

All of Us Researcher Workbench Overview



Create a new Workspace

A Workspace is your place to store and analyze data for a specific project. Each Workspace is a separate Google bucket that serves as a dedicated space for file storage. You can share this Workspace with other users, allowing them to view or edit your work. Your Workspace is where you will go to build concept sets and cohorts and launch Notebooks for performing analyses on your cohorts.

Creating a new Workspace:

1. Name your Workspace
2. Select dataset
3. Choose your billing account
4. Select the primary purpose of your project
5. Provide the reason for choosing *All of Us* data for your investigation
6. Describe the specific scientific question(s) you intend to study – this will be published on the Research Hub
7. Explain your anticipated findings from the study
8. Indicate if your study will focus on specific populations
9. If desired, request a review of your research purpose

Create a new Workspace (Required)

Workspace Name ⓘ

Billing Account ⓘ
National Institutes of Health

Research Use Statement Questions
The All of Us Research Program requires each user of All of Us data to provide a meaningful description of the intended purpose of data use for each workspace they create. The responses provided below will be posted publicly in the All of Us Research Hub website to inform research participants. Therefore, please provide sufficiently detailed responses at a 5th grade reading level. Your responses will not be used to make decisions about data access.

Note that you are required to create separate Workspaces for each project for which you access All of Us data, hence the responses below are expected to be specific to the project for which you are creating this particular Workspace

1. What is the primary purpose of your project? (Required)
Please select as many options below as describe your research purpose.

Disease-focused research
The primary purpose of the research is to learn more about a particular disease or disorder (for example, type 2 diabetes), a trait (for example, blood pressure), or a set of related conditions (for example, autoimmune diseases, psychiatric disorders).

Methods development/validation study
The primary purpose of the use of All of Us data is to develop and/or validate specific methods/tools for analyzing or interpreting data (e.g. statistical methods for describing data trends, developing more powerful methods to detect gene-environment or other types of interactions in genome-wide association studies).

Research Control
All of Us data will be used as a reference or control dataset for comparison with another dataset from a different resource (e.g. Case-control studies).

Genetic Research
Research concerning genetics (i.e. the study of genes, genetic variations and heredity) in the context of diseases or ancestry.

Social/Behavioral Research
The research focuses on the social or behavioral phenomena or determinants of health.

Population Health/Public Health Research
The primary purpose of using All of Us data is to investigate health behaviors, outcomes, access and disparities in populations.

Drug/Therapeutics Development Research
Primary focus of the research is drug/therapeutics development. The data will be used to understand treatment-gene interactions or treatment outcomes relevant to the therapeutic(s) of interest.

For-Profit Purpose
The data will be used by a for-profit entity for research or product or service development (e.g. for understanding drug responses as part of a pharmaceutical company's drug development or market research efforts).

Educational Purpose
The data will be used for education purposes (e.g. for a college research methods course, to educate students on population-based research approaches).

Other Purpose
If your Purpose of use is different from the options listed above, please select "Other Purpose" and provide details regarding your purpose of data use here (500 character limit).

2. Provide the reason for choosing All of Us data for your investigation (Required)
(Free text, 500 Character limit)

3. What are the specific scientific question(s) you intend to study? (Required)
If you are exploring the data at this stage to formalize a specific research question, please describe the reason for exploring the data, and the scientific question you hope to be able to answer using the data.
(Free text, 500 Character limit)

4. What are your anticipated findings from this study? (Required)
(Layperson language, 2000 Character limit)

5. Will your study or data analysis focus on specific population(s)? Or do you intend to study your phenotype, disease, or condition of interest with a focus on comparative analysis of a specific demographic group (for example a group based on race/ethnicity, gender, or age)? (Required)

More info on underserved populations ▾

No, I am not interested in focusing on specific population(s) in my research.
 Yes, I am interested in the focused study of specific population(s), either on their own or in comparison to other groups.

If "Yes": Please specify the demographic category or categories of the population(s) that you are interested in exploring in your study. Select as many as applicable.

Race/Ethnicity

Disability status

Age Groups

Access to care

Sex

Education level

Gender Identity

Income level

Sexual Orientation

Other

Geography (e.g. Rural, urban, suburban, etc.)

