Framework for Access to *All of Us* Data Resources *v1.1*

Updated on 8/17/2021

Table of Contents

Executive Summary

This document describes the framework by which individuals will access *All of Us* data resources within the *All of Us* Research Hub. Salient features include:

- The framework is inclusive; in particular, data will be accessible by citizen and community scientists and individuals conducting research outside of academic medical centers.
- There will be three tiers of data access: public data (no login required); registered data (login required); controlled data (additional approval required).
- Authorization for access to the registered and controlled data tiers will be user-based, rather than project-based. Authorized users will receive a "data passport." A data passport is required for access to the registered and controlled data tiers and to set up workspaces to carry out research projects .
- As one of the first steps in initiating an *All of Us* registered and controlled tier project and setting up a workspace, users will be required to submit a description of their project. These descriptions will be made public and searchable for auditing purposes to facilitate

public engagement. This helps to further the program's commitment to partnership and transparency with participants (Precision Medicine Initiative: Privacy and Trust Principles), and for compliance with the 21st Century Cures Act (Pub.L. 114-255).

- Policies will be under the aegis of the Committee on Access Privacy and Security (CAPS). Operational aspects of this framework will be the responsibility of the Resource Access Board (RAB) that reports to the Steering Committee.

Introduction

In constructing a framework for access to *All of Us* Research Program data resources, we have been guided by the following principles:

- Participants are research partners
    - Participant privacy must be protected to the greatest extent possible, in accordance with the PMI Privacy and Trust Principles, and their data must be kept secure, as dictated by the PMI Data Security Policy Principles and Framework.
    - Participants' data may not be used for research purposes that could harm or stigmatize them.
    - Participants must be able to find information on the data that *All of Us* data users have accessed, and for what purpose. Data users must self-declare research purpose when accessing *All of Us* data, and this information must be auditable and made accessible to the participants on a per-research project basis.
- Unlike biospecimens, data is a *non-scarce* resource; as such, it should be as accessible as possible for authorized uses.
    - Data should be available to not only researchers at academic medical centers, but also to users affiliated with industry and citizen and community scientists with no institutional affiliation.
    - No restrictions are placed on the use of *All of Us* resources to develop commercial products and tests to meet public health needs. *All of Us* claims no intellectual property rights on such commercial products developed from research use of *All of Us* data.
    - *All of Us* resources should be accessible to users around the world regardless of country of origin, although the access process may be modified to allow appropriate user authentication.
- We should continuously seek to remove unnecessary barriers to accessing *All of Us* data.
    - For research studies, no group of data users should have privileged access to *All of Us* resources based on anything other than data protection criteria; this includes researchers based at institutions participating in the *All of Us* Research Program. Users must be both trusted and trustworthy.
    - A clear Data User Code of Conduct (DUCC) is needed (stated below) that defines the appropriate use of *All of Us* data resources.
    - We should make resources as broadly available as possible with the expectation that most users can be trusted to follow the DUCC, but we must also have appropriate deterrence and auditing measures in place to promote responsible stewardship of *All of Us* resources..
    - If a user is concerned that their research purpose might stigmatize research participants or be inconsistent with the DUCC, there will be a mechanism to receive guidance from the *All of Us* RAB.
    - If users violate the DUCC, their case will be reviewed by the RAB, which will consider appropriate sanctions.

Within these principles, there is an inherent tension in the need to honor participants' wishes for their data to be both widely accessible and securely protected. Therefore, we must balance these principles to create a framework for achieving these goals. This document describes our proposed framework and addresses the following questions:

1. Who is eligible to become an authorized data user?
2. What steps must individuals take to become authorized users of the *All of Us* data resources?
3. What are some of the potential consequences of violating the DUCC, either intentionally or unintentionally?

Note that this document focuses only on access to data resources. Access to participants and biospecimens will be addressed through subsequent policies.

