All of Us Research Program
Apple Health Data
Data Management & Sharing Plan (DMSP), CDR v9

Cover Sheet

Summary Information	
Data type	Apple HealthKit
Data source	For all participants who agreed to share, data were pulled using an Application Programming Interface (API) via the *All of Us* Participant Portal supported by the program's Participant Technology Systems Center (PTSC) (Vibrent Health) and CareEvolution (CE) from December 2020 through November 2024.
Dates reflected in data type	Data from July 2014 through December 2024 collected and shared by PTSC and CE portals.
Cadence for updates	No planned update.
High-level summaries of the data type (materials/links)	Healthkit is data collected from users showing daily activity (both daily summary and intraday values) and physical and vital health measurements. ● ***Activity_measurements:*** Provides the summary of activity for the participants for various types of health data at different specified intervals. ● ***Device and source:*** Shows the type of device used to collect the data including how the data was collected (e.g. Apple watch). ● ***Physical Measurements:*** Provides measurements of the physical characteristics of the participant (e.g. body mass, blood type, height). ● ***Vital Measurements:*** Provides measurements of the vital metrics of the participant throughout the day (e.g. heart rate every 5 seconds during workouts or heart rate every 10 minutes during rest/daily activity).
Data quality (materials/links) ● **Summary** ● **Issue Tracker** ● **Change Logs**	● Summary: o Release Notes: All of Us v9 CDR Release Notes (C2025Q4R6 & R2… o Data Dictionaries: ▪ Registered Tier: All of Us Registered Tier Dataset v9 CDR Da… ▪ Controlled Tier: All of Us Controlled Tier Dataset v9 CDR Dat… ● 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

All of Us Research Program
Apple Health Data
Data Management & Sharing Plan (DMSP), CDR v9

	• Change logs: Any updates to the data are listed in the CDR release notes
***All of Us* Contact**	DRC: support@researchallofus.org

DMSP Version Number and Date:

Summary of Revisions

Version	Date	Section	Revisions Made	Rationale
1.0	14 October 2026	All	N/A	Initial version generated

All of Us Research Program
Apple Health Data
Data Management & Sharing Plan (DMSP), CDR v9

Table of Contents

Cover Sheet **1**
1. Types of Data Produced 4
2. Related Tools, Software, and/or Code 5
3. Standards for Data and Metadata 5
4. Data Preservation, Access, and Associated Timelines 6
5. Access, Distribution, or Reuse Considerations 6
6. Oversight of Data Management and Sharing 8

All of Us Research Program
Apple Health Data
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.

- Apple Healthkit data collected from Apple devices.
- Amount/volume of data to be collected and shared.

Data Type	Table	Number of Participants	Total Row Count	Estimated Size (GB)
Healthkit	**All Healthkit associated tables**	19,922	4,018,080,090	456.74
	activity_measurements	19,859	2,931,284,290	348.84
	device_and_source	N/A	1,091,081	0.10
	physical_measurements	8,762	1,697,379	0.16
	vital_measurements	14,915	1,084,007,340	107.64

- Data is either aggregated to summary per day or intraday for each participant.
- The data is unprocessed.
- The data can be taken as a summary for the day or at specified intervals (intraday) for the participant.

Healthkit is data collected by Participant Technology Systems Center (PTSC) and CareEvolution (CE) from users showing daily activity (both daily summary and intraday values) and physical and vital health measurements.

- ***Activity_measurements:*** Provides the summary of activity for the participants for various types of health data (e.g., StepCount, ActiveEnergyBurned, ExerciseMinutes, etc.) at different specified intervals (e.g., daily, 5 second to 1 minute, etc.).
-
- ***Device and source:*** Shows the type of device used to collect the data (e.g. Apple watch). Note: the person ID is not part of this table. The device source ID field functions as the primary key, and each participant may be linked to multiple devices.
- ***Physical Measurements:*** Provides measurements of the physical characteristics of the participant (e.g. body mass, blood type, height)
- ***Vital Measurements):*** Provides measurements of the vital metrics of the participant throughout the day (e.g. heart rate every 5 seconds during workouts or heart rate every 10 minutes during rest/daily activity).

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.

- During the data transformation process, the following steps are completed to ensure data conforms to both Observational Medical Outcomes Partnership (OMOP) and the *All of Us* Research Program (*All of Us*) standards. Details are referenced in the Curated Data Release (CDR) data dictionary release notes.
    - Validation checks that occur before and after data are transferred to the *All of Us* Data

and Research Center (DRC) for CDR inclusion. The goal is to ensure that data conforms to OMOP naming conventions as well as data type and column structures.
    - Characterization checks that occur after data are transferred to the DRC and are completed before and after CDR inclusion. The goal is to use OMOP-supported tools like the Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (ACHILLES) to assess person-level data characterization and density.
    - Select tables, fields, and concepts are suppressed (e.g., removed). The goal is to ensure that data conforms to the *All of Us* privacy protection principles and rules.
- Data Curation Process – User Support

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

Data provenance including sources:
- Release Notes: All of Us v9 CDR Release Notes (C2025Q4R6 & R2025Q4R6)
- Data Dictionaries:
    - Registered Tier:
      All of Us Registered Tier Dataset v9 CDR Data Dictionary (R2025Q4R6)
    - Controlled Tier:
      All of Us Controlled Tier Dataset v9 CDR Data Dictionary (C2025Q4R6)

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 special software is 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?

No custom-created code or software is needed.

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.

- **Common data model standard:** *All of Us* generally uses the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) for data released on the Researcher Workbench.The data is represented in the OMOP Common Data Model (CDM) structure.
- **Standardization:** The data is in standard OMOP conventions.
- **Metadata and documentation:** Data dictionaries and OMOP table definitions describe table purpose, field meaning, provenance, and any relevant derived-data indicators to support reuse after publication.

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.

- **Data and file formats:** The data is presented in a big query data table.

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.

For all off-cycle releases, researchers must directly query the off-cycle dataset to be able to access and join the off-cycle datasets to the major release dataset. Off-cycle releases will be a part of subsequent major release packages. When transitioning between subsequent major releases, researchers may wish to review their queries to ensure that they have excluded referencing the previous off-cycle releases from their analyses to avoid querying the data types multiple times.

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.

The following will be made available in both Registered Tier and Control Tier.

- ***Activity_measurements:*** Provides the summary of activity for the participants for various types of health data (e.g., StepCount, ActiveEnergyBurned, ExerciseMinutes, etc.) at different specified intervals (e.g., daily, 5 second to 1 minute, etc.).
-
- ***Device and source:*** Shows the type of device used to collect the data including how the data was collected (e.g. Apple watch).
- ***Physical Measurements:*** Provides measurements of the physical characteristics of the participant (e.g. body mass, blood type, height).
- ***Vital Measurements):*** Provides measurements of the vital metrics of the participant throughout the day (e.g. heart rate every 5 seconds during workouts or heart rate every 10 minutes during rest/daily activity).

Data will be released into the production public environment on 9/29/2026.

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 are 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

- Data Curation Process – User Support

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, EHR Operations 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.