All of Us Researcher Workbench Overview

All *of* **Us**
RESEARCH PROGRAM

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.

<u>Creating a new Workspace:</u>

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

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:

1. Select criteria
   - INCLUDE participants who meet a specific criteria
   - EXLCUDE participants who meet a specific criteria

2. 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.

3. 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.

4. Define temporal relationships between criteria. For example, a condition diagnosis must have occurred within X days of a selected procedure.

5. 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.

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

Back to review set
Male Type 2 Diabetes

no metformin

‹	Participant 10390	›

Date Range	Age Range	Visits

Date Range: [   ] 📅 and [   ]

| **Summary** | **All Events** | **Conditions** | **Procedures** | **Drugs** | **Observations** | **Physical Measurements** | **Labs** | **Vitals** | **Surveys** |

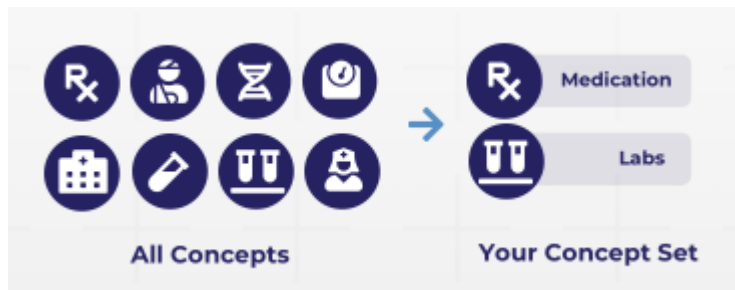| Date	Standard Code ▼	Standard Vocabulary ▼	Standard Name ▼	Age At Event	Visit Type
2006-05-26	190237002	SNOMED	Non-toxic uninodular goiter	62	
2006-05-27	36241006	SNOMED	Non-toxic multinodular goiter	62	
2006-06-23	94098005	SNOMED	Primary malignant neoplasm of thyroid gland	62	
2006-06-23	94098005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

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.

All Concepts → Your Concept Set

| DATA | ANALYSIS | ABOUT |

Search Concepts

Search concepts in domain   ☑ Standard concepts only

EHR Domain

Conditions
22668
concepts in this domain.

946237 participants in domain.
Browse Domain

Drug Exposures
3812
concepts in this domain.

892671 participants in domain.
Browse Domain

Measurements
2623
concepts in this domain.

944511 participants in domain.
Browse Domain

Procedures
15879
concepts in this domain.

940058 participants in domain.
Browse Domain

Survey Questions

The Basics
122
survey questions with
567437 participants

Survey includes participant demographic information.

Browse Survey

Overall Health
26
survey questions with
567857 participants

Survey provides information about how participants report levels of individual health.

Browse Survey

Lifestyle
62
survey questions with
568120 participants

Survey includes information on participant smoking, alcohol and recreational drug use.

Browse Survey

All of Us
RESEARCH PROGRAM

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:

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

Your Cohort + **Data About Your Cohort** = **Your Dataset**
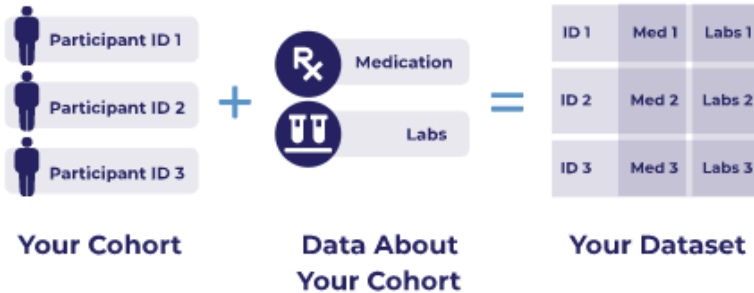
DATA   ANALYSIS   ABOUT

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
- [x] diabetes cases

2 Select Concept Sets (Rows)

Prepackaged Concept Sets
- [x] Demographics
- [ ] All Surveys

Workspace Concept Sets
- [ ] Biometrics

3 Select Values (Columns)   [x] Deselect All

Person
- [x] PERSON_ID
- [x] GENDER_CONCEPT_ID
- [x] GENDER
- [x] DATE_OF_BIRTH
- [x] RACE_CONCEPT_ID
- [x] 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

[x] 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)

Notebook Name

Test Notebook

Programming Language:
- ( ) Python
- ( ) R

CANCEL   **SAVE AND ANALYZE**

All of Us RESEARCH PROGRAM

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.