1. Data User Code of Conduct and Definition of "Authorized User"

Any individual may receive access to *All of Us* data resources (i.e., become an "authorized user") if they follow the appropriate process. An authorized data user is a person who is authorized to access and/or work with registered or controlled tier data from the *All of Us* Research Program. Authorized users receive a data passport that allows them to create workspaces and conduct research. Initially, a data user's institution must enter into an institutional data use agreement with the *All of Us* Research Program for an individual to become an authorized user. Once their institution has entered into the agreement, the individuals must take the following steps to become authorized data users:

1. Provide their identity to the *All of Us* Research Program.
    a. Applications must include proof of identity in accordance with the National Institute of Standards and Technology (NIST) Identity Assurance Level 2 (IAL2) standard; pseudonyms are not allowed. Identity will be verified electronically using a third-party electronic ID verification platform. A manual process for ID verification will be available for individuals who cannot verify using the electronic platform.
    b. Applications must include contact information (email, address), as well as any professional affiliations.
    c. Individuals without traditional forms of identification may petition the program for access. The program, or a specified body within the program, will review such petitions and decide whether or not to grant access.
2. Provide consent for public display of their name and affiliations along with plain language descriptions of their research projects.
3. Provide consent for public release of name and affiliation if the RAB finds that they have violated the DUCC.
4. Complete the *All of Us* Responsible Conduct of Research Training, including modules on data security and participant privacy awareness, and renew this training on an annual basis.
5. Provide a signature that codifies that the user has read, understood, and agrees to abide by the DUCC, and has completed the requisite training.

Within the DUCC, Authorized Data Users attest that "I will":

- read and adhere to the *All of Us* Research Program core values.

- follow all laws and regulations regarding research involving human data and data privacy that are applicable in the area where I am conducting research.
    - In the US, this includes all applicable federal, state, and local laws.
    - Outside of the US, other laws will apply.
- conduct research that follows all policy requirements and conforms to the ethical principles upheld by the *All of Us* Research Program.
- respect the privacy of research participants at all times.
    - I will **NOT** use or disclose any information that directly identifies one or more participants.
        - If I become aware of any information that directly identifies one or more participants, I will notify the *All of Us* Research Program immediately using the appropriate process.
    - I will **NOT** attempt to re-identify research participants or their relatives.
        - If I unintentionally re-identify participants through the process of my work, I will contact the *All of Us* Research Program immediately using the appropriate process.
        - If I become aware of any uses or disclosures of *All of Us* Research Program data that could endanger the security or privacy of research participants, I will contact the *All of Us* Research Program immediately using the appropriate process.
- use the *All of Us* Research Program data **ONLY** for the purpose of biomedical or health research.
- provide a meaningful and accurate description of my research purpose every time I create an *All of Us* Research Program Workspace.
    - Within each Workspace, I will use the *All of Us* Research Program data only for the research purpose I have provided.
    - If I have a new research purpose, I will create a new Workspace and provide a new research purpose description.
- take full responsibility for any external data, files, or software that I import into the *All of Us* Researcher Workbench and the consequences thereof.
    - I will follow all applicable laws, regulations, and policies regarding access and use for any external data, files, or software that I upload into my Workspace.
    - I will **NOT** upload data or files containing personally identifiable information (PII), protected health information (PHI), or identifiable private information (IPI).
    - I will **NOT** use external data, files, or software that I upload into my Workspace for any malicious purpose.
    - If any import of data, files, or software into my Workspace results in unforeseen consequences and/or unintentional violation of these terms, I will notify the *All of Us* Research Program as soon as I become aware using the appropriate process.
- use a version of the *All of Us* Research Program database that is current at or after the time my analysis begins.
- follow all provisions of the *All of Us* Publication and Presentation Policy.

