

Research Information Series

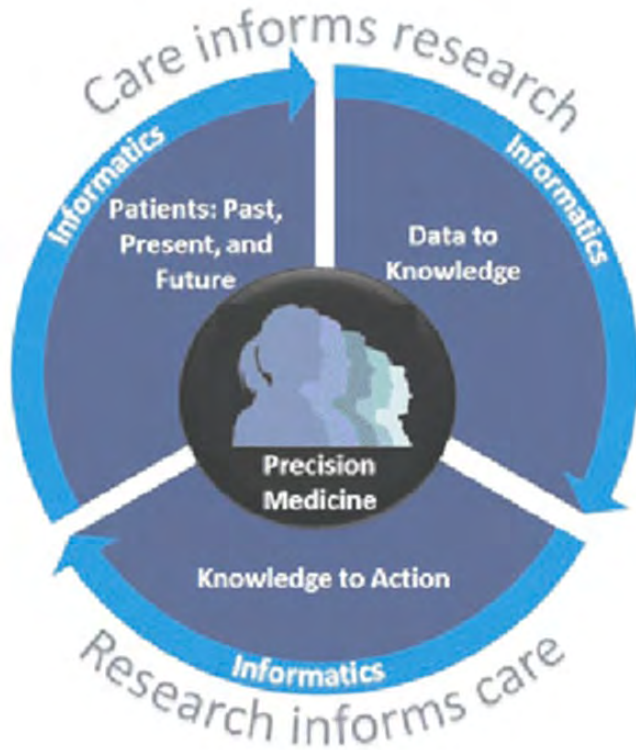
Introduction to Data Science Access and Resources

FAU
SCHMIDT COLLEGE
OF MEDICINE
Florida Atlantic University

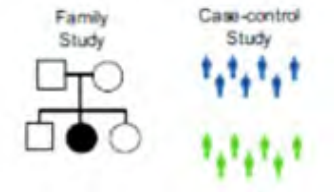
Janet Robishaw, Ph.D.
Joanne Krasnoff, Ph.D.
Oge Marques, Ph.D.

November 28, 2022

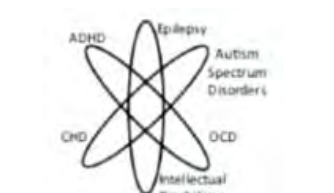
Transformational Research Enterprise



Identifying Causal Genes and Gene-X Interactions



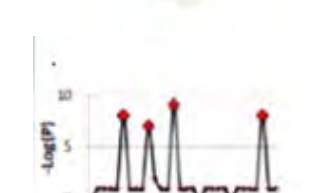
Understanding Phenotypic and Genetic Heterogeneity



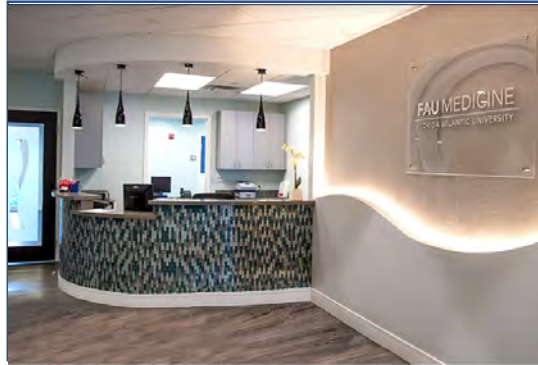
Developing New Diagnostic Tools and Drugs



