All of Us Research Program
Participant Provided Information
Data Management & Sharing Plan (DMSP), CDR v9

Summary Information	
Data type	Participant Provided Information (PPI)
Data source	Beginning at program establishment in May 2017, all participants – whether enrolled through their participating Healthcare Provider Organizations (HPOs) or not (direct volunteers) – responded to surveys via the *All of Us* Participant Portal supported by the program's Participant Technology Systems Center (Vibrent Health). Beginning in December 2019, a new version of the Participant Portal, supported by Scripps University with CareEvolution (CE), was introduced as a pilot approach to enrolling a limited number of direct volunteer participants.
Dates reflected in data type	All surveys collected up to v9 CDR cut-off date, January 1, 2025
Cadence for updates	PPI was delivered in near real time from PTSC and CE to the Data and Research Center (DRC).

High-level summaries of the data type (materials/links)	• Introduction to *All of Us* Survey Collection and Data Transformation Methods
	• Estimated number of participants in the CDR v9 survey data:
	  o Basics: 751,242
	  o Overall Health: 684,082
	  o Lifestyle: 678,028
	  o Health Care Access and Utilization (HCAU): 370,956
	  o Family Health History (FHH): 151,161
	  o Personal Medical History (PMH): 147,449
	  o Personal and Family Health History (PFHH): 204,636
	  o Social Determinants of Health (SDOH): 318,335
	  o Emotional Health History and Well-Being (EHHWB): 137,946
	  o Behavioral Health and Personality (BHP): 139,637
	• Surveys – User Support

Data quality (materials/links)
• **Summary**
• **Issue Tracker**
• **Change Logs**
	• Summary: *All of Us* Survey Changes Over Time
	• Issue tracking:
	  o Data quality (DQ) feedback is shared back with Data Submitters after review
	  o Inquiries from researchers are tracked in this flow chart.
	  o DRC SOP for Incremental Data Amendment
	  o Issues with the data are listed in the CDR release notes
	• Change logs: Any updates to the data are listed in the CDR release notes

***All of Us* Contact**	support@researchallofus.org

All of Us Research Program
Participant Provided Information
Data Management & Sharing Plan (DMSP), CDR v9

SUMMARY OF REVISIONS

Version	Date	Section	Revisions Made	Rationale
1	22 December 2025	All	N/A	1st internal version generated
1.1	17 June 2026	All	Modified the internal version for researchers	1st public version generated

All of Us Research Program
Participant Provided Information
Data Management & Sharing Plan (DMSP), CDR v9

Table of Contents

All of Us Research Program
Participant Provided Information
Data Management & Sharing Plan (DMSP), CDR v9

1. Types of Data Produced

1.1. Summarize the types and estimated amount of scientific data expected to be generated.

- Introduction to *All of Us* Survey Collection and Data Transformation Methods
- The level of aggregation – individual
- Estimated survey data in CDR v9 :
    - Basics: 751,242
    - Overall Health: 684,082
    - Lifestyle: 678,028
    - Health Care Access and Utilization (HCAU): 370,956
    - Family Health History (FHH): 151,161
    - Personal Medical History (PMH): 147,449
    - Personal and Family Health History (PFHH): 204,636
    - Social Determinants of Health (SDOH): 318,335
    - Emotional Health History and Well-Being (EHHWB): 137,946
    - Behavioral Health and Personality (BHP): 139,637
- Surveys – User Support
- Development of the Initial Surveys for the *All of Us* Research Program - PubMed

1.2. Describe which scientific data will be preserved and shared, and provide the rationale for this decision based on ethical, legal, and technical factors.

- Introduction to *All of Us* Survey Collection and Data Transformation Methods
- Survey questions and answers are transformed into structural survey metadata and stored and transmitted from the online Participant Portals to the *All of Us* Data and Research Center's Raw Data Repository via an artifact referred to as a "survey codebook." Survey codebooks are created using REDCap data dictionary format and contain not only survey content, but also assigned variables (e.g., codes), branching logic, and field validation rules. Besides serving as the codebook from survey data collection and transmission, resulting REDCap data dictionary artifacts are also made available through the REDCap Consortium's Shared Library for reuse in other studies. Codebook variables (e.g., source codes) are deposited into the Observational Medical Outcomes Partnership (OMOP) Common Data Model's observation table and stored as an *All of Us* specific "PPI" vocabulary. This vocabulary is both browsable and downloadable via Athena, the online OMOP repository. Survey response data is extracted from the *All of Us* program's Raw Data Repository (RDR) and transmitted to the Curated Data Repository (CDR) for the purpose of research-grade dataset creation. Extracted response data is mapped according to the codebook source codes, as well as OMOP concept codes and IDs. All survey-specific OMOP concepts are assessed for Registered and Controlled Tier privacy methodology application and resulting decisions are logged and implemented within the corresponding CDRs.