Authorized Data Users further attest that “I will”:

- **NOT** share my login information with anyone, including another Authorized Data User of the *All of Us* Research Program.
    - I will **NOT** create any group or shared accounts.
- **NOT** use *All of Us* Research Program data or any external data, files, or software that I upload into the Researcher Workbench for research that is discriminatory

or stigmatizing of individuals, groups, families, or communities in accordance with the *All of Us* Policy on Stigmatizing Research.
    - I will contact the *All of Us* Research Program Resource Access Board (RAB) for further guidance on this point as needed.
- **NOT** attempt to contact *All of Us* Research Program participants.
- **NOT** take screenshots or attempt in any way to copy, download, or otherwise remove any participant-level data from the *All of Us* Researcher Workbench.
    - I will **NOT** publish or otherwise distribute any participant-level data from the All of Us Research Program database.
    - I will **NOT** publish or otherwise distribute any data or aggregate statistics corresponding to fewer than 20 participants unless expressly permitted under the terms of the *All of Us* Data and Statistics Dissemination Policy.
- **NOT** redistribute or publish any data or statistics with the intent of reproducing the *All of Us* Research Program database or part of the database outside of the *All of Us* Researcher Workbench.
- **NOT** attempt to link registered or controlled tier *All of Us* Research Program data at the participant-level with data from other sources.
- **NOT** use *All of Us* Research Program data or any part of the Research Hub for marketing purposes.
- **NOT** represent that the *All of Us* Research Program endorses or approves of my research unless such endorsement is expressly provided in writing by the *All of Us* Research Program.

On an annual basis, users must re-attest to the DUCC. Each approved data user is also required to update their contact information, affiliation, and research purpose descriptions annually if they have changed. They will be required to complete refresher training modules on the responsible conduct of research, and report any publications resulting from the use of *All of Us* Research Program data. The publications will be posted publicly on the Research Hub website. This will be an additional mechanism to ensure appropriate use of *All of Us* data.

2. Framework for Access to Data Resources

The framework by which a user can access *All of Us* data resources is comprised of several interacting components: a governing body, tiers of data access, and a software platform. In this section, we describe each of these components, beginning with the governing body.

2a: Governance

CAPS is responsible for developing, or assisting in the development of, all policies and procedures relating to the Data Access Framework. CAPS is comprised of representatives from the consortium and HHS with relevant expertise in such fields as data science, privacy and security, and health disparities research; participant representatives; and non-voting liaisons from the *All of Us* Research Program staff. Diversity will also be an important consideration when selecting CAPS membership. Members will serve up to two-year terms, which may be renewable. In addition, committee membership may be expanded, either permanently or temporarily, as necessary to include requisite subject matter expertise in policy- and decision-making processes.

The Resource Access Board (RAB), which also reports to the *All of Us* Steering Committee, is authorized to operationalize decisions regarding this Data Access Framework and DUCC. The RAB is comprised of individuals including:

- investigators with expertise in human subjects research
- individuals representing *All of Us* research participants, including at least one representing a population or group considered to be underrepresented in biomedical research
- individuals to provide perspective from underrepresented populations in biomedical research
- individuals with expertise in ethical/legal/social issues, including those faced by populations that are underrepresented in biomedical research (UBR)
- individuals with expertise in Privacy and Security
- an HHS employee
- a representative of the Citizen or Community Science community

Every attempt will be made for the composition of the RAB to reflect the diversity of the American populace. A quorum will be reached when a majority of the RAB members are present, and at least one of them represents research participants.

The RAB may call ad hoc members to assist with review of specific applications when needed.

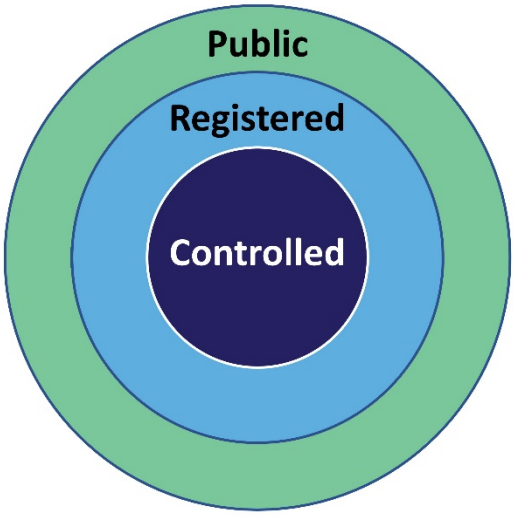
The responsibilities of CAPS and RAB include:

1. Devise policy requirements for registration of new authorized users and for validation of identities (CAPS). For access to controlled data, devise policies for access, including validation of the status of eRA Commons account or other backing, and the requirements for validation of affiliation by an institutional signing official.
    a. CAPS will develop alternate paths to access for diverse types of users, including individuals who are not affiliated with an institution and cannot obtain an eRA Commons Account.
2. Oversee the process of assigning *All of Us* data types to the appropriate tier (i.e., public vs registered vs controlled, defined below; CAPS).
3. Review potential violations of the DUCC and recommend penalties commensurate with the violations. (RAB).