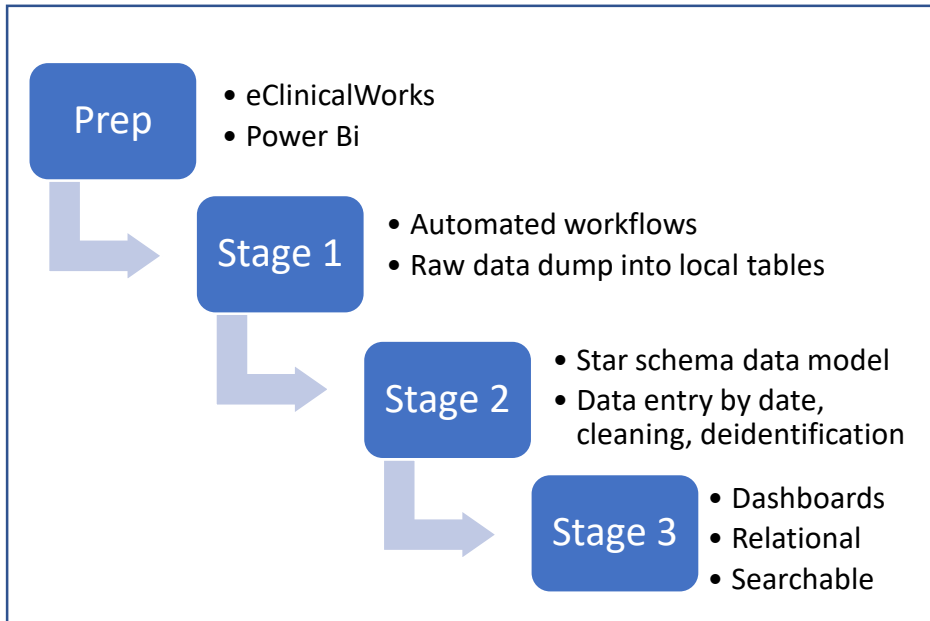
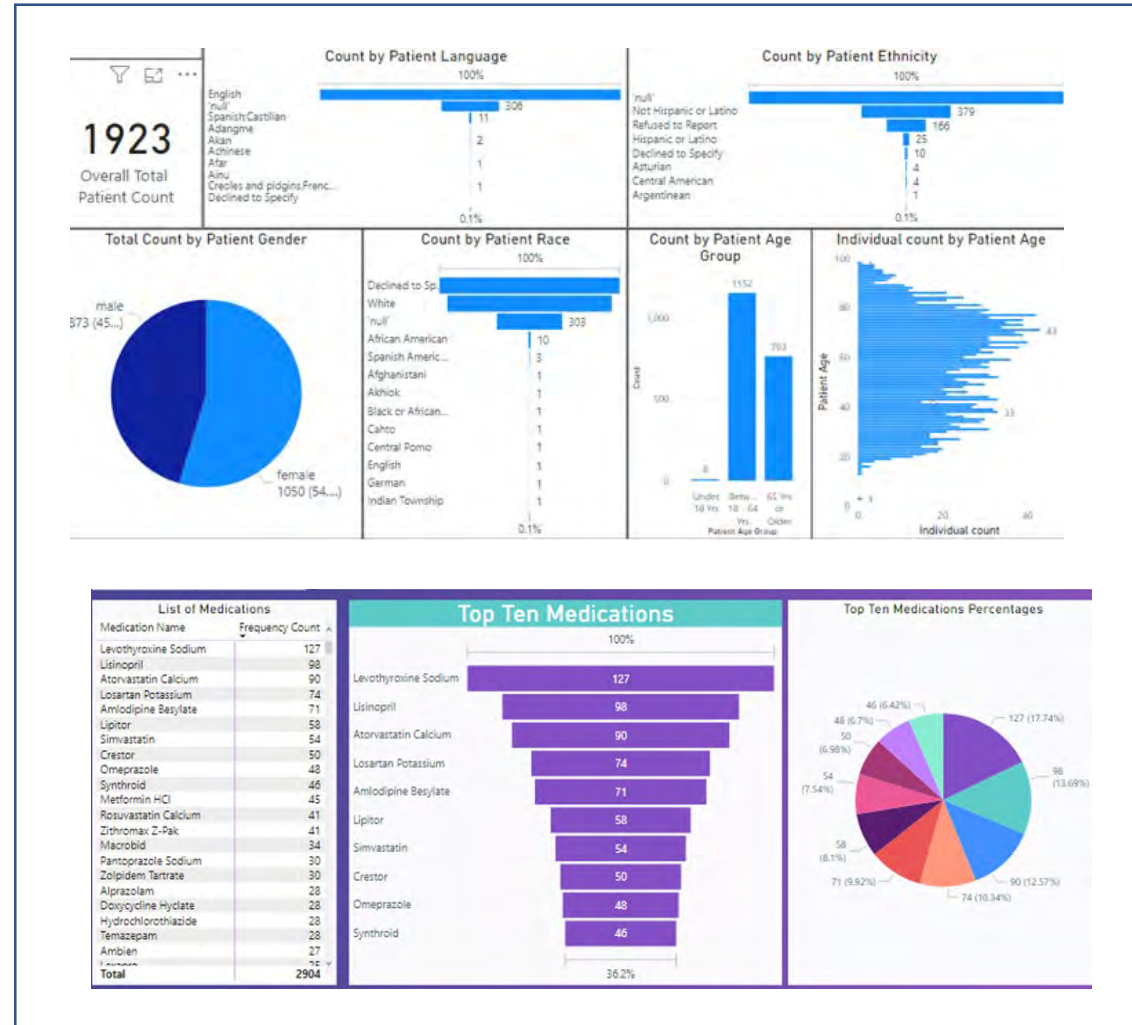
Creating Polygenic Risk Algorithms



