Building a digital health research platform to enable recruitment, enrollment, data collection and follow-up for a highly diverse longitudinal US cohort of 1 million people: The All of Us Research Program

Mark Begale, Dave Klein, Dave Klein, Sherilyn Sawyer, Jacob McCauley, Sepideh Shokouhi, Mark Verdecia, Scott Sutherland, Letheshia Husbands, Deepti Joshi, Alan Ashbeck, Marcy Palmer, Praduman Jain

Submitted to: Journal of Medical Internet Research
on: May 03, 2024

Table of Contents

Building a digital health research platform to enable recruitment, enrollment, data collection and follow-up for a highly diverse longitudinal US cohort of 1 million people: The All of Us Research Program

Mark Begale[1] BS; Dave Klein[1] MBA; Dave Klein[1] MBA; Sherilyn Sawyer[2] PhD; Jacob McCauley[3] PhD; Sepideh Shokouhi[1] PhD; Mark Verdecia[4] PhD, MBA; Scott Sutherland[1] BS; Letheshia Husbands[5] MPH; Deepti Joshi[1] PhD; Alan Ashbeck[1] BS; Marcy Palmer[1] BS; Praduman Jain[1] MSc

[1]Vibrent Health, Inc Fairfax US
[2]Boston VA Research Institute VA Boston Health Care Veterans Administration Boston US
[3]University of Miami, Miller School of Medicine Miami US
[4]Food and Drug Administration Silver Spring US
[5]The Task Force for Global Health, Inc Decatur US

Abstract

Background: Longitudinal cohort studies have traditionally relied on clinic-based recruitment models, which limit cohort diversity and the generalizability of research outcomes. Digital platforms can be used to increase participant access, improve study engagement, streamline data collection, and increase data quality; however, the efficacy and sustainability of digitally enabled studies rely heavily on the design, implementation, and management of the digital platform being used.

Objective: The National Institutes of Health’s (NIH) All of Us Research Program (AOU) is an ongoing national, multiyear study aimed at building a large research cohort that reflects the diversity of the United States, including minority, health disparate, and other populations underrepresented in biomedical research (UBR). We sought to design and build a highly secure, privacy-preserving, validated, participant-centric digital research platform to recruit, enroll, and engage AOU participants from diverse backgrounds.

Methods: AOU applied digital research methods to facilitate multi-site, hybrid, and remote study participation and multimodal data collection. We collaborated with community members, healthcare provider organizations, and NIH leadership to design, build, and validate a secure, feature-rich digital research platform based upon the core values of AOU. Participants were recruited by many methods, including in-person, print, and online digital campaigns. Participants accessed a secure digital research platform via web and mobile applications, either independently or with research staff support. The participant-facing tool facilitated electronic consent, multi-source data collection, including surveys, genomic results, wearables, electronic health records, and ongoing participant engagement. We also built tools for study staff and researchers to provide remote participant support, study workflow management, participant tracking, data analytics, data harmonization, and data management tools.

Results: We built a secure, participant-centric digital research platform with engaging functionality used to recruit, engage, and collect data from diverse participants throughout the United States. As of April 2024, 87% of participants enrolled via the platform are from UBR groups, including racial and ethnic minorities (46%), rural dwelling individuals (8%), those over the age of 65 (31%), and individuals with low socioeconomic status (20%).

Conclusions: This digital research platform demonstrated successful use among diverse participants. We built a user-friendly, participant-centric digital platform with tools to enable engagement with individuals from different racial, ethnic, socioeconomic, and other UBR groups. These findings could be used as best practices for effective use of digital platforms to build and sustain cohorts of various study designs to increase engagement with diverse populations in health research. Clinical Trial: N/A

(JMIR Preprints 03/05/2024:60189)
DOI: https://doi.org/10.2196/preprints.60189

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Building a digital health research platform to enable recruitment, enrollment, data collection and follow up for a highly diverse longitudinal US cohort of 1 million people: The All of Us Research Program

Mark Begale[1], Dave Klein, MBA[1], Aisha Montgomery, MD, MPH[1], Sherilyn Sawyer, PhD[2], Jacob L. McCauley, PhD[3], Sepideh Shokouhi, PhD[1], Mark Verdecia, PhD, MBA[4], Scott Sutherland[1], Letheshia Husbands, MPH[5], Alan Ashbeck[1], Marcy Palmer[1], Praduman Jain, MS[1]

[1]Vibrent Health, Fairfax, VA, USA
[2]VA Boston Health Care, Veterans Administration, Boston, MA, USA
[3]University of Miami, Miller School of Medicine, Miami, FL, USA
[4]US Pharmacopeia, Rockville, MD, USA
[5]The Task Force for Global Health, Inc., Decatur, GA, USA

ABSTRACT

Background

Longitudinal cohort studies have traditionally relied on clinic-based recruitment models, which limit cohort diversity and the generalizability of research outcomes. Digital platforms can be used to increase participant access, improve study engagement, streamline data collection, and increase data quality; however, the efficacy and sustainability of digitally enabled studies rely heavily on the design, implementation, and management of the digital platform being used.

Objective

The National Institutes of Health’s (NIH) All of Us Research Program (AOU) is an ongoing national, multiyear study aimed at building a large research cohort that reflects the diversity of the United States, including minority, health disparate, and other populations underrepresented in biomedical research (UBR). We sought to design and build a highly secure, privacy-preserving, validated, participant-centric digital research platform to recruit, enroll, and engage participants from diverse backgrounds.

Methods

AOU applied digital research methods to facilitate multi-site, hybrid, and remote study participation and multimodal data collection. We collaborated with community members, healthcare provider organizations, and NIH leadership to design, build, and validate a secure, feature-rich digital research platform based upon the core values of AOU. Participants were recruited by many methods, including in-person, print, and online digital campaigns. Participants accessed a secure digital research platform via web and mobile applications, either independently or with research staff support. The participant-facing tool facilitated electronic consent, multi-source data collection, including surveys, genomic results, wearables, electronic health records, and ongoing participant engagement. We also built tools for study staff and researchers to provide remote participant support, study workflow management, participant tracking, data analytics, data harmonization, and data management tools.

Results

We built a secure, participant-centric digital research platform with engaging functionality used to recruit, engage, and collect data from diverse participants throughout the United States. As of April 2024, 87% of participants enrolled via the platform are from UBR groups, including racial and ethnic minorities (46%), rural dwelling individuals (8%), those over the age of 65 (31%), and individuals with low socioeconomic status (20%).

Conclusions

This digital research platform demonstrated successful use among diverse participants. We built a user-friendly, participant-centric digital platform with tools to enable engagement with individuals from different racial, ethnic, socioeconomic, and other UBR groups. These findings could be used as best practices for effective use of digital platforms to build and sustain cohorts of various study designs to increase engagement with diverse populations in health research.

Keywords

Longitudinal studies; Cohort studies; Health disparate, minority, and vulnerable populations; Precision medicine; Biomedical Research; Decentralization; Digital Health Technology; Database management systems –

INTRODUCTION

Longitudinal cohort studies have historically relied on clinic-based recruitment models which limit participation and reduce cohort diversity in many ways. [1, 2] System level barriers, such as distance to the research site and clinic-based eligibility, can negatively affect participant recruitment and study retention. Socioeconomic barriers such as income, education, and health insurance status also limit research participation in minority and other health disparate groups. [3-5] The resultant homogeny in research cohorts reduces the generalizability and validity of research outcomes. [6-8] Digital platforms can minimize barriers to research participation, such as transportation costs, site access, and time commitment, and significantly improve participant access and engagement. [9-11] As a result, digital platforms are being increasingly used in health research to enable accelerated, more accessible, and more reliable real-world data collection. However, as longitudinal cohorts increase in breadth and scale of data, the efficacy of digital platforms can vary profoundly depending upon the platform's capability to meet the ever-increasing needs of the study over time.

Appropriate implementation of study activities via digital platforms can minimize many of the commonly reported barriers to research participation. Recent studies, such as MyHeart Counts and Health eHeart, show how digital enablement of traditional research methods may improve engagement of persons from diverse backgrounds, including those who do not commonly engage with healthcare systems. [12-14] Digital technologies also reduce stakeholder burden related to collection, curation, and sharing of health data.[15] Therefore, digital platforms may be effective tools to improve participant engagement, increase cohort diversity, broaden geographical reach, and streamline curation of diverse research datasets.

Digital research platforms must also accommodate users of varying digital aptitudes. This requires consideration of sociodemographic factors that impact the use of digital technology, such as age, disability, rurality, education, income, culture, digital access, native language, and literacy. [16, 17] Digital research platforms can be adapted to accommodate these differences and thereby reduce participant burden and build trust with communities that are underrepresented in health research. [18] Existing commercial and academic tools for electronic data capture were built primarily for research teams, are not designed to meet the preferences of participants interacting with them.[19] Therefore, it is important to create digital tools with engaging functionality for diverse participants that also meet the evolving needs of longitudinal cohort studies. We aimed to build such a platform to support nationwide enrollment of 1-million diverse participants in a novel multiyear, longitudinal cohort study.