2b: Tiers of Data Resources

There are three tiers of data resources within the *All of Us* Research Hub:
i) Public - Aggregate-level data that poses negligible risks to the privacy of research participants. It can be accessed without logging into the *All of Us* Research Hub at https://databrowser.researchallofus.org/.
ii) Registered- Data that poses a low risk to the privacy of research participants. It can only be accessed after logging into the *All of Us* Researcher Workbench; all access will be logged and may be audited.
iii) Controlled - Data that poses the most significant risks, although still low, to the privacy of research participants; CAPS will define a path to user authorization that appropriately establishes user trustworthiness, with requirements that build on those for the registered tier.

Registered Data can be accessed by any individual who has completed the steps outlined above to become an authorized user. Initially, access to any non-public *All of Us* data, including registered tier data, will be accessible only to authorized users with an eRA commons ID and an Institutional Data Use and Registration Agreement. This requirement will ensure institutional engagement in any research utilizing non-public *All of Us* resources. Following an initial 'beta' period, access to registered tier data will be broadened to include researchers without institutional affiliation and community scientists. The controlled tier data, once available, will include more stringent access requirements to ensure appropriate protection for the release of more granular, and hence, more risky data.



Authorized users will be asked to renew their attestation to the DUCC and refresh their *All of Us* training on the responsible conduct of research annually, along with updating their user profile. In the future, users who do not have an eRA commons ID and are not affiliated with an academic institution will have the opportunity to become authorized users through a process that will be determined by the CAPS.

During the course of the *All of Us* Research Program, data may be shifted between the public, registered, and controlled tiers; CAPS is charged with overseeing the process of how data is assigned to the appropriate tier. In Appendix A, we provide an initial triaging of data types to help make this discussion concrete, but we note that it may evolve during the course of the program.