Access to Patient Data: FAU Medicine & Hospital Partners



FAU Medicine
Marcus Center
for Integrative Health
Mangurian Center
for Brain Health



Broad Access to Patient Data:



Strengths of EHR Warehouses

- No need for patient recruitment or data collection
- Large sample size
- Diverse population
- Generalizability
- Multi-site studies



All of Us Research Hub | NIH National Institutes of Health All of Us Research Program



FAU Medicine Data Warehouse

Provider name	Patient
	1927
	30
	11
	26
	1509
	2
	357
	356
	100
	77
	369
Total	4070

Race	Patients
	32
ACOMA	2
AFGHANISTANI	2
AFRICAN	1
AFRICAN AMERICAN	14
AKHIOK	1
Total	4078

Ethnicity	Patients
	500
2135-2,2186-5	2
2135-2,ASKU	1
2145-2	173
2145-2,2135-2	1
2145-2,2186-5	37
2145-2,ASKU	1
2155-0,2186-5	1
2186-5,2145-2	1
2186-5,2155-0	1
2186-5 ASKU	3
Total	4078

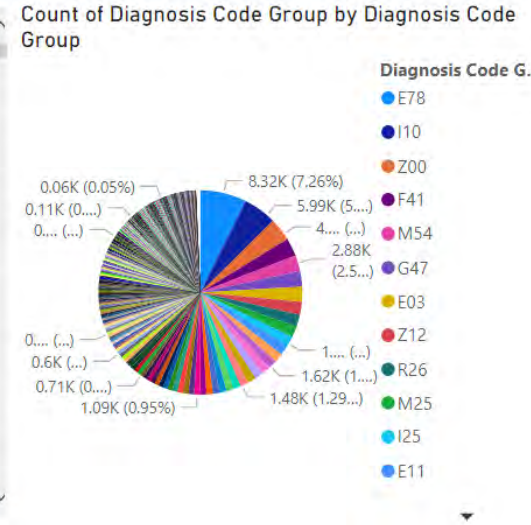
Gender	Patients
F	2214
M	1864
Total	4078