Request a review of your research purpose for potential stigmatization of research participants ⓘ

More info on stigmatization ▾

I am concerned about potential stigmatization of research participants; I would like the All of Us Resource Access Board (RAB) to review my research purpose. (This will not prevent you from creating a workspace and proceeding.)

Build a Cohort

A “cohort” is a group of participants that a group of researchers are interested in. The cohort builder allows you to create and review cohorts and annotate participants in a researcher’s study group. Cohorts can be built based on several criteria, including demographics, conditions, procedures, drugs/medications, survey responses, etc. Your saved cohort is a list of Participants that meet your defined criteria.



Building a cohort:

- Select criteria
 - INCLUDE participants who meet a specific criteria
 - EXCLUDE participants who meet a specific criteria
- Criteria within a group use the **OR** operator. Participants must meet one or the other criteria but not necessarily meet all the criteria in that group.
- Criteria in separate groups use the **AND** operator, such that the both groups of criteria must be met for a participant to be added to the cohort.
- Define temporal relationships between criteria. For example, a condition diagnosis must have occurred within X days of a selected procedure.
- Review and annotate saved cohorts (see Cohort Builder Tutorial for details).

Review a Cohort

The review set feature allows you to select a subset of your cohort to review participants row-level data and add notes and annotations.

Reviewing a cohort:

1. After saving your cohort, you can do a detailed review of a selected number of participant records by creating a review set.
2. You have the option to view your Cohort Description, including cohort definition, demographic breakdown, top conditions, drugs, procedures and labs found in the EHR records of the cohort.
3. You can also review individual participant records and add record annotations.

[Back to review set](#)
Male Type 2 Diabetes
 no metformin

Participant 10390

Date Range: and

[Summary](#)
[All Events](#)
[Conditions](#)
[Procedures](#)
[Drugs](#)
[Observations](#)
[Physical Measurements](#)
[Labs](#)
[Vitals](#)
[Surveys](#)

Date	Standard Code	Standard Vocabulary	Standard Name	Age At Event	Visit Type
2006-05-26	190237002	SNOMED	Non-toxic uninodular goiter	62	
2006-05-27	36241006	SNOMED	Non-toxic multinodular goiter	62	
2006-06-23	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-06-23	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-06-23	36241006	SNOMED	Non-toxic multinodular goiter	62	
2006-07-21	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-07-21	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-08-05	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-08-21	27059002	SNOMED	Postoperative hypothyroidism	62	
2006-08-21	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	

Cohort Saved Successfully
 The cohort **Test 4** has been saved and can now be used in analysis and concept sets.

What Next?

Create Review Sets
 The review set feature allows you to select a subset of your cohort to review participants row-level data and add notes and annotations.

[CREATE REVIEW SETS](#)

[i](#)

[📄](#)

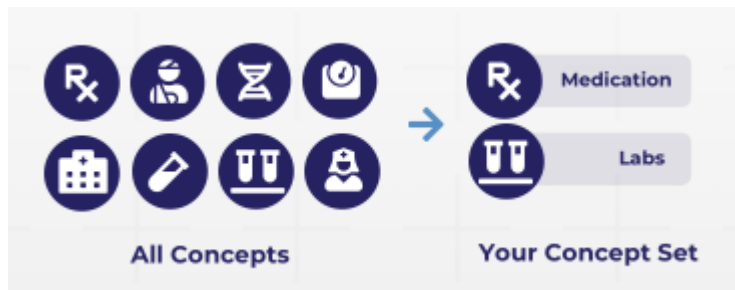
[✍️](#)

Build a Concept Set

Concepts describe information in a patient’s medical record, such as a condition they have, a prescription they are taking or their physical measurements. Subject areas such as conditions, drugs, measurements, etc. are called "domains." Concept sets are saved collections of concepts from a single domain. The set is used when building a dataset to define the variables that you want to see about your cohort of interest.

Building a concept set:

1. Search for your concept of interest in the search bar.
2. Browse domains to see the top concepts.
3. Select a survey to add specific questions to a concept set.
4. Add concepts from a single domain to a concept set.