1.3. List the metadata, other relevant data, or associated documentation that will be made accessible to facilitate interpretation of the scientific data.

- Introduction to *All of Us* Survey Collection and Data Transformation Methods
- Data Dictionaries – User Support
- *All of Us* Survey Codebooks

2. Related Tools, Software, and/or Code

2.1. Indicate whether specialized/licensed software or tools are needed or recommended to access or manipulate data generated by the implementation partner.

No specialized/licensed software or tools are needed.

2.2. Indicate whether custom-created code or in-house software or tools will be needed or created to access or manipulate data generated by the implementation partner. If so, how will these be made available to researchers?

Not applicable to this data. All standard Researcher Workbench functions are used for this data type.

3. Standards for Data and Metadata

3.1. Describe what standards, if any, will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers) to enable interoperability of datasets and resources. If not applicable, indicate that no consensus standards exist.

Physical measurements provided by participants were mapped to standard Logical Observation Identifiers, Names and Codes (LOINC) codes and curated into the OMOP Common Data Model:

Field	PMI Code	LOINC
Heart Rate	heart-rate	8867-4
Blood pressure systolic and diastolic	blood-pressure	55284-4
Systolic Blood Pressure	blood-pressure-systolic	8480-6
Diastolic Blood Pressure	blood-pressure-diastolic	8480-4
Height	height	8302-2
Weight	weight	29463-7
Waist Circumference	waist-circumference	56086-2
Hip Circumference	hip-circumference	62409-8
Head Circumference	head-circumference	9843-4
Body mass index	bmi	39156-5
Are you currently pregnant?	pregnancy-status	66174-4
Growth Percentile Weight for Age	growth-percentile-weight-for-age	22222-0
Growth Percentile Height for Age	growth-percentile-height-for-age	33333-0
Growth Percentile Weight for Length	growth-percentile-weight-for-length	44444-0
Growth Percentile Head Circumference for Age	growth-percentile-head-circumference-for-age	55555-0
Growth Percentile BMI for Age	growth-percentile-bmi-for-age	66666-0

3.2. Describe data formats created, used, and stored. If unusual data formats will be used, discuss the proposed solution for converting data into more accessible formats.

- Introduction to *All of Us* Survey Collection and Data Transformation Methods
- You can view survey questions and learn more about their sources by browsing the *All of Us* Research Hub Survey Explorer.

4. Data Preservation, Access, and Associated Timelines

4.1. Describe where the data will be made publicly available.

Data will be made available to researchers on the *All of Us* Researcher Workbench.

4.2. Describe how the scientific data will be findable and identifiable for researchers.

- Introduction to *All of Us* Survey Collection and Data Transformation Methods
  - Extracted response data is mapped according to the codebook source codes, as well as OMOP concept codes and IDs.
- You can view survey questions and learn more about their sources by browsing the *All of Us* Research Hub Survey Explorer.
- Surveys – User Support

4.3. Describe what scientific data from 1.1. and 1.2 will be made available to other users and when data will become available.

- **Researchers may access the *All of Us* data through three separate tiers:**
  - **Public Tier** data includes only anonymized, aggregated data and is accessible to the public through the *All of Us* Research Hub's Data Browser.
  - **Registered Tier** data includes participant-level data with added transformations to protect participant privacy and is only available to registered researchers.
  - **Controlled Tier** data includes participant-level data with fewer transformations than those in the Registered Tier.