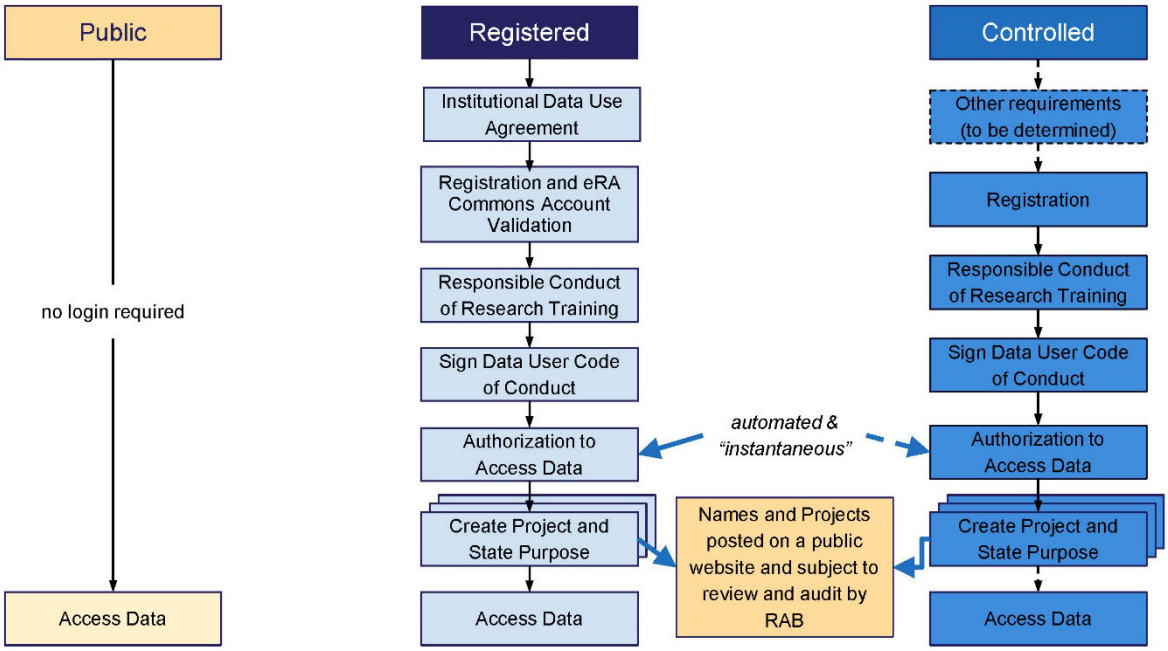
2c: Use of the All of Us Data Resources

All of Us participants consent to the use of their samples and data for general secondary research use without specific data use restrictions. This allows access to *All of Us* data resources to be user-based, rather than project-based. In order to access registered- and/or controlled tier datasets, users will need to complete the data access process described above.

The Data and Research Center (DRC) is charged with building a cloud-based platform for storing *All of Us* data resources and enabling their use by the research community.

The architecture of this platform is closely coupled to the framework proposed here. Specifically, all registered and controlled tier data resources will be stored in a *Curated Data Repository* (CDR), operated by the DRC. To work with the data, authorized users will create new *workspaces* within the platform. These workspaces will serve as analytical sandboxes where users can virtually pull in subsets of data from the CDR and perform analyses within their workspace(s). All analyses will occur within the workspace; data cannot be removed from the data resource with the exception of research analyses with aggregate statistics in buckets of 20 or more individuals.

All of Us Data Access Process



Each workspace will have one or more user/owners that set permissions on who may enter to view data and perform analyses. All users must be data passport holders. These workspaces will account for the tiers of data access and research purposes through the following mechanisms:

1. When a workspace is created, the owner states whether it will contain controlled tier data. If so, the owner and all people granted permission to enter the workspace must be approved to access controlled tier data resources.
2. When a workspace is created, irrespective of whether it contains registered or controlled tier data resources, the owner must provide a summary of their research project via both a structured ontology (which facilitates aggregate analyses on how *All of Us* data is being used), as well as a plain language description. This information will be posted on the Research Hub and the *All of Us* website.
3. Once a workspace has been created, its owner(s) may grant access to additional users, provided all contributing users have been approved to access the appropriate tiers of data within the workspace.

2d: Potentially Stigmatizing Research

As stated above, *All of Us* participants consent to allow their data to be made available for general secondary research use. As such, there are no restrictions on secondary data use, such as restrictions for specific diseases or prohibitions on commercial use. However, there is the potential that some research purposes could lead to stigmatization of research participants or of a particular demographic.

As part of the *All of Us* Responsible Conduct of Research Training, users will be educated that studies of subpopulations or groups have the potential to be stigmatizing, whether these populations are defined by sex, gender, ethnicity, culture, disease status, tribal organization, or any other mechanism. This training will explain to users that it is incumbent upon them to understand the populations and groups they choose to study well enough to identify topics and attributes that are potentially sensitive to that group.

Because stigmatizing research is, to some degree, subjective, it is likely not possible to completely prevent it. The *All of Us* Research Program wants to first encourage, in the strongest possible terms, that users engage directly with the communities they are studying. We believe this is the best possible way to preempt stigmatizing research and involve *All of Us* participants and their communities as true research partners.