OBJECTIVES

In this article, we will describe various tools of the digital health research platform designed to effectively engage diverse participants from various backgrounds and collect multisource health data in a large community-based longitudinal cohort study. The All of Us Research Program (AOU) is an ongoing nationwide initiative aimed to recruit 1-million participants from diverse cultural, socioeconomic, demographic, and geographic backgrounds and to collect data that is generalizable, accelerates biomedical research, and improves outcomes for all groups.[20,21] The broadly inclusive eligibility, recruitment, and data curation needs of AOU required the development of a highly adaptable, comprehensive digital platform able to accommodate multifaceted, complex study requirements. The study also needed a platform with modifiability that would allow study staff across 1,200 sites to conduct study operations and rapidly deliver customized, community-specific engagement (Figure 1). These complex needs led to the development of a comprehensive digital platform design capable of supporting data expansion and the addition of ancillary studies over the extended study timeline of ten or more years.[22]

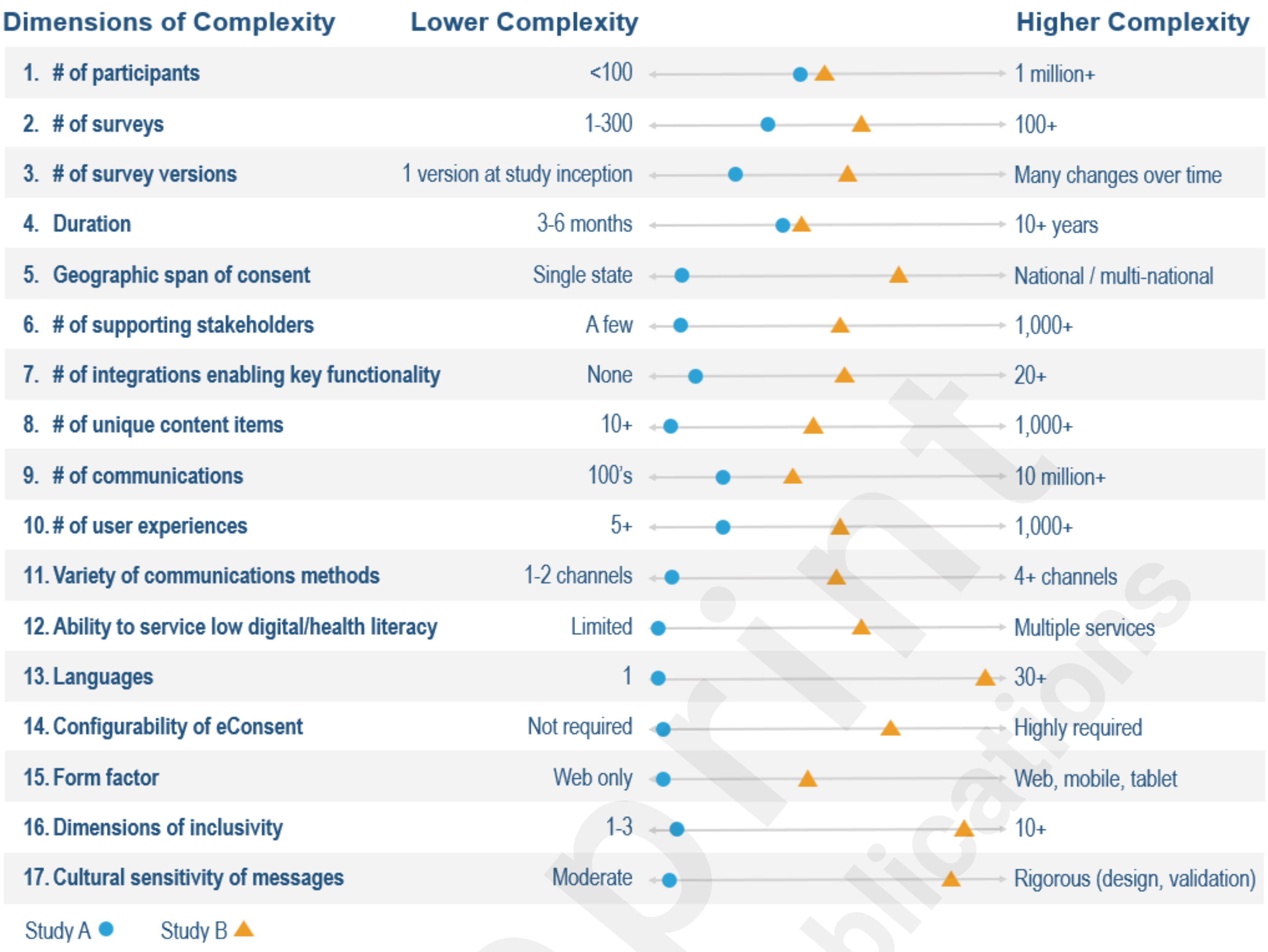


Figure 1: Illustration of how an all-in-one platform requires broad functionality to accommodate the varying dimensions of complexities for different study protocols.

Vibrent Health, a digital health technology company, was selected as the participant technology systems center and tasked with developing a secure digital platform and tools for participant experience, study management, and data analysis. [23] To align with its core values, AOU needed a secure digital health research platform that 1) could be broadly deployed nationwide to support study management, 2) could host several program touch points within a flexible participant journey, 3) was accessible to different levels of digital access, literacy, and comfort, and 4) was cyber secure and robust (Figure 2).

AOU core values		DHRP design principles
Participation is open to all. People of every race, ethnicity, sex, gender, and sexual orientation are welcome. No health insurance is required.	➡	**Participant-centric design providing consumer-grade user experiences and interfaces**

Participants reflect the rich diversity of the United States. Having a diverse group of participants can lead to important breakthroughs.	**User-friendly tools designed for all populations**
Participants are partners. Participants shape the program with their input.	**Provide tools for long-term engagement**
Transparency earns trust. We inform participants about how their data are used. Participants choose how much information to share.	**Usability for all stakeholders.** **Adaptable, scalable tools support research flexibility and efficiency**
Participants have access to their information. All of Us lets participants see their own information and records.	**Secure participant portal providing access to individual data**
Data are broadly accessible for research purposes. All of Us makes information about participants as a group available in a public database.	**Accessible data and secure data sharing**
Security and privacy are of highest importance. Data are stored in a secure, cloud-based, HIPAA compliant database.	**High level data security and privacy**
The program will be a catalyst for positive change in research. Working together, All of Us researchers, partners, and participants can build a better future for health research and care.	**Multisource data collection, data harmonization, and data management tools**

SYSTEM DESCRIPTION

Digital Health Research Platform Design and Architecture

Vibrent designed and built the Digital Health Research Platform (DHRP) to facilitate participant recruitment, enrollment, multisource data collection, and long-term engagement using web-capable digital devices (computers, mobile devices, tablets). The DHRP was developed via close collaboration between government, academia, community stakeholders, and industry using validated user experience (UX) and user interface (UI) research. (Figure 3).

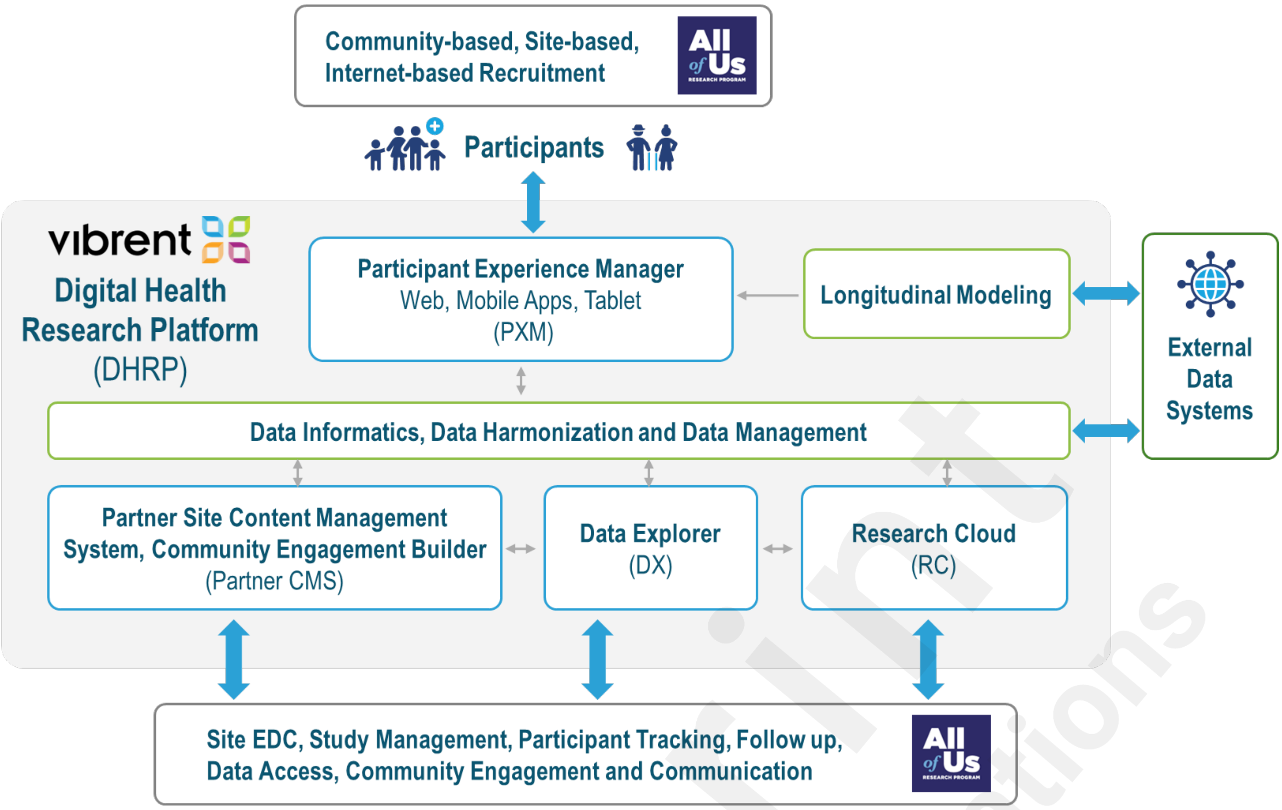


Figure 3: Integration of DHRP tools within the AOU digital infrastructure

We developed participant-facing tools for engagement and data collection, as well as researcher-facing study management tools. The data collected within DHRP was stored in a cyber-secure cloud environment where it could be harmonized, cleaned, and integrated with other data systems.

The technical architecture of DHRP is highly configurable, utilizing a low-code approach and promoting an open ecosystem without the need for custom software development. AOU utilized a range of DHRP tools to facilitate the study: 1) Participant Experience Manager (PXM), 2) Research Cloud (RC), 3) Data Explorer (DX), and 4) Community Engagement Builder (CEB). DHRP tools are shown in Figure 4.