DATA
ANALYSIS
ABOUT

Search Concepts

Standard concepts only

EHR Domain

<p style="margin: 0;">Conditions 22668 concepts in this domain.</p> <p style="margin: 0;">946237 participants in domain. Browse Domain</p>	<p style="margin: 0;">Drug Exposures 3812 concepts in this domain.</p> <p style="margin: 0;">892671 participants in domain. Browse Domain</p>	<p style="margin: 0;">Measurements 2623 concepts in this domain.</p> <p style="margin: 0;">944511 participants in domain. Browse Domain</p>	<p style="margin: 0;">Procedures 15879 concepts in this domain.</p> <p style="margin: 0;">940058 participants in domain. Browse Domain</p>
--	---	---	--

Survey Questions

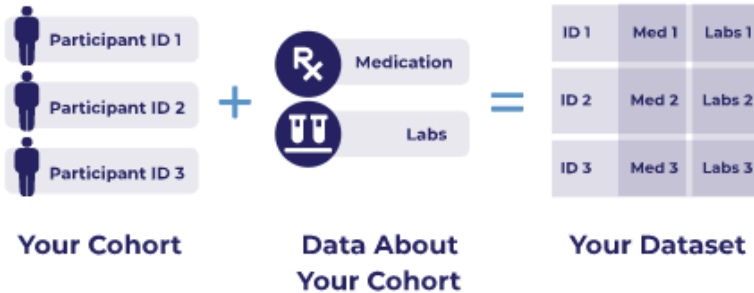
<p style="margin: 0;">The Basics 122 survey questions with 567437 participants</p> <p style="margin: 0;">Survey includes participant demographic information.</p> <p style="margin: 0;">Browse Survey</p>	<p style="margin: 0;">Overall Health 26 survey questions with 567857 participants</p> <p style="margin: 0;">Survey provides information about how participants report levels of individual health.</p> <p style="margin: 0;">Browse Survey</p>	<p style="margin: 0;">Lifestyle 62 survey questions with 568120 participants</p> <p style="margin: 0;">Survey includes information on participant smoking, alcohol and recreational drug use.</p> <p style="margin: 0;">Browse Survey</p>
---	--	---

Build a Dataset

Datasets are analysis-ready tables that can be exported to a Notebook for analysis. Users can build and preview a dataset for one or more cohorts by selecting the desired concept sets and values for the cohorts.

Building a Dataset:

- Select a cohort(s) and concept set(s) and then select the values you want to be added to your dataset.
 - Your cohort provides the list of participants that meet your criteria
 - The concept set defines the concepts you want to see about each of the participants in your cohort
 - Decide which columns you want to see in your dataset
- Preview the dataset before you export it to a notebook.
- Clicking Save and Analyze will export your dataset to an existing or new notebook.



DATA
ANALYSIS
ABOUT

Datasets

Build a dataset by selecting the variables and values for one or more of your cohorts. Then export the completed dataset to Notebooks where you can perform your analysis

1 Select Cohorts (Participants)

Prepackaged Cohorts

All Participants

Workspace Cohorts

diabetes cases

2 Select Concept Sets (Rows)

Prepackaged Concept Sets

Demographics

All Surveys

Workspace Concept Sets

Biometrics

3 Select Values (Columns) Deselect All

Person

PERSON_ID

GENDER_CONCEPT_ID

GENDER

DATE_OF_BIRTH

RACE_CONCEPT_ID

RACE

4 Preview Dataset A visualization of your data table based on concept sets and values you selected above. Once complete, export for analysis

[View Preview Table](#)

PERSON					
PERSON_ID	GENDER_CONCEPT_ID	GENDER	DATE_OF_BIRTH	RACE_CONCEPT_ID	RACE
360870	8507	MALE	1969/12/20 11:13:55	1901001	
361112	8507	MALE	1969/12/26 23:55:40	8516	Black or African American
558740	8507	MALE	1969/12/22 22:42:14	8516	Black or African American
944360	8507	MALE	1969/12/27 05:06:43	8516	Black or African American
158049	8507	MALE	1969/12/24 03:01:26	8516	Black or African American
612558	8507	MALE	1969/12/25 05:47:02	601000	
766651	8507	MALE	1969/12/31 19:04:48	601000	
217931	8507	MALE	1969/12/26 14:31:12	601000	
668697	8507	MALE	1969/12/20 10:26:24	601000	
365975	8507	MALE	1969/12/22 21:15:50	601000	
643525	8507	MALE	1970/01/05 14:02:24	601000	
337939	8507	MALE	1969/12/28 05:52:48	601000	
699589	8507	MALE	1969/12/30 16:49:26	601000	
138039	8507	MALE	1969/12/26 10:32:09	601000	
60341	8507	MALE	1969/12/25 03:10:04	601000	
290648	8507	MALE	1969/12/22 06:25:55	601000	
509385	8507	MALE	1969/12/22 20:48:28	601000	