To further mitigate the risk stemming from potentially stigmatizing research, the program will implement three additional mechanisms for surfacing those research endeavors that would benefit from further oversight:

1) Users will be given the opportunity to have their research purpose and proposed methods reviewed by the RAB. When a user creates a new workspace following the flow described above, especially one that might involve analyses with potentially stigmatizing connotations for research participants, they will be given the opportunity to submit a description of the research to the RAB for review. The RAB will then evaluate the research purpose and document their concerns or lack thereof that the researcher or program can then point to should questions as to the ethics of the research later arise. If the user elects to have a workspace reviewed in this fashion, the RAB will provide guidance on best practices for complying with the values of *All of Us*. If the research is not deemed consistent with the values of *All of Us,* they will be told that they cannot proceed with it.
2) All persons - including other users, research participants, and members of the public - will have the opportunity to request that the RAB review a given research project for its potential to stigmatize. As stated below, all descriptions of research purpose will be posted on the Research Hub and on allofus.nih.gov, providing transparency into the research that users are conducting with the *All of Us* resources.
3) The RAB will have the authority and ability to review descriptions of research purpose through both manual (i.e., reading the descriptions of research purpose) as well as automated (e.g., keyword searches) mechanisms. They will have the authority to initiate a deeper review of research purpose to determine whether the research is potentially stigmatizing. During the course of the program, we anticipate that the RAB will develop facility at performing screens to effectively identify research purposes that require further oversight.

2e: Compliance with 21st Century Cures Act

Section 2011 of the 21st Century Cures Act requires that ""the Secretary shall…ensure that only authorized individuals may access controlled or sensitive, identifiable biological material and associated information collected or stored in connection with" *All of Us*. Of note, Section 2011(d)(6) requires that in implementing the *All of Us* Research Program, the Secretary shall:

> *"...on the appropriate Internet website of the Department of Health and Human Services, identify any entities with access to such information and provide information with respect to the purpose of such access, a summary of the research project for which such access is granted, as applicable, and a description of the biological material and associated information to which the entity has access."*

Accordingly, data passport holders must create a new workspace for each research project using registered or controlled tier *All of Us* data and provide a plain language description of the research project. .

For each workspace that is created in the *All of Us* Researcher Workbench, the following information will be listed on the Research Hub and on https://allofus.nih.gov/.

1. The investigators who have access to the workspace, their institutional affiliation and role.
2. The research purpose description for the project workspace provided by the researcher in plain language.
3. Relevant publications linked to the workspace, once the researcher submits publication resulting from the work done in the workspace.

3. Role for IRB

As described in Appendix A, we do not anticipate releasing data that directly identifies participants (i.e., data that can be linked to specific individuals by users either directly or indirectly through coding systems), even to users with access to controlled data. These inaccessible data elements include, but are not limited to, personal names, addresses of residence or employment, medical record numbers, and social security numbers. Furthermore, these individual-level data will be coded and authorized users will not be given the key to this code. Therefore, the research that occurs within the Workbench is not subject to IRB review or approval. Users may be bound by institutional policies governing research, which may include local IRB review. Such users should continue to follow their institution's policies and procedures governing research where they are not in conflict with those policies stipulated by *All of Us* in the DUCC and other user-facing program materials.

While not involved in approvals of individual user access applications or research proposals, the *All of Us* IRB has several responsibilities with regards to data access. First, the *All of Us* IRB will be responsible for reviewing the Data Access Framework to ensure appropriate measures have been taken to safeguard participants and their data.

4. Violations of the Data User Code of Conduct

The DUCC Compliance Review Policy lays out the process by which the RAB will evaluate violations of the DUCC and recommend penalties.This process includes, but is not limited to, such stipulations as:

- The RAB will determine whether a data user or group of data users have violated the DUCC and will notify the *All of Us* Research Program of the violation.
- The RAB may:

- Notify the data user(s) of any violation of the *All of Us* DUCC.
- Determine whether any action by the data user is required to remedy the violation, and ensure that the data user has taken the recommended actions.
- Recommend that the Data and Resource Center (DRC) revoke and/or deny access of the data user to all non-public *All of Us* data.
- Recommend that the Data and Resource Center (DRC) post the name and affiliation, if applicable, of the data user on a public *All of Us* Research Program webpage.

Based on the scope and impact of the violation, the *All of Us* Research Program may choose to employ additional sanctions against the authorized user who has violated the DUCC.

Appendix A: Data triaging into data tiers

This revised data triage summarizes the general framework guiding the data tier definition and de-identification rules developed by the DRC.