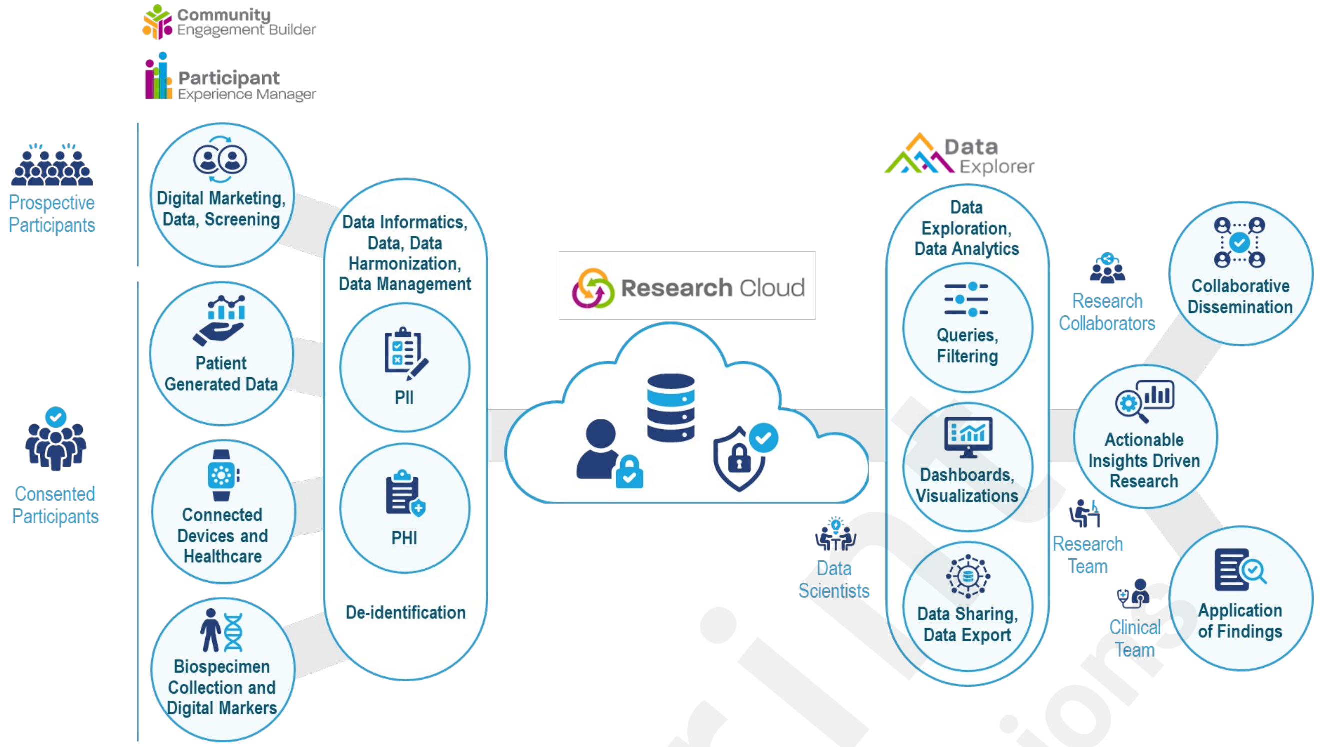


Figure 4: Schematic of DHRP end-to-end unified data informatics platform and Participant Experience Manager, Research Cloud and Data Explorer tools facilitating collaboration amongst research teams in a large multi-site consortium.

The DHRP adopted a hybrid technical strategy, blending readily accessible commercial tools and subsystems with customized application layer enhancements and integrations. This approach provided a tailored solution of high quality explicitly crafted tools that are also generalizable to meet the needs of various study designs. The platform supported an open ecosystem to integrate with additional devices and to embed experimental digital health technology modules in study protocol pathways.

DHRP utilized multi-layered enterprise architecture to achieve secure storage of data along with flexibility and scalability. The Cloud Services and Infrastructure Management Layer provided dynamic auto-scaling to meet the changing demands of studies over time. The Data Lake and Database Layer managed and stored large volumes of multisource data, including participant data and operational data. Key functions of DHRP included data integration, data transformation, advanced analytics, data governance, real-time data processing, and metadata management. The platform used centralized cloud hosted databases (MySQL, Snowflake (NoSQL), Redshift) for secure data storage, and data warehouse. The Functional Modules Layer facilitated interconnection of DHRP to multiple third-party application programming interfaces (APIs) and end-user applications.

The DHRP employs several methodological testing frameworks and capabilities as well as scientific testing support protocol including user research, functional randomization testing, randomization for scientific testing, survey response randomization, and bias mitigation. In addition, Agile software development life cycle (SDLC) and continuous integration/continuous delivery (CI/CD) processes facilitated delivery of high-quality enterprise software. The "shift-left" SDLC reduces complexity of software development, thus enabling the study to respond quickly to protocol amendments without disrupting the participant experience. Examples of configuration updates include the addition of new surveys, workflow changes, or revision of study-related content.

Containerization and microservices methods were applied in DHRP to optimize flexibility, scalability, efficiency, system performance, security, and privacy. Containerization provides a standardized, efficient approach to software deployment by encapsulating applications within isolated environments. This ensures consistent performance and compatibility across diverse computing platforms, which is essential for reproducible and scalable scientific research. [24] Microservices architecture breaks down complex software into smaller, independent parts that work together, making it easier to update and scale applications. DHRP microservices are described in Table 1.

Table 1: Microservices used in DHRP

Microservice	Function
Ancillary Study Integration	Facilitates integration with third-party study tools
Appointment Management	Schedule, view and update appointments with map and geolocation.
Asynchronous Messaging	Advanced data synchronization service enabling large scale cross

	platform integration.
Authentication Services	Enables authentication and single sign on.
Case Management	Create cases with specific actions to take for a group of participants.
Communications Services	Transmit targeted email or SMS or push communications to mobile apps.
Comparative Insights	Provide demographic-based results (by age, location, etc.) for survey responses.
Dashboards	Canned visualizations of protocol activities to assist study staff and researchers in overseeing program activities. Utilize default dashboards or create your own customized dashboard.
Data Integration	Integrate with external data hub
Data Management	Scalable storage of large volumes of multisource data
Data Retrieval Services, multiple	Retrieve data from various sources, Medicare, Epic, Cerner, Athena, Fitbit, Apple HealthKit.
Deep Linking	Communications (email and SMS) link to specific steps in the participant workflow to improve participant engagement.
eConsents and Agreements	Manages primary and additional consent agreement modules
File Sharing Services	Ability to manage upload and sharing of test results with participants.
Fulfillment Services	Order management and delivery tracking of bio sample kits and other assets required to support data collection.
Identity Verification	Secure, remote verification of participant for account access
Insights	Search, analyze and visualize data about participants and study activities. Custom dashboards, reports and data exports can be created based on the unique needs of a study.
Mobile Apps	Study branded mobile apps for participant access. Supports iOS and Android clients and adaptive web for participants choosing not to use mobile apps.
No Login Experiences	Enable participants to complete survey activities without authentication.
Reports and Exports	Create customized reports and exports for use with study engagement activities.
Segmentation	Create complex filtering criteria to identify a group of participants to target for a protocol activity, engagement and may be used to create a case or transmit communications.
Survey Import	Import survey definitions and participant survey data from 3rd party hosted survey platforms.
Task Management	Create and view appointments, cases and follow ups.
Time-based Event Management	Create user engagement workflows based on time-based protocol requirements.
User experience and advance workflow customization	Enable the creation of simplified user engagements and support for complicated time and event driven workflow. Includes configurable page and navigation items, business rules and conditions.

DHRP Participant Experience Tool

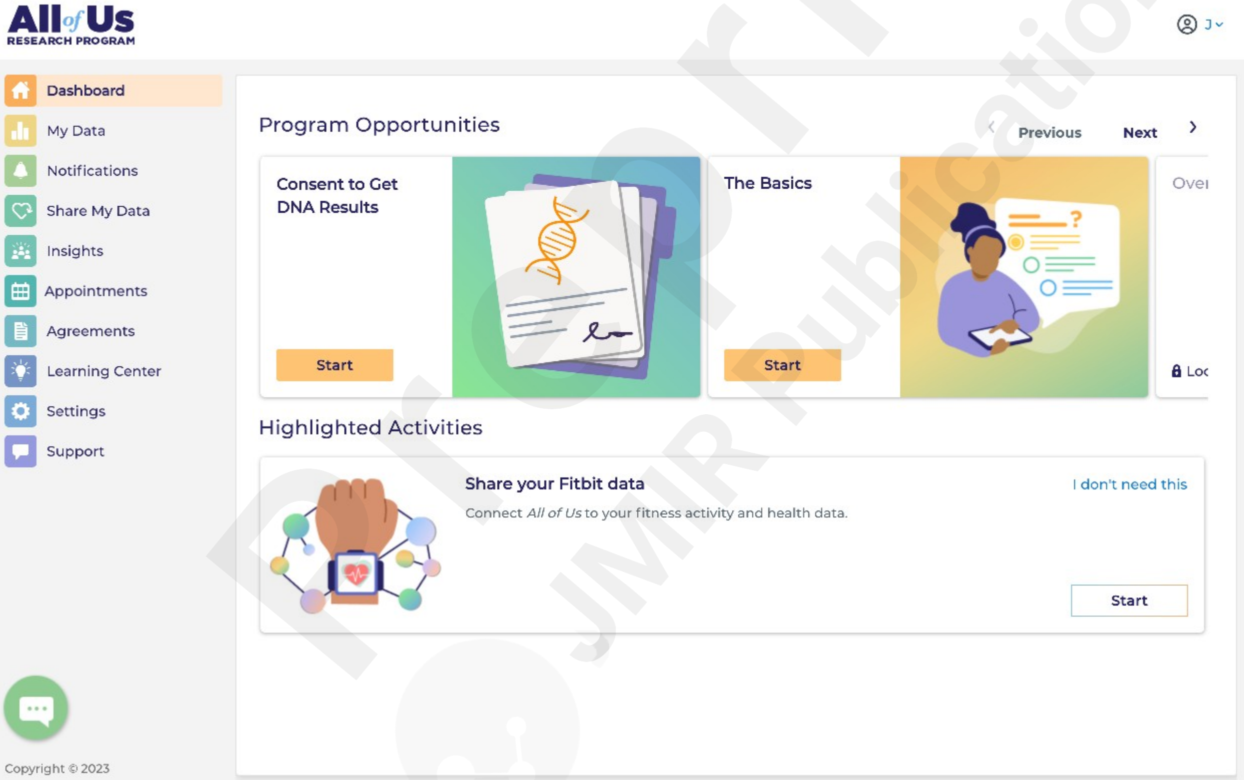
The Participant Experience Manager (PXM) facilitated all participant-facing study protocol activities, including 1) electronic informed consent (eConsent), 2) study data collection and data sharing, and 3) secure bi-directional communication. PXM operates in English and Spanish and can be translated to more than 30 additional languages. Table 2 describes the third-party services that were integrated with DHRP to support PXM.