5. Access, Distribution, or Reuse Considerations

5.1. Describe and justify any factors affecting subsequent access, distribution, or reuse of scientific data.

- Overview of *All of Us* Research Program Policies for Researchers
  - **Data User Code of Conduct (DUCC):** Researchers must read and sign the DUCC, adhere to ethical guidelines, comply with laws, respect participant privacy, and use data solely for biomedical or health research. It prohibits discriminatory research and sets restrictions on data sharing without explicit permission.
  - **Data and Statistics Dissemination Policy:** Restrictions on publishing or distributing participant counts of 1 to 20 to protect privacy, with strategies provided for obscuring such data. Exceptions are rare but can be requested.
  - **Publication and Presentation Policy:** Requires submission of peer-reviewed manuscripts to PubMed Central for immediate public access, notification of upcoming publications or presentations, and adherence to Data and Statistics Dissemination Policy regarding participant counts.

- **Stigmatizing Research and Ethical Conduct Policies:** Prohibits research that could lead to social detriment or stigma. Emphasizes inclusivity, ethical research practices, and adherence to the Federal Policy for the Protection of Human Subjects (The Common Rule).
- **Data Access Framework and Egress Alert Policy:** Inclusive access framework with three data tiers (public, registered, controlled). Users must obtain a "data passport" for higher access levels. Egress alerts are triggered by security policy violations, primarily related to data downloads.

- *All of Us* Consent Process: https://allofus.nih.gov/article/all-us-consent-process
- Participant Privacy Protections – User Support

5.2. Describe how access to scientific data will be controlled (i.e., made available by a data repository only after approval).

Access to individual-level participant data is controlled via *All of Us* Data Use and Registration Agreements, which must be completed and approved before researchers gain access to the *All of Us* Researcher Workbench. Additionally, researchers must complete identity verification, responsible conduct of research training, and the Data User Code of Conduct. Additional training is required for researchers with access to the Controlled Tier of the Workbench. More information on data access can be found here.

5.3. Describe protections for privacy, rights, and confidentiality of human research participants.

- Participant Privacy Protections
  - **De-identification and Data Transformation:** The *All of Us* Research Program employs rigorous methods to remove personally identifying information (PII) from participant records. Data is also subject to transformations such as date shifting and generalization to protect privacy and minimize re-identification risks.
  - **Tiered Data Access:** Data are categorized into three tiers—Public, Registered, and Controlled—based on privacy risks and access requirements. Public Tier includes only anonymized, aggregated data; Registered Tier includes participant-level data with significant privacy transformations; Controlled Tier includes more detailed participant-level data with fewer transformations.
  - **Suppression of Sensitive Information:** Explicit identifiers, free-text fields, and sensitive information (e.g., geolocation data, race/ethnicity subcategories, and certain diagnosis codes) are suppressed or generalized to protect participant confidentiality. For example, geographic data smaller than the US state level and specific race/ethnicity subcategories are suppressed in the Registered Tier.
  - **Additional Protective Measures:** The program includes privacy experts to ensure data curation methods effectively protect participant privacy. The Controlled Tier data undergoes fewer transformations but continues to exclude direct identifiers and apply necessary suppressions to protect participant confidentiality.

- *All of Us* Data and Statistics Dissemination Policy

5.4. Describe security measures that will be in place to protect the data from unauthorized access.

- Access to individual-level participant data is controlled via *All of Us* Data Use and Registration Agreements, which must be completed and approved before researchers gain access to the *All of Us* Researcher Workbench. Additionally, researchers must complete identity verification, responsible conduct of research training, and the Data User Code of Conduct. Additional training is required for researchers with access to the Controlled Tier of the Workbench. More information on data access can be found

here.
- https://www.researchallofus.org/privacy-security-protocols/

5.5. Describe if there are factors that limit the ability to share data (e.g., proprietary nature or commercialization of the data).

All of Us Research Program Policies for Researchers

6. Oversight of Data Management and Sharing

6.1. Provide details/roles of personnel responsible for implementing the DMSP, in line with allowable cost guidance for data management and sharing (e.g., data curation, quality checks, developing supporting documentation).

- DRC Roles: Curation team, Curation project management
- NIH Roles: review, finalize, and approve the DMSP

6.2. Provide details of data management and sharing oversight practices throughout the research project workflow. State how often compliance with the DMSP will be verified (e.g., every X months, on the first of each month).

- The DMSP will be updated when changes are made. It will also be verified annually until the next CDR release, when the DMSP is updated.