Age Group	Patients
< 18	5
>=65	2010
18-64	2063
Total	4078

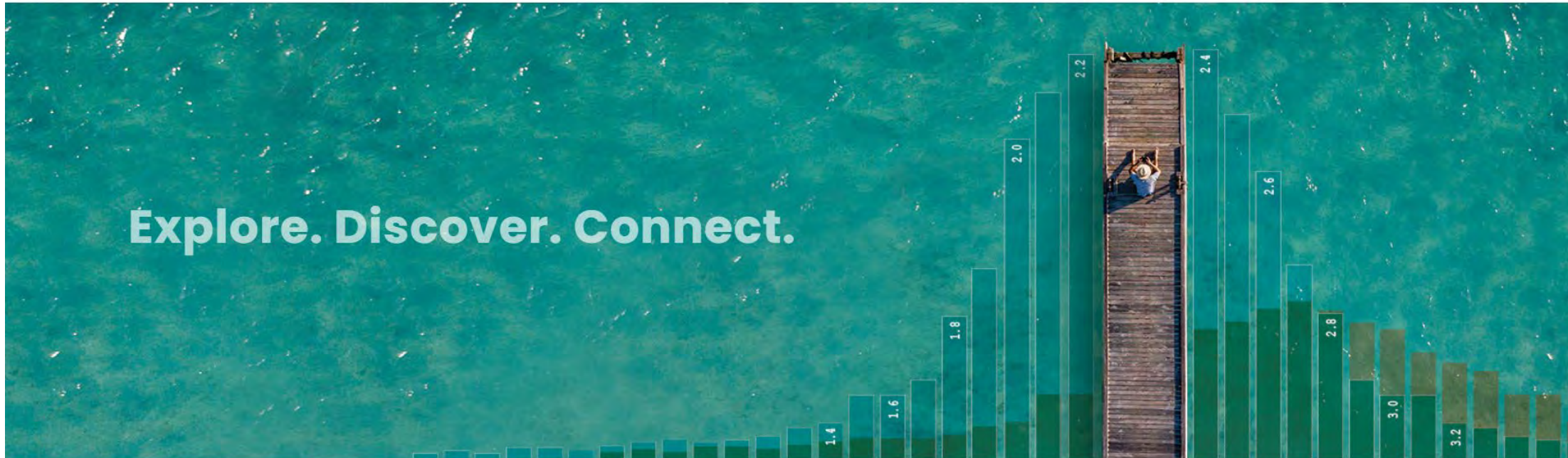
Facility Name	Patients
CCBH (Archive)	403
Florida Atlantic University Integrative Health	268
Florida Atlantic University Primary Care	3577
Florida Atlantic University Primary Care - HOME	8
Total	4078

Diagnosis Code	Diagnosis Description	Patient Count
Z00.00	Encounter for general adult medical examination without abnormal findings	1536
E78.5	Hyperlipidemia, unspecified	1454
I10	Essential (primary) hypertension	1142
M25.50	Pain in unspecified joint	543
Z12.11	Encounter for screening for malignant neoplasm of colon	516
E03.9	Hypothyroidism, unspecified	506
F41.1	Generalized anxiety disorder	439
Z12.31	Encounter for screening mammogram for malignant neoplasm of breast	432
R73.9	Hyperglycemia, unspecified	414
I25.10	Atherosclerotic heart disease of native coronary artery without angina pectoris	403
I78.1	Nevus, non-neoplastic	397
Z00.01	Encounter for general adult medical examination with abnormal findings	395
K21.9	Gastro-esophageal reflux disease without esophagitis	365
Total		4070

Diagnosis Code Group	Patients
Z00	1857
E78	1798
I10	1169
Z12	938
M54	755
F41	716
M25	689
G47	684
R94	533
E03	522
R73	495
Z01	429
R26	426
I25	420
R53	412
I78	399
K21	384
Total	4070



