extension and expansion of the Data Browser, with examples and structures to assist community members in asking research questions
- Provide opportunities for nontraditional partners to collaborate with AoU, such as entrepreneurs and technology partners
- Establish a method to adjust AoU data to be representative of the U.S. population

**All of Us Research Program
Overcoming Obstacles: A Facilitated Discussion for Beginning
Workbench Users
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021**

Panelists: *Brandy Mapes, M.L.I.S., Melissa Patrick, M.B.A., Francis Ratsimbazafy, Ph.D., Cassie Springer, M.P.H., JoAnn Smith, M.P.H., and Gage Rion, M.Eng., NIH All of Us Data and Research Center–Vanderbilt University Medical Center*

Facilitator: *Kelsey Mayo, Ph.D., NIH All of Us Data and Research Center–Vanderbilt University Medical Center*

Discussion Summary

This hour-long breakout session shared information on how to use the *All of Us* Researcher Workbench. This session began with an explanation of the three different tiers of the Workbench, the four categories of data types, and several tools within the Researcher Workbench designed to help researchers create a study cohort using program data. In addition, attendees received information on importing and storing external data within a Google bucket. Amid the discussion, participants recommended the following ways to further develop the Workbench:

- Harmonization of data from different sources, including lab data
- Share more information on the release of the controlled tier and the types of data available in it
- Make participant nutritional information and dietary intake data available on the Researcher Workbench
- Addition of extensions to the Jupyter Notebook
- Host additional learning sessions on how to apply weights or codes to *All of Us* participant survey data
- Distribute visuals showing how common dataset models and categorizations work using OMOP

All of Us Research Program
Facilitated Discussion: Efforts to Engage Diverse Researchers & Institutions
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021

Panelists: *Melissa Basford, M.B.A., Vanderbilt Institute for Clinical and Translational Research–Vanderbilt University Medical Center; Justin Hentges, M.P.P., Division of Engagement and Office of Finance and Management, NIH All of Us Research Program*

Facilitator: *Courtney Wallin, Ph.D., Division of Engagement, NIH All of Us Research Program*

Discussion Summary

This hour-long breakout session highlighted the *All of Us* Research Program's commitment to building an accessible and robust Research Hub and Workbench, as well as a diverse and inclusive researcher cohort. Panelists described how the *All of Us* Consortium is working with researchers to understand how the program can create both an infrastructure and an ecosystem for diverse researchers at different career stages, with different backgrounds and expertise, and affiliated with different organizations. These efforts are synced with the program's diversity, inclusion, equity, and accessibility efforts and NIH's UNITE Initiative. Amid the discussion, participants recommended the following ways to encourage additional researchers to join the Workbench:

- Take steps to support the use of the platform, including funding opportunities for junior faculty
- Motivate institutions to get involved and secure a DURA through tailored pilot programs, especially with HBCUs
- Offer additional programming languages, such as STATA and S.A.S.
- Offer training tools, e.g., a synthetic dataset for undergraduate students

**All of Us Research Program
Facilitated Discussion: Generating More Impactful Hypotheses Using
the All of Us Research Hub Tools
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021**

Panelists: *Paul Harris, NIH All of Us Data and Research Center–Vanderbilt University Medical Center; Andrea Ramirez, NIH All of Us Research Program; Sheri Schully, Division of Medical and Scientific Research, NIH All of Us Research Program*

Facilitator: *Adrienne Roman, Ph.D, NIH All of Us Data and Research Center–Vanderbilt University Medical Center*

Discussion Summary

This hour-long breakout session provided an opportunity for attendees, including new and prospective Workbench users, to ask questions and learn more about how to use different components of the Research Hub, such as the [Data Dictionary](#), [Data Browser](#), and other *All of Us* tools to generate a hypothesis and assess feasibility of their proposed research. Panelists promoted use of the Data Browser as a first step in the process of generating workspace descriptions and projects. Panelists described the various types of data available to researchers, such as EHR data, physical measurement data, Fitbit data, and survey responses on a variety of demographics. During the discussion, attendees suggested the following additions to the Workbench:

- Information on adverse drug reactions
- Environmental exposures of participants
- The linkage of genetic and clinical data
- Information on the role of *All of Us* data in NIH grants and funding announcements
- Time and date stamps of data

All of Us Research Program
Facilitated Discussion: Meet the Researchers & Trainees on the
Workbench
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021

Panelists: *Ámbar Marrero Sierra, University of Puerto Rico; Kathryn McDougal, Towson University; Nicholas Giangreco, Columbia University*

Facilitators: *Ashley Green, NIH All of Us Data and Research Center–Vanderbilt University Medical Center; Minnkyong Lee, Ph.D, Division of Engagement, NIH All of Us Research Program*

Discussion Summary

This hour-long breakout session provided an opportunity for attendees, including trainees interested in pursuing research with *All of Us* data, to meet current researchers on the Workbench. Panelists described the process for registering on the Workbench and encouraged attendees to review the project directory of existing research projects. Panelists also shared possible improvements to the Workbench, such as streamlining the electronic health record data and fixing technical errors. Amid the discussion, participants were interested in learning about the following topics:

- Overcoming technical difficulties while using data on the Workbench
- Resources to support new researchers on the workbench, such as the Help Desk, Slack, code snippets, tutorials, and featured workspaces
- The research publishing process
- Creating a community among the researcher user base, possibly by emphasizing current Workbench features more (i.e. the User Support Hub Community Forum)
- New opportunities for researchers upon release of genomics data

**All of Us Research Program
Facilitated Discussion: Conducting Responsible Research:
Understanding Program Requirements and Users' Responsibilities
Researcher Workbench Onramp Breakout
Meeting Summary
June 24, 2021**

Panelists: *Katherine Blizinsky, Ph.D., Office of Policy, NIH All of Us Research Program; Sonya Jooma, M.A., Office of Policy, NIH All of Us Research Program; Subhashini Chandrasekharan, Ph.D., Office of Policy, NIH All of Us Research Program*

Facilitator: *Subhashini Chandrasekharan, Ph.D., Office of Policy, NIH All of Us Research Program*

Discussion Summary

This hour-long breakout session provided an opportunity for attendees to learn about the *All of Us* Research Program's requirements and expectations of researchers to conduct ethical and responsible research with the program's scientific resources. Panelists emphasized the importance of protecting participant privacy and preventing harm to individuals, groups, and communities, provided information on relevant *All of Us* data use policies, Responsible Conduct of Research training course, and assistance available to users. Amid the discussion, participants recommended the following:

- Increased access for citizen scientists and researchers from institutions without an established research program
- More readily-available information regarding the program's plan to update data regularly