First, note that collected data elements that directly identify a participant will NOT be released in the CDR. These are data elements include, for example:

a) Participant name or alternate contact names
b) Participant contact or alternate contact information, including full mailing addresses, email addresses, and phone numbers
c) Participant IDs, such as medical record numbers, Social Security Numbers, etc.
d) IP Addresses of participant computers or URLs indicative of a participant's identity
e) Raw medical records data that potentially includes patient identifiers

Access to such identifying data elements is outside the scope of the current document. Although *All of Us* may ultimately create a mechanism, that would include IRB review, by which users can access such identifiable information, these data elements are not covered by this framework.

As overviewed earlier in this document, there will be three tiers of data that are released by the program: Public, Registered and Controlled

1. **Public Tier:** No individual participant level information is included in this tier. This tier contains only summary statistics and aggregate information. Aggregate bin size will be set at 20 individuals. Counts lower than 20 will be displayed as 20; Counts higher than 20 will be rounded up to the closest multiple of 20 (e.g., a count of 1245 will be displayed as 1260). Examples of data within this tier include:
    a. Enrollment data, such as counts by state and census region, stratified by demographics such as gender, age group, and race/ethnicity
    b. Derived or analyzed data, such as genomic summary statistics (in alignment with any outcomes of the NIH proposed policy on data sharing), returning numbers of participants below a threshold to be determined
    c. Precomputed results, such as a genome-wide association study for a given phenotype
    d. Medical records data including binned counts of lab results, medications, and diagnosis and procedure codes. Text fields will not be included.
    e. Aggregate counts of participant provided structured field data such as demographics, access to health care, general health questionnaires, and others. Text fields will not be included.

Ages will be binned using groups spanning 10 years of life, with ages <18 omitted and ages ≥90 bucketed together. (eg: 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+)

2. **Registered Tier**: The Registered Tier data includes participant-level data with a number of transformations to protect participant privacy. The transformations are based on empirical analysis of re-identification risks associated with each data element. The empirical analysis specifically corresponds to the chance that the values encountered in the combination of fields (e.g., age, sex, U.S. state of residence) would uniquely distinguish an individual within the resource.

   The transformations can be summarized as follows:
   Fields that are removed:
   - All free-text fields (PPI) and unstructured documents (EHR)
   - All geo location data smaller than US state except EHR site
   - Race and Ethnicity subgroups (PPI) *(e.g., Hmong, Filipino, or Caribbean)*
   - Living situation (PPI)
   - Active duty military status (PPI)
   - Death causes (EHR) *(e.g., diagnosis codes specifying cause of death)*
   - Diagnosis codes subject to public knowledge (EHR)
   - Billing Codes specifying sex, sexuality or gender categories that are generalized in PPI data (EHR)

   Date Transformation:
   - All dates are shifted backwards by a random number between 1 to 365. The shift is constant for each participant, such that temporality of events is preserved.
   - All participants aged > 89 are removed

   (Demographic) Fields that are generalized:
   Note: Race/Ethnicity, Sex/Gender data from EHR is excluded, and only PPI is included as the primary source for these fields in the Registered Tier.
   - Race/Ethnicity *(Less common races grouped together, selections of two or more races other than Hispanic bundled together as 'Mixed racial group')*
   - Sex at Birth *(Grouped into 'Male', 'Female' and a generalized group)*
   - Gender Identity *(Grouped into 'Man', 'Woman' and a generalized group)*
   - Sexual Orientation *(All selections other than 'Straight' grouped together)*
   - Education *(All categories for less than high school/GED grouped together)*
   - Employment *(Grouped into 'Employed for wages/self-employed' and 'Not currently employed for wages' and 'Prefer not to answer')*

3. **Controlled Tier**: This tier contains data elements that may not, in their own right, readily identify individual participants, but may increase the risk of unapproved re-identification when combined with other data elements. All included data types will be appropriately

pre-processed to minimize that risk, based on input from domain experts, prior to inclusion in this tier of the data resource. Examples of possible data types include:

a. Narrative Participant Provided Information (PPI) (excluding explicit identifiers, such as those enumerated above)
b. Medical records data that has been algorithmically cleaned of direct personal identifiers but which may retain sensitive information in free-text fields, such as clinic notes
c. All dates including month and year, date-shifted to preserve participant privacy
d. Census tract and/or 3-digit zip codes
e. Exact ages, including age > 89
f. Individual-level genomic data (reads and variants)
g. Gene expression, metabolome, proteome data

Appendix B. Glossary of Terms

Authorized User: A person who has been approved to access data resources through the *All of Us* Researcher Workbench. An authorized user may be granted access to the registered or controlled tier.