Table 2: Integrations with third-party services in DHRP

Service	Purpose
Color	Genomics counselling
DNA Genotek, Mirimus, Molecular Testing Labs (MTL), Tasso, Quest Diagnostics	Remote bio-sample collection (saliva, blood, nasal swab)
Epic, Cerner, Athena Health, Others	Electronic health records
Fitbit, Apple HealthKit, Google Fit	Wearables and biometric devices
IDnow	Identity verification
Iterable	Email communications to participants
Lob	Postal mail to participants
Metabase, Tableau, Google Analytics	Data visualizations, dashboards, data analytics, data exploration
REDCap, Qualtrics	Surveys, codebooks
Rybbon	Participant incentives
Snowflake	Data Lake informatics
Twilio	SMS communications to participants
US Postal Service, Aftership, FedEx	Shipment and logistics
Other 3rd Party Data Lake / Database	External data repositories

Through PXM, we delivered a modular eConsent process to facilitate a self-paced, dynamic, and adaptive experience for participants. [25] The electronic informed consent (eConsent) process was implemented to increase geographic inclusivity and access to participation. The eConsent functionality, content, and user experience were developed through collaborative UX research with participants, researchers, and NIH staff. Key functionality of the eConsent included 1) web-based and mobile device accessibility, 2) built-in knowledge reinforcement, and 3) content accessible by text, audio, or video. A formative consent evaluation module was included to assess participant informedness following review of each eConsent module. Before providing digital signature, participants were required to correctly answer 80% of the evaluation questions. Participants were able to withdraw consent from any individual consent module at any time, with the option to re-consent again in the future. Digital copies of completed consent forms were stored in PXM.

PXM allowed participants to securely access individual participant portals to complete study-related data collection activities, including submission of survey responses, scheduling appointments to complete study measures, and communication with research staff. PXM supported both “repeated surveys” and “longitudinal data collection” with defined events including complex window management that made certain study tasks available in the participant dashboard during the appropriate window of time. Participants were able to securely review their personal data and genetic results through PXM. PXM was also integrated with a range of third-party fast healthcare interoperability resources (FHIR) systems enabling participants to share electronic health records and medical claims data to contribute to the study. This secure, participant-focused design yielded user-friendly, tailored experiences to foster ease-of-use throughout the study (Figure 5).



Connect *All of Us* to your fitness activity and health data.

Figure 5:Screenshot of the Participant Experience Management (PXM) dashboard

As part of an end-to-end readiness assessment for Pediatric enrollment, PXM supports engaging with parents of eligible children as proxy users. Parents of children under the age of 5 years old were invited to enroll their children in the study and were authorized to complete permission, authorization, and data collection modules on behalf of their child(ren). Tools enable study staff to engage with children for bio sample and physical measurement collection processes using PXM. Parents can enroll multiple children and select these children in a drop-down menu.

DHRP Research Administration Tools

The Research Cloud (RC) empowers research staff to complete direct engagement, monitoring, and support to assist participants with completion of study activities. RC enables research staff to communicate with participants via automated digital marketing, email, and SMS messaging, and to securely collect participant data using built-in Computer Assisted Telephone Interviewing (CATI) capabilities. RC allowed staff to assist participants with account access and scheduling of study measures visits (in-clinic or remote). This tool provided tracking of participant progress via customizable dashboards. RC was critical in building relationships to keep participants engaged and ensure that study tasks were completed.

For security and privacy, RC implemented a Roles-Functions-Permissions based framework system which granted limited access to select data and workflows based upon the research staff’s assigned role within the platform. The thirteen (13) distinct user roles available in RC included program coordinators, data analysts, research assistants, and system administrators. After completing two-step authentication login, RC users could review a role-based data dashboard and/or download participant-level research data. To facilitate adoption, a training environment was created which allowed RC users to practice their workflows without compromising real study data. The Research Cloud can also support expansion to ancillary studies and can extend user roles to allow access to multiple projects through one dashboard.

DHRP Data Management tool

Data Explorer (DX) is a self-service analytics tool that enabled study staff to access structured or unstructured data to construct custom reports, dashboards, and insights of collected data and metadata. DX allows researchers to review and analyze study data from all sources (surveys, collected biometric measures, and third-party shared data) in real time within one dashboard. DX also supports the use of data analytics tools (RStudio, Python, JupyterHub) within the platform on live data. This capability allows researchers to assess data quality and preliminary study results in real-time. Modular data pipelines were developed in Python, allowing for scalability and adaptability to many different data sources and data analytics functions.

DHRP Content Management tool

The content management tool, called Community Engagement Builder (CEB), delivered study outreach materials, and hosted the study recruitment website. It enabled rapid and low-cost creation and deployment of customized recruitment websites, event marketing, and landing pages tailored to a site's specific catchment area. CEB included UX-validated templates to optimize content with appropriate levels of cultural sensitivity and language translation. CEB was used to generate websites that educate and recruit people to AoU or ancillary studies and was also used to build the nationwide recruitment website.[26]

CEB integrated with Research Cloud to provide real-time enrollment metrics to support strategic adjustment of recruitment strategies for study success. Digital marketing analytics generated by CEB enabled staff to track conversions from both online (recruitment website, lead generation, events landing pages, and social networking links) and offline sources (QR codes from posters and flyers). By leveraging these analytics, staff gained insights into the effectiveness of different recruitment methods and made informed decisions to optimize recruitment efforts and meet enrollment and diversity targets.

DHRP Research staff training and support tools

After the study launched, there was a recognized need to provide comprehensive and efficient support and training to research staff using DHRP tools. The Help Center and the Vibrent Research Academy were digital training resources created to support researchers' executing study management and outreach tasks. The Help Center served as a knowledge base for site staff using the RC or CEB tools and provides detailed explanation of DHRP platform features, along with supplemental articles, troubleshooting tips, and other resources. The Vibrent Research Academy offered short courses which provide detailed, multimedia demonstration and instruction on the use of DHRP tools and features.

Data security, encryption, regulatory, and compliance

The DHRP maintains "defense in depth" security posture to keep participant data safe and private. Data security and privacy is implemented in accordance with established industry-standard policies, procedures, and technology system guidelines such as FISMA and NIST Cybersecurity Framework for Security and Privacy Controls (NIST SP 800-53 r5, NIST SP 800-39, NIST SP 800-37) and was HIPAA compliant. Technical safeguards include strong data encryption (AES256, TLS 1.2 and higher) in flight and at rest. The DHRP is compliant with 21 CFR part 11 and ensures entries come from an authenticated and authorized source. Role-based permissions ensure that only authorized users are permitted to change settings.

MATERIALS AND METHODS

Community engagement

AOU is a population-based precision medicine initiative operating on a community engagement structure. To operationalize the program’s core values, a community based participatory research approach was implemented to develop a national network of actively engaged community organizations and individuals to build Community Advisory Boards (CAB). CABs collaborated with HPOs, local community centers, academic and industry partners to provide study guidance.

Participant recruitment and enrollment

Participant recruitment, eligibility, and enrollment were delivered through in-person and digital methods facilitated by the DHRP. Outreach was conducted at local community events, conferences, word-of-mouth, digital advertising, and physical distribution of study-branded flyers and promotional items. The CEB tool of DHRP provided customized, culturally appropriate recruitment webpages in multiple languages. Participants who enrolled remotely were able to communicate with staff by phone or email via PXM. Remote participants were also able to use PXM to locate the nearest site at which to complete their physical study measurements. Participants were also recruited in-person at HPO and mobile clinic sites throughout the US with enrollment completed in the DHRP, with or without assistance from site staff.

Study data collection

All study data was collected in the DHRP. Participants used the PXM tool to submit surveys and complete data sharing from third-party services. To minimize participant burden, all surveys were available in separate, self-paced modules allowing participants to save their progress and return to modules for completion later. To support data collection, research staff used the DX tool to identify participants with incomplete study tasks, such as eConsent or surveys. RC was then used to develop targeted engagement campaigns to remind participants of remaining incomplete study tasks requiring attention. Description of sources and types of data collected in DHRP is available in Table 3.

Table 3: Multisource data types collected in DHRP.

Data Source	**Data formats**
e-Consent signature	PDF
Electronic health records	JSON formatted FHIR payloads
Consumer wearables (Fitbit, Apple HealthKit)	Apple HealthKit health data - JSON Apple HealthKit EHR data - JSON
Biometric measures	JSON formatted FHIR payloads
Genetic results	PDF formatted reports
Surveys	JSON formatted FHIR payloads - Questionnaire Response
Personally identifiable information	JSON formatted FHIR payloads
Participant communications	CSV, JSON
Participant communication	CSV
Prospective participants	CSV
Cognitive assessments	JSON formatted FHIR payloads
Pediatric assent and surveys	PDF, JSON formatted FHIR payloads - Questionnaire Response
Digital marketing data	CSV, UTM tracking parameters

RESULTS

The DHRP launched for use in May 2018. As of April 2024, a total of 705,719 participants had enrolled in AOU through the DHRP. Of these participants, 74% (n=524,264) completed all study activities in the core study protocol, including all required survey modules. In the current cohort, 55% identified as women (n=386,377), 51% were between ages 35-64 years (n=360,303), and 48% identified as White (n=340,423). The platform was used by participants and research teams in each of the 50 US states and territories with 8% (n=58,211) representing rural areas (Figure 6).