Save Dataset

Test Dataset

Export to notebook

HIDE PREVIEW

Python
R

```
import pandas

dataset_person_sql = """SELECT
person.BIRTH_DATETIME as DATE_OF_BIRTH,
person.GENDER_CONCEPT_ID, person.PERSON_ID,
person.RACE_CONCEPT_ID,
person.ETHNICITY_CONCEPT_ID,
```

(Create a new notebook) v

Notebook Name

Test Notebook

Programming Language:

Python

R

CANCEL
SAVE AND ANALYZE

Analyze Data in Notebooks

Through the built-in application Jupyter Notebook, users can perform comprehensive analyses on cohorts and data sets using programming languages R or Python. Teams of researchers with various areas of expertise can work together on data cleaning and transformation, statistical modeling, machine learning, and more.

Data Analysis:

1. Export your dataset to a notebook.
2. Run the code to display a dataframe containing the values you selected from your concepts set(s) about your cohort.
3. From here you can run all types of analysis, clearly markdown your notebook, and share with collaborators.

The screenshot shows a Jupyter Notebook titled "Test Notebook (autosaved)". The code cell contains the following Python code:

```
In [1]: import pandas

# This query represents dataset "Test Dataset 1" for domain "person"
dataset_60217493_person_sql = """SELECT person.BIRTH_DATETIME as DATE_OF_BIRTH, person.GENDER_CONCEPT_ID, person.PERSON_ID, p
WHERE person.PERSON_ID IN (select person_id
from `all-of-us-ehr-dev.synthetic_cdr20180606.person` person
where
person.person_id in (select person_id
from `all-of-us-ehr-dev.synthetic_cdr20180606.person` p
where
p.gender_concept_id in unnest(@p0_3062)
)
and person.person_id not in
(select person_id
from `all-of-us-ehr-dev.synthetic_cdr20180606.person` p
where
p.race_concept_id in unnest(@p1_3062)
)
)"""

dataset_60217493_person_query_config = {
  'query': {
    'parameterMode': 'NAMED',
    'queryParameters': [
      {
        'name': "p0_3062",
        'parameterType': {'type': "ARRAY", 'arrayType': {'type': "INT64"}},
        'parameterValue': {'arrayValues': [{'value': 8532}]}
      },
      {
        'name': "p1_3062",
        'parameterType': {'type': "ARRAY", 'arrayType': {'type': "INT64"}},
        'parameterValue': {'arrayValues': [{'value': 8657}]}
      }
    ]
  }
}

dataset_60217493_person_df = pandas.read_gbq(dataset_60217493_person_sql, dialect="standard", configuration=dataset_60217493_
dataset_60217493_person_df.head(5)
```

The output of the code cell is a DataFrame with 5 rows and 8 columns:

```
Out[1]:
```

	DATE_OF_BIRTH	GENDER_CONCEPT_ID	PERSON_ID	RACE_CONCEPT_ID	ETHNICITY_CONCEPT_ID	RACE	GENDER	ETHNICITY
0	1976-02-25 00:00:00+00:00	8532	320685	8515	38003563	Asian	FEMALE	Hispanic or Latino
1	1994-06-04 00:00:00+00:00	8532	838217	8515	38003563	Asian	FEMALE	Hispanic or Latino
2	1954-12-10 00:00:00+00:00	8532	502307	8515	38003563	Asian	FEMALE	Hispanic or Latino
3	1965-04-29 00:00:00+00:00	8532	838129	8515	38003563	Asian	FEMALE	Hispanic or Latino
4	1948-07-30 00:00:00+00:00	8532	176553	8515	38003563	Asian	FEMALE	Hispanic or Latino