CAPS: See ***Committee on Access, Privacy, and Security***

Citizen or Community Scientist: Individuals unaffiliated with traditional academic or industrial research institutions who are interested in accessing the data resource to answer research questions; such individuals are broadly defined to include self-experimenters, DIY patient/researchers, professional scientists and trainees, and others who identify as citizen scientists. “Citizen Scientist” is a conventional term used within this field; no official citizenship is expected or implied by this term.

Committee on Access, Privacy, and Security (CAPS): The committee that develops and assists in the development of the policies and implementation of data access for the *All of Us* Research Program; CAPS is overseen by the *All of Us* Steering Committee.

Controlled Tier: The data tier that contains data elements that may not, in their own right, readily identify individual participants, but may increase the risk of unapproved re-identification when combined with other data elements; such data include individual-level genomic data, clinical notes, and narrative data; users must be approved to access the controlled tier, and all access will be logged and may be audited.

Curated Data Repository (CDR): Cloud-based infrastructure where *All of Us* data resources are stored and maintained in a usable format.

Data Passport: An approval model that approves user access to create *All of Us* research projects/workspaces based on user credentials.

Data User Code of Conduct (DUCC): An agreement between Vanderbilt University Medical Center and authorized users of data from the *All of Us* Research Program. The DUCC sets out the terms with which authorized data users must comply, and prospective data users must sign the DUCC to gain access to the registered and controlled tiers.

Direct Identifiers: These are variables that explicitly reveal the identity of an individual (e.g., personal name), provide an ability to readily ascertain an identity (e.g., social security number), or permit direct communication with an individual (e.g., a cell phone number). For purposes of this framework, some data variables and elements may be considered identifiable elements for other purposes and/or when handled by other entities outside the DRC. For example, a direct identifier does not include all the eighteen identifiers under the "Safe Harbor" method as defined under § 164.514(b)(2)(i).

Public Tier: The data tier containing only summary statistics and aggregate information that poses minimal risks to the privacy of research participants; the public tier can be accessed by anyone without logging into the *All of Us* Researcher Workbench.

RAB: See ***Resource Access Board***

Registered Tier: The data tier that contains data elements that have a lower risk of unapproved re-identification as compared with controlled tier data, and thus carries some risk to the privacy of research participants; registered tier data can only be accessed by data passport holding users after logging into the *All of Us* Research Platform and creating an *All of Us* workspace; all access will be logged and may be audited.

Research Hub: A cloud-based portal that acts as an interface between users and the curated data repository; in addition to resource access, the research platform hosts user workspaces.

Researcher Workbench: the cloud-based research platform that the *All of Us* Research Program has created, where users can request access to the data, and once approved, create project-specific workspaces in which to access and analyze the data.

Resource Access Board (RAB): The board that operationalizes decisions regarding data access; responsibilities include: oversight of registration procedures of new data users, review of potentially stigmatizing research proposals, and review of potential violations of the DUCC; the RAB reports to the *All of Us* Steering Committee.

Stigma: A mark of disgrace associated with a particular circumstance, quality, or person; stigmatizing research confers or reinforces stigma and runs counter to the *All of Us* Research Program DUCC.

Workspace: A user-created analytical sandbox within the Researcher Workbench where users can virtually pull in subsets of data from the *All of Us* Research Program database and perform analyses; users must create a new workspace for each research project using *All of Us* Research Program data and provide a plain language description of the research project, as well as other project information, that will be published publicly on an *All of Us* Research Program website.