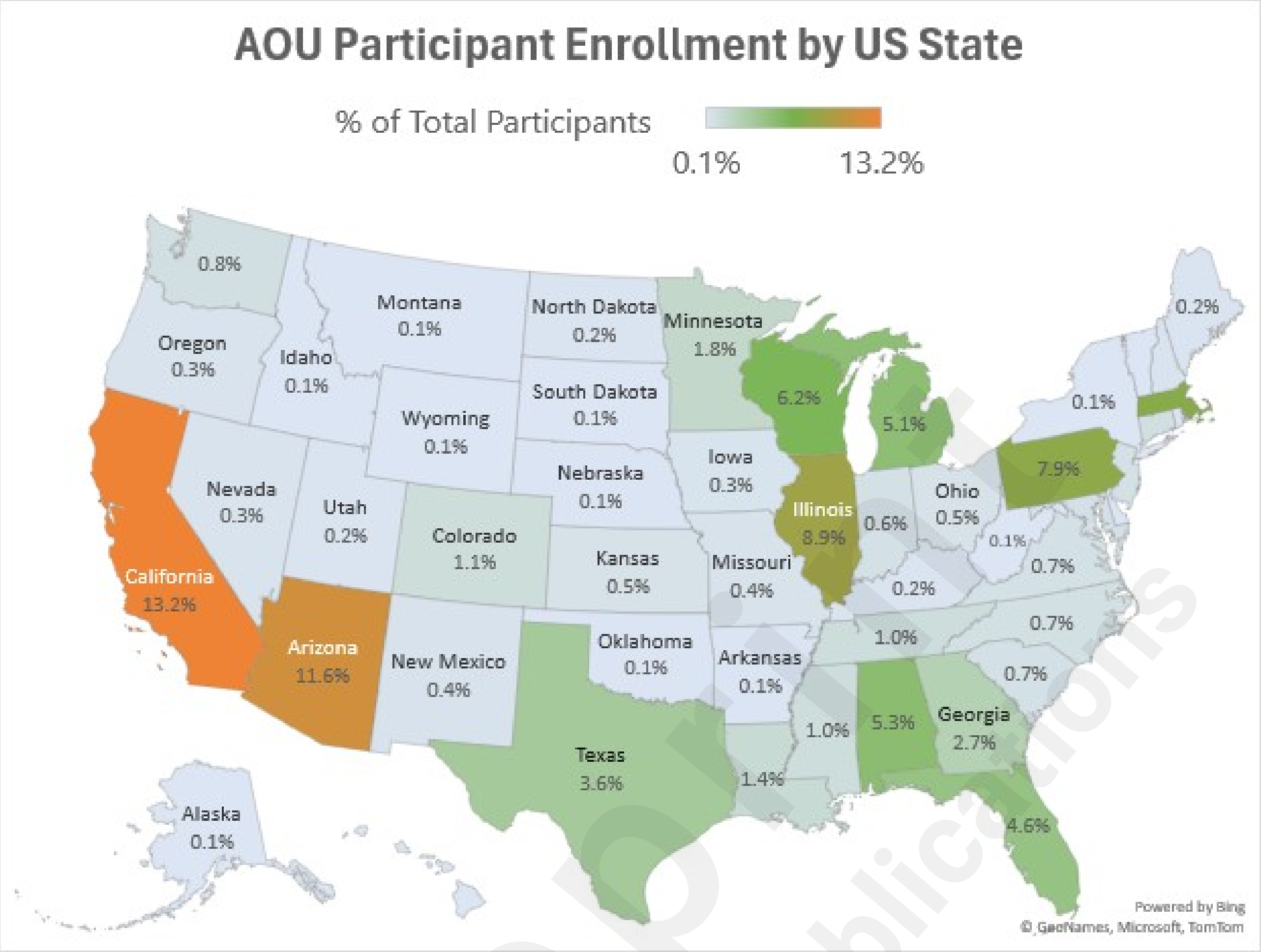


Figure 6: Geographic distribution of AOU participants enrolled through DHRP by US State as of March 21, 2024. State enrollment varied from 13.2% (California) to 0.1%. The West US region had the highest enrollment (28.1%).

AOU has utilized DHRP to build a diverse research cohort in which 87%(n=457,514) identify as members of groups underrepresented in biomedical research in at least one demographic category (age, race, sex, gender, income, education, or disability). More specifically, 46% of participants (n=243,296) identified as racial minorities, 36% (n=189,087) report low income, and 20% (n=104,916) report a disability. One third of the cohort (n=219,355) are aged 65+ years, 8% (n=58,211) identify as rural residents, and 28% (n=197,601) are persons with limited access to healthcare. More than 5% of all program participants have opted to participate in the program in Spanish (n=~37,000). UBR diversity metrics of the current cohort are available in Table 4.

Table 4: AOU participant sociodemographic characteristics as of April 2024.

Total, enrolled through DHRP	n=705,719
Characteristic	n (%)
Age, years	
18-24	25,707 (3.6%)
25-44	213,958 (30.3%)
45-64	246,699 (35%)
65-84	205,072 (29%)
85+	14,283(2%)
Gender Identity	
Man	237, 070 (33.6%)
Woman	386,377 (54.7%)
Non-binary	2,729 (0.4%)
Transgender	1,169 (0.2%)
Other/Multiple selections	4,010 (0.5%)
Unspecified/No answer	74,364 (10.5%)
Racial Identity	
White	340,423 (48.2%)
Black or African American	102,371 (14.5%)
Asian	22,281 (3.2%)
Multi-race/ Other	55,872 (7.9%)
None/no answer	86,477 (12.3%)
Education	
Less than high school	51, 236 (7.3%)
Grade 12 or GED	113,545 (16.1%)
College, 1-3 years	165,317 (23.4%)
College, 4 years or more	148,910 (21.1%)
Advanced degree	42,504 (20.2%)
Unspecified/ no answer	84,207 (11.9%)
Income, annual household	
≤ $24,999	143,939 (20.1%)
$25,000-$49,999	96,800 (16%)
$50,000-$74,999	70,599 (10%)
$75,000-$99,999	55,0752 (7.8%)
$100,000-$199,000	101,965 (14.4%)
$200,000 or more	45,136 (6.4%)
Unspecified/ no answer	188,205 (26.7%)
Rurality	
Rural	58,211 (8.2%)
Non-rural (Metro/Suburban)	644,852 (91.4%)
Unspecified/ no answer	2,656 (0.4%)

Approximately 4,500,000 combined survey and eConsent modules have been completed through DHRP. Analysis of eConsent formative evaluation questions revealed that after reviewing eConsent modules, over 95% of participants were able to differentiate the research program from medical care, understand the voluntary nature of their involvement, and comprehend their right to withdraw at any point in time.[27] Complete data collection metrics can be reviewed in Table 5.

Table 5: Data collected and tool usage metrics from the DHRP for AOU

Measure	**Data (**n)
Participant Data sources	
Fitbit	49,627
Apple HealthKit	19,940
Electronic Health records (EHR)	14,431
Survey modules completed, total	3,685,608
Physical Biometric study measures, total	538,160
Saliva kits	59,819
Electronic Informed Consent (eConsent)	
Primary eConsent modules completed, total	2,015,622
EHR eConsent module	632,644
Genomics return of results eConsent module	406,781
Cognitive assessment eConsent module	85,572
Digital Marketing	
Prospective participants contacted	378,010
Automated email and SMS campaigns delivered	3,677
Custom HTML templates created by sites in the Partner CEB tool	1,838
Multi-channel communication automations (bi-lingual)	300+
Participant Multimodal Communications	
emails sent to participants	64,837,083
SMS messages sent to participants	7,794,054
Mobile app push notification messages sent to participants	463,975
Direct mail sent to participants	1,531,814
Site staff remote engagement with participants	1,168,079
Protocol step completions logged by site staff due to remote engagements	412,026
CATI sessions	42,795
Appointments scheduled (remote and site-based)	301,393

Another demonstration of the effective engagement capabilities of the DHRP is shown in the completion rates of the optional COVID-19 Pandemic Evaluation (COPE) survey. The COPE survey was added in response to the COVID-19 pandemic which was first confirmed in the United States in January 2020. The COPE survey launched in DHRP in May 2020 and six (6) iterations were developed, modified, and released to participants through February 2021. Previous study of the

COPE survey showed that the modification of surveys led to a significant increase ($p < 0.001$) in the participant response rate over 10 months of survey delivery. [28] Overall, completion rates for the COPE survey among individuals enrolled under UBR categories showed an increase from 11% in the first survey to 16% in the final survey. These findings underscore the significance of utilizing a digital platform able to accommodate rapid implementation of protocol amendments.

DISCUSSION

Principal Results

The design and implementation of the DHRP in this study demonstrates that a digital health research platform with technical architecture for participant-focused design can effectively be used to recruit, collect health data from geographically, racially, and ethnically and otherwise diverse participants for longitudinal cohort research. The collaborative, community-informed approach used to develop the DHRP combined with validation through user experience research contributed to its effective use. The participant-centric design of DHRP may be broadly applicable to increase participation of persons from diverse backgrounds in other research study designs, including cross-sectional, case-control, and experimental clinical trials.

The principles of generalizability, usability, cultural sensitivity, and accessibility were consistently maintained in the forefront of DHRP design. This led to the creation of digital tools that were both participant-centered and researcher-friendly. Furthermore, integrating PXM with RC and CEB in a containerized manner protected participant autonomy and privacy while also maintaining workflow automation and bidirectional communication between participants and study staff. RC also shows how analytics and dashboards can be used to increase participant retention and protocol completion to enhance data quality.

From an engagement perspective, it was important to establish a standardized electronic consent process that was accessible and comprehensive for all individual platform users, regardless of their geographical location, method of enrollment, preferred language, reading level or digital device used. PXM eConsent modules were successfully used with high levels of participant informedness demonstrated by high rates of completion of the formative consent evaluation. Well-designed and delivered eConsent provides transparency to participants at study initiation and can increase volunteerism, participant understanding of research, and participant satisfaction. This can lead to increased trust and future participation in research amongst groups historically underrepresented in the field. [25, 27, 29]

This demonstrated use of DHRP supports the idea this digital research platform can enhance delivery of longitudinal studies and can be used by participants with different levels of digital literacy. It is important to note that a combined approach of both digital and in-person engagement may be best for building trust with UBR communities. While the DHRP did facilitate and streamline participant-to-site staff engagement in AOU, the DHRP did not replace the need for research staff. For instance, direct messaging through email and SMS was used to engage participants, however research staff were needed to accommodate participants with low digital literacy using in-person and person-to-person communications such as phone calls or CATI.[30] Further study is needed to separately quantify the contributory effects of digital enablement with or without the support of research staff.