- Demographics
- Diagnoses
- Procedures
- Medications
- Labs & vitals

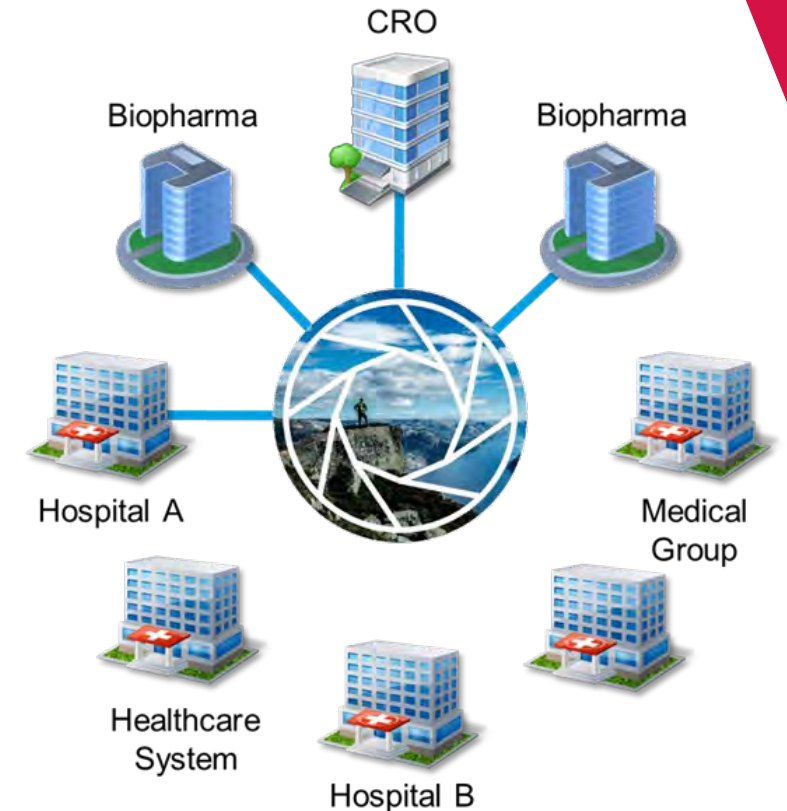


Welcome to the world's largest, living ecosystem of real-world data and evidence for the life sciences and healthcare industries.

<https://trinetx.com/>

TriNetX

- Network is constantly refreshed and growing
 - 120+ HCOs in 19 Countries
 - Anonymized EMR-based-networks of 100m+ patients
 - 70B+ date and patient-indexed clinical observations
- Free of charge to healthcare organizations



TriNetX

Local Research

Provide data and analytics to develop study protocols and pursue grants

Collaboration

Participate in multi-site investigator-initiated trials with peer research organizations

Sponsored Trials

Attract industry-sponsored clinical trials and funding for local research and development

FAU Available Networks

- **Florida Atlantic University**: FAU Medicine patients, on-line analytics, download data sets
- **APAC**: Asia-Pacific healthcare organizations, 6 HCOs, on-line analytics, no dataset download
- **COVID-19 Research**: Research network participants plus several EU sites, which is why there are more patients on this one than plain Research. On-line analytics only, no dataset download
- **EMEA**: Europe/Middle-East/Africa healthcare organizations, 21 HCOs, on-line analytics, no dataset download
- **Global Collaborative**: global healthcare organizations, 93 HCOs, on-line analytics, no dataset download
- **LATAM**: Latin America healthcare organizations, 14 HCOs, on-line analytics, no dataset download
- **Research**: US healthcare organizations plus other global HCOs that allow download data set capabilities, 72 HCOs, on-line analytics, data set download
- **US Collaborative**: Only US sites, 55 HCOs, on-line analytics, no data set download



Live Demo

Next steps:

- Working with TriNetX and OIT – FAU Medicine QA and QC + SSO login
- Platform information and training site
- Honest Data Broker process for providing datasets

Stay tuned for further information on roll-out...

All of USSM

THE FUTURE OF HEALTH BEGINS WITH YOU

The
Precision
Medicine
Initiative[®]

Oge Marques, PhD

Professor

College of Engineering and Computer Science

Charles E. Schmidt College of Medicine

Florida Atlantic University



The future of health begins with you.

The *All of Us* Research Program is inviting one million people across the U.S. to help build one of the most diverse health databases in history. We welcome participants from all backgrounds. Researchers will use the data to learn how our biology, lifestyle, and environment affect health. This may one day help them find ways to treat and prevent disease.

JOIN NOW! 

<https://allofus.nih.gov>

How the Program Works



Participants Share Data



Data is Protected



Researchers Study Data