Conclusions

We created and validated a digital research platform to support the development of a large, nationwide, community-engaged, longitudinal cohort study (AOU) of diverse participants from UBR groups. This study demonstrates the promising potential of digital platforms to enable the curation, harmonization, and data management of high-volume longitudinal multisource datasets from diverse populations for precision medicine research. The technology-enabled research approach is generalizable to both small and large research studies, observational studies, and clinical trials. Digital platforms can provide solutions to researchers who aim to broaden their catchment area in site-based, decentralized and hybrid research methods to develop more diverse datasets inclusive of underrepresented groups, supporting a fundamental shift towards research that generates more generalizable scientific outcomes and addresses health disparities. With continued development, digital research platforms, such as the DHRP described here, may lead to successful establishment of best approaches to engage health disparate, minority, vulnerable, and other underrepresented populations into clinical research to broadly improve the overall health of all communities.

Acknowledgements

We would like to thank all collaborators who contributed to the design and implementation of Vibrent's digital health research platform. We would like to acknowledge all participants of the All of Us Research Program.

Conflict of Interest

M. Begale, A. Montgomery, S. Shokouhi, D. Klein, S. Sutherland, D. Joshi, A. Ashbeck, M. Palmer, and P. Jain are employees of Vibrent Health, Inc. M. Begale, S. Sawyer, J. McCauley, D. Klein, and P. Jain are affiliated with the All of Us Research Program.

Author Contributions

Manuscript concept and design were completed by PJ, DK, MB, AM, SSh in collaboration with AOU stakeholders. Data acquisition was completed by SSc and JM. Data analysis and interpretation was completed by MB, AM, and SSh. The manuscript was drafted by AM, DK, MB, SSh, PJ, and MV. Manuscript revision was led by AM, DK, PJ, and MB and completed by all co-authors. Theoretical support was provided by DK, SSa, JM, DK, and PJ.

Funding

OTA-20-001

RFA-PM-16-003

REFERENCES

1. Caruana, EJ, Roman, M, Hernández-Sánchez, J, & Solli, P. Longitudinal studies. Journal of thoracic disease. 2015; 7(11): E537–E540. doi: 10.3978/j.issn.2072-1439.2015.10.63

2. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016; 538(7624): 161-164. doi: 10.1038/538161a
3. Allison K, Patel D, Kaur R. Assessing Multiple Factors Affecting Minority Participation in Clinical Trials: Development of the Clinical Trials Participation Barriers Survey. Cureus. 2022; 14(4), e24424. doi: 10.7759/cureus.24424
4. Hamel L M, Penner LA, Albrecht TL, et. al. Barriers to Clinical Trial Enrollment in Racial and Ethnic Minority Patients With Cancer. Cancer control: Journal of the Moffitt Cancer Center. 2016; 23(4): 327–337. doi: 10.1177/107327481602300404
5. Ejiogu N, Norbeck JH, Mason MA, et. al. Recruitment and retention strategies for minority or poor clinical research participants: lessons from the Healthy Aging in Neighborhoods of Diversity across the Life Span study. The Gerontologist. 2011; 51 Suppl 1(Suppl 1): S33–S45. doi: 10.1093/geront/gnr027
6. Allmark P. Should research samples reflect the diversity of the population? Journal of medical ethics. 2004; 30(2): 185–189. doi: 10.1136/jme.2003.004374
7. Gross AS, Harry AC, Clifton CS, Della Pasqua O. Clinical trial diversity: An opportunity for improved insight into the determinants of variability in drug response. Br J Clin Pharmacol. 2022; 88(6): 2700-2717. doi:10.1111/bcp.15242
8. Clark L, Watkins I, Piña L, et. al. Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. Current Problems in Cardiology. 2009; 44(5): 148-172. doi: 10.1016/j.cpcardiol.2018.11.002.
9. McConnell MV, Shcherbina A, Pavlovic A, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart Counts Cardiovascular Health Study. JAMA Cardiol; 2017;2:67-76
10. Spartano NL, Lin H, Sun F, Lunetta KL, et. al. Comparison of On-Site Versus Remote Mobile Device Support in the Framingham Heart Study Using the Health eHeart Study for Digital Follow-up: Randomized Pilot Study Set Within an Observational Study Design, JMIR Mhealth Uhealth. 2019; 7(9):e13238. doi: 10.2196/13238
11. Hwang DA, Lee A, Song JM, Han H. Recruitment and Retention Strategies Among Racial and Ethnic Minorities in Web-Based Intervention Trials: Retrospective Qualitative Analysis, J Med Internet Res. 2021; 23(7):e23959 doi: 10.2196/23959
12. Dockendorf MF, Hansen BJ, Bateman KP, et. al. Digitally Enabled, Patient-Centric Clinical Trials: Shifting the Drug Development Paradigm. Clin Transl Sci. 2020;14: 445-459. doi: 10.1111/cts.12910
13. Greenhalgh T, Wherton J, Papoutsi C, et. al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res. 2017;19(11):e367. doi: 10.2196/jmir.8775
14. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of Mobile Health in Communities Underrepresented in Biomedical Research: Barriers and Ethical Considerations for Scientists, JMIR Mhealth Uhealth. 2017; 5(6):e87 doi: 10.2196/mhealth.6494
15. All of Us Research Program Investigators, et. al. The "All of Us" Research Program. The New England journal of medicine. 2019; 381(7), 668–676. doi: 10.1056/NEJMsr1809937
16. Armenta A, Serrano A, Cabrera M, Conte R. The new digital divide: the confluence of broadband penetration, sustainable development, technology adoption and community participation. Information Technology for Development. 2012; 18(4), 345–353. doi:

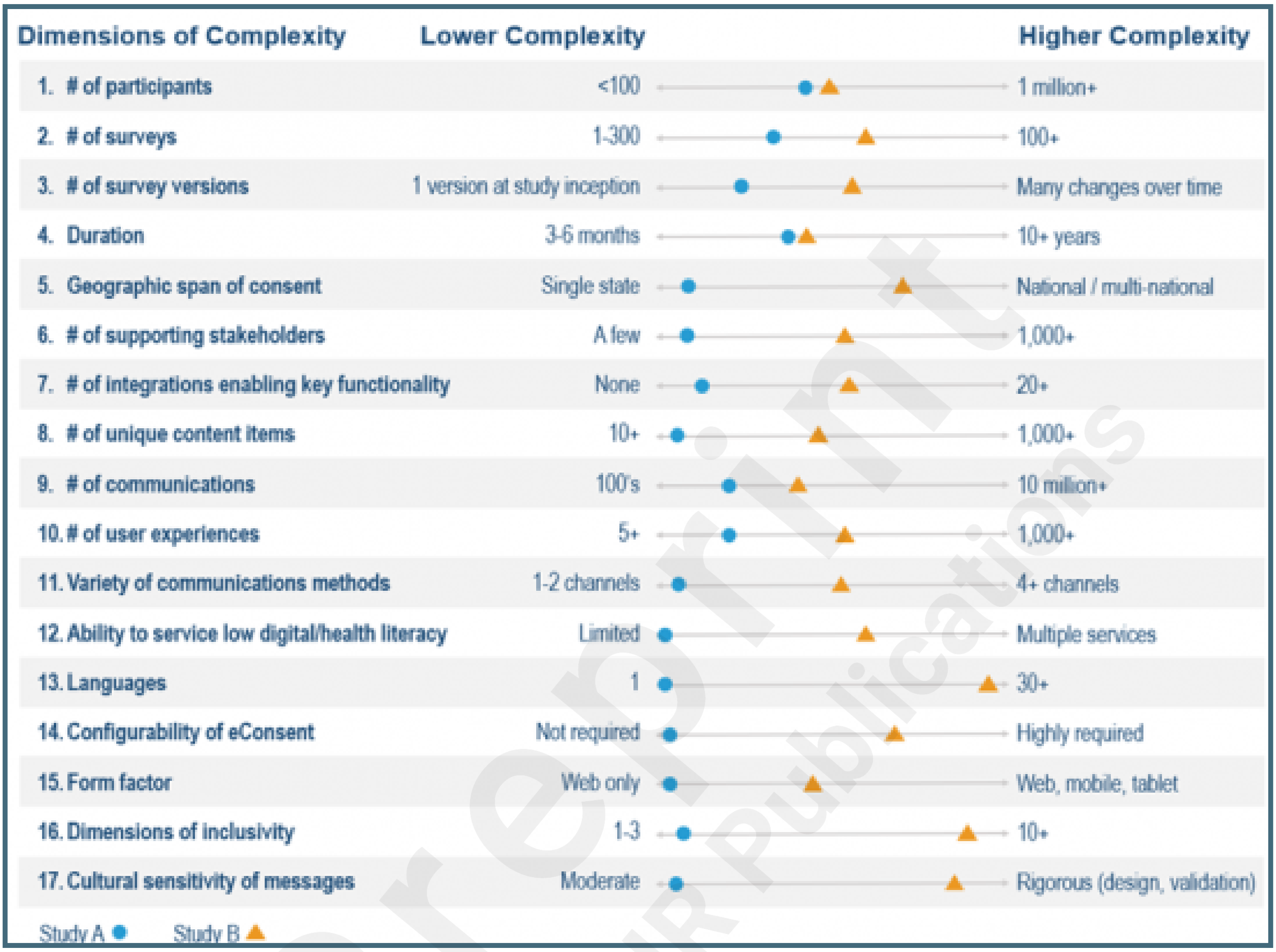
10.1080/02681102.2011.625925

17. Anderson-Lewis C, Darville G, Mercado R, Howell S, Di Maggio S. (2018). mHealth Technology Use and Implications in Historically Underserved and Minority Populations in the United States: Systematic Literature Review. JMIR Mhealth Uhealth, 6(6): e128. DOI: 10.2196/mhealth.8383
18. Mapes BM, Foster CS, Kusnoor SV, Epelbaum MI, AuYoung M, et al. Diversity and inclusion for the All of Us research program: A scoping review. PLOS ONE. 2020; 15(7): e0234962. doi: 10.1371/journal.pone.0234962
19. Ghadessi M, Di J, Wang C, et al. Decentralized clinical trials and rare diseases: a Drug Information Association Innovative Design Scientific Working Group (DIA-IDSWG) perspective. Orphanet J Rare Dis. 2023; 18, 79. doi: 10.1186/s13023-023-02693-7
20. Nijhawan LP, Janodia MD, Muddukrishna BS, Bhat KM, Bairy KL, Udupa N, Musmade PB. Informed consent: Issues and challenges. J Adv Pharm Technol Res. 2013; (3):134-40. doi: 10.4103/2231-4040.116779.
21. National Institutes of Health All of Us Research Program. (2024, April 21). All of Us Research Program Overview. https://allofus.nih.gov/about/program-overview
22. Doerr M, Grayson S, Moore S, Suver C, Wilbankson J, Wagner J. Implementing a universal informed consent process for the All of Us Research Program. Biocomputing 2019. 2018; 427-438. doi: 10.1142/9789813279827_0039
23. National Institutes of Health All of Us Research Program. (2024, April 21). Participant Technology Systems Center. https://allofus.nih.gov/funding-and-program-partners/participant-technology-systems-center
24. Casalicchio E, Iannucci S. The state-of-the-art in container technologies: Application, orchestration and security. Concurrency Computat Pract Exper. 2020; 32:e5668. doi: 10.1002/cpe.5668
25. Doerr M, Moore S, Barone V, Sutherland S, Bot BM, Suver C, Wilbankson J. Assessment of the All of Us research program's informed consent process. AJOB Empirical Bioethics, 2021; 12:2, 72-83, doi: 10.1080/23294515.2020.1847214
26. National Institutes of Health. (20204, May 1). All of Us Research Program. https://www.joinallofus.org/
27. Earl CE, Penney PJ. The Significance of Trust in the Research Consent Process with African Americans. Western Journal of Nursing Research. 2001;23(7):753-762. doi:10.1177/01939450122045528
28. Schulkey CR, Litwin TR, Ellsworth G, et. al. Design and Implementation of the All of Us Research Program COVID-19 Participant Experience (COPE) Survey, American Journal of Epidemiology.2023; 192(6), 972–986. doi: 10.1093/aje/kwad035
29. Guarino, Jennifer, et al. "Characteristics of electronic informed consent platforms for consenting patients to research studies: a scoping review." MEDINFO 2021: One World, One Health–Global Partnership for Digital Innovation; 2022) 777-781.
30. Peterson, Rachele, et al. Rethinking Data Collection Methods During the Pandemic: Development and Implementation of CATI for the All of Us Research Program. Journal of Public Health Management and Practice. 2024; 30(2), 195-199. doi: 10.1097/phh.0000000000001846

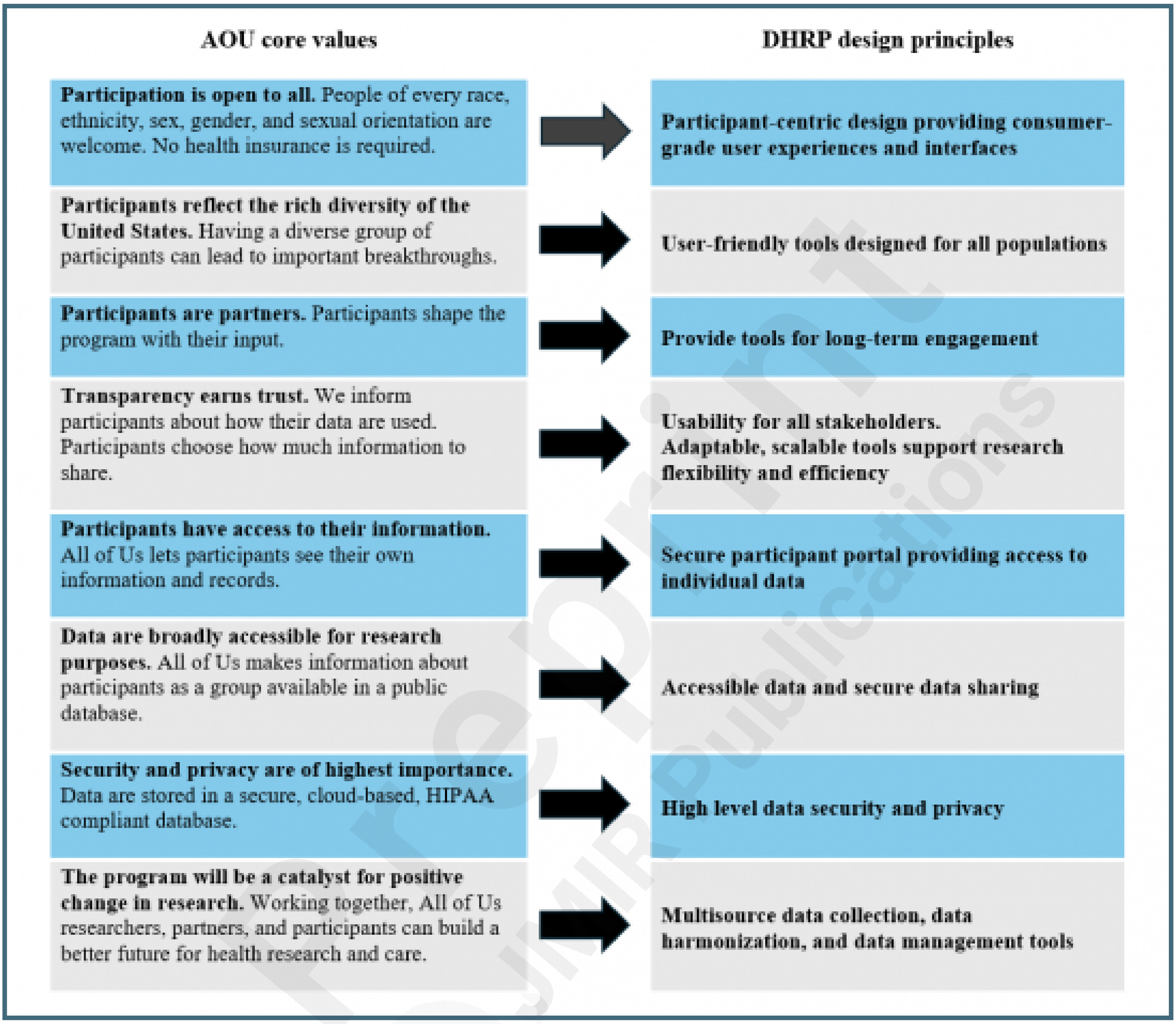
Supplementary Files

Figures

Illustration of how an all-in-one platform requires broad functionality to accommodate the varying dimensions of complexities for different study protocols.

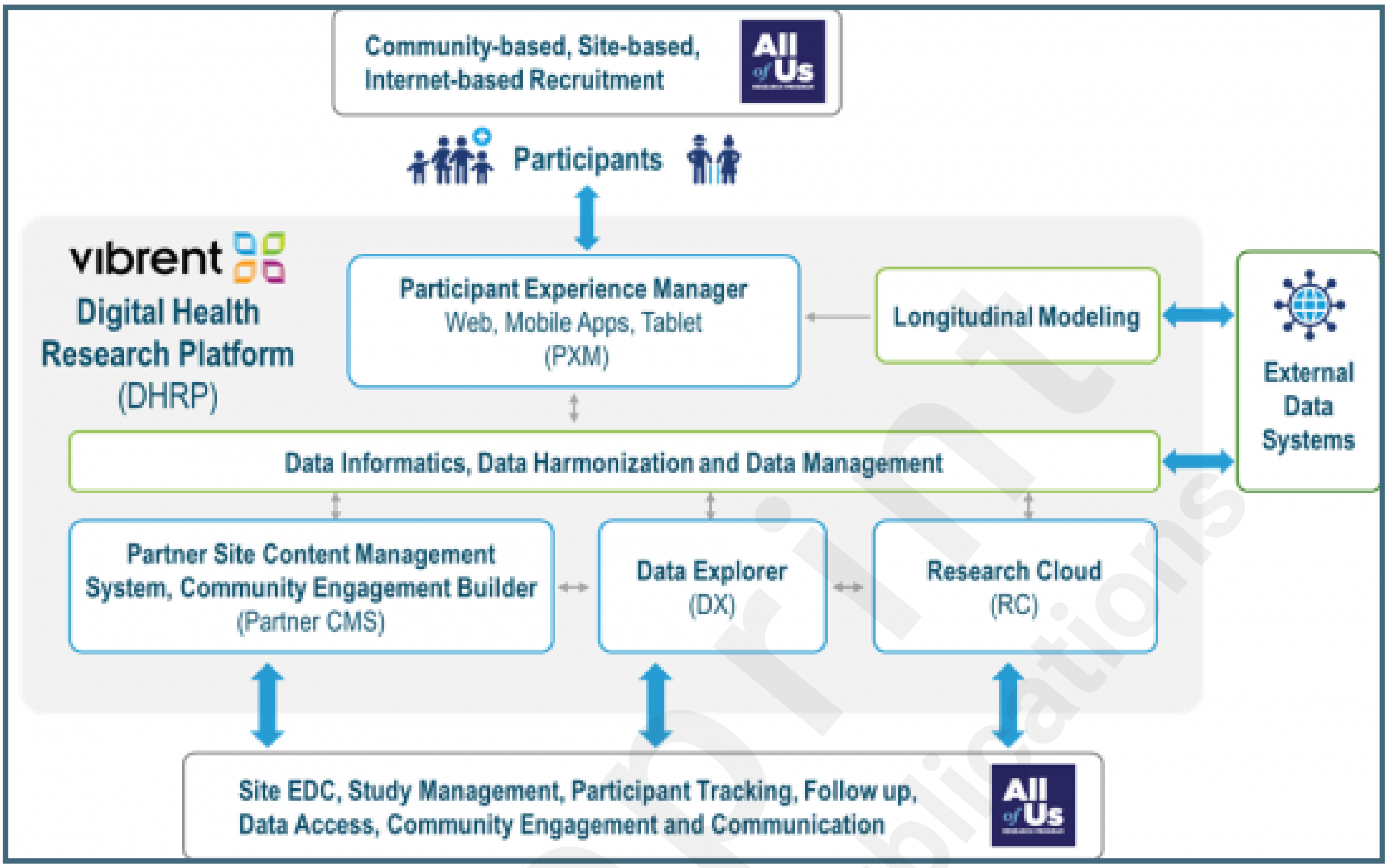


All of Us research program core values were used to develop key design features of the digital health research platform.

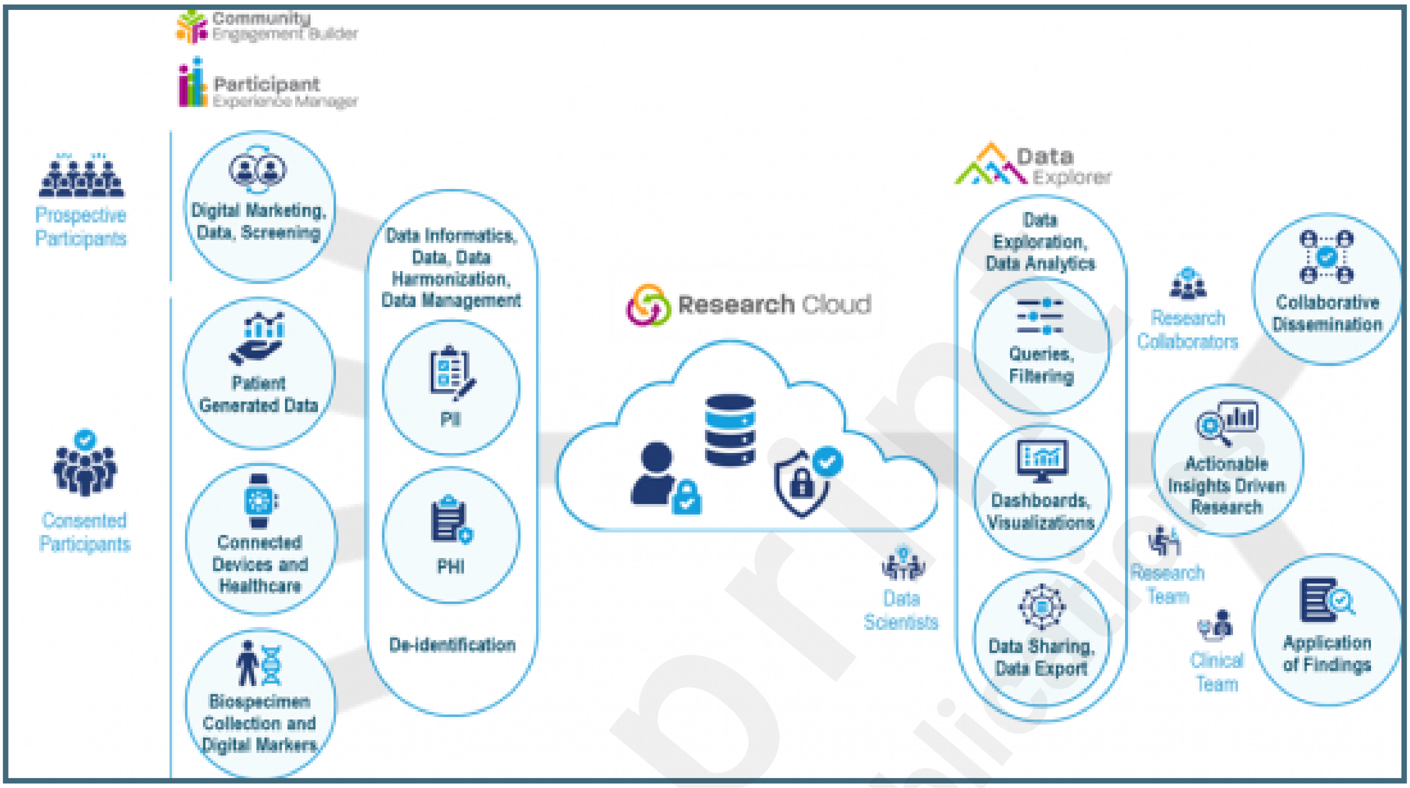


AOU core values	DHRP design principles
Participation is open to all. People of every race, ethnicity, sex, gender, and sexual orientation are welcome. No health insurance is required.	**Participant-centric design providing consumer-grade user experiences and interfaces**
Participants reflect the rich diversity of the United States. Having a diverse group of participants can lead to important breakthroughs.	**User-friendly tools designed for all populations**
Participants are partners. Participants shape the program with their input.	**Provide tools for long-term engagement**
Transparency earns trust. We inform participants about how their data are used. Participants choose how much information to share.	**Usability for all stakeholders. Adaptable, scalable tools support research flexibility and efficiency**
Participants have access to their information. All of Us lets participants see their own information and records.	**Secure participant portal providing access to individual data**
Data are broadly accessible for research purposes. All of Us makes information about participants as a group available in a public database.	**Accessible data and secure data sharing**
Security and privacy are of highest importance. Data are stored in a secure, cloud-based, HIPAA compliant database.	**High level data security and privacy**
The program will be a catalyst for positive change in research. Working together, All of Us researchers, partners, and participants can build a better future for health research and care.	**Multisource data collection, data harmonization, and data management tools**

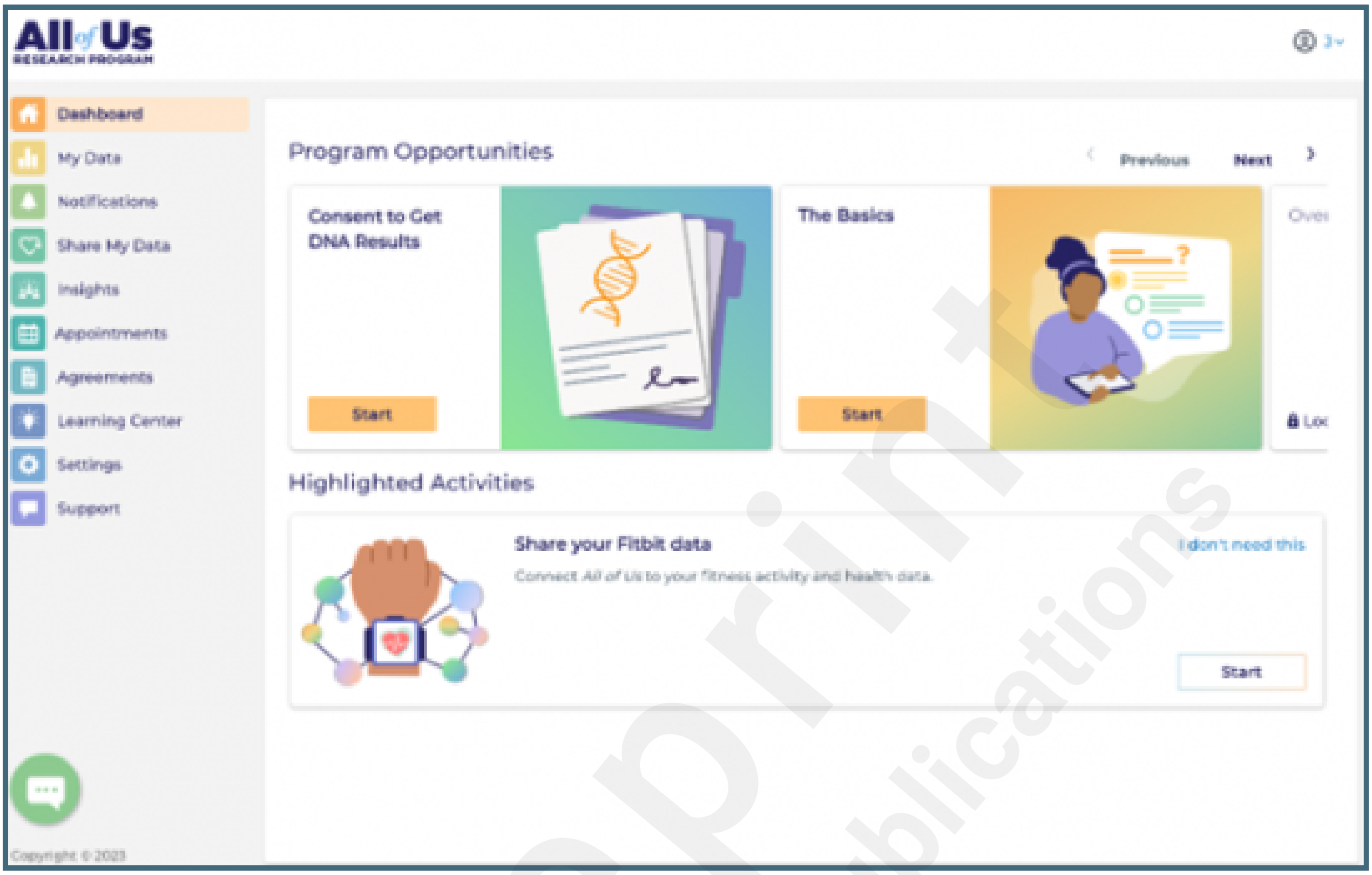
Integration of DHRP tools within the AOU digital infrastructure.



Schematic of DHRP end-to-end unified data informatics platform and Participant Experience Manager, Research Cloud and Data Explorer tools facilitating collaboration amongst research teams in a large multi-site consortium.



Screenshot of the Participant Experience Management (PXM) dashboard.



Connect *All of Us* to your fitness activity and health data.

Geographic distribution of AOU participants enrolled through DHRP by US State as of March 21, 2024. State enrollment varied from 13.2% (California) to 0.1%. The West US region had the highest enrollment (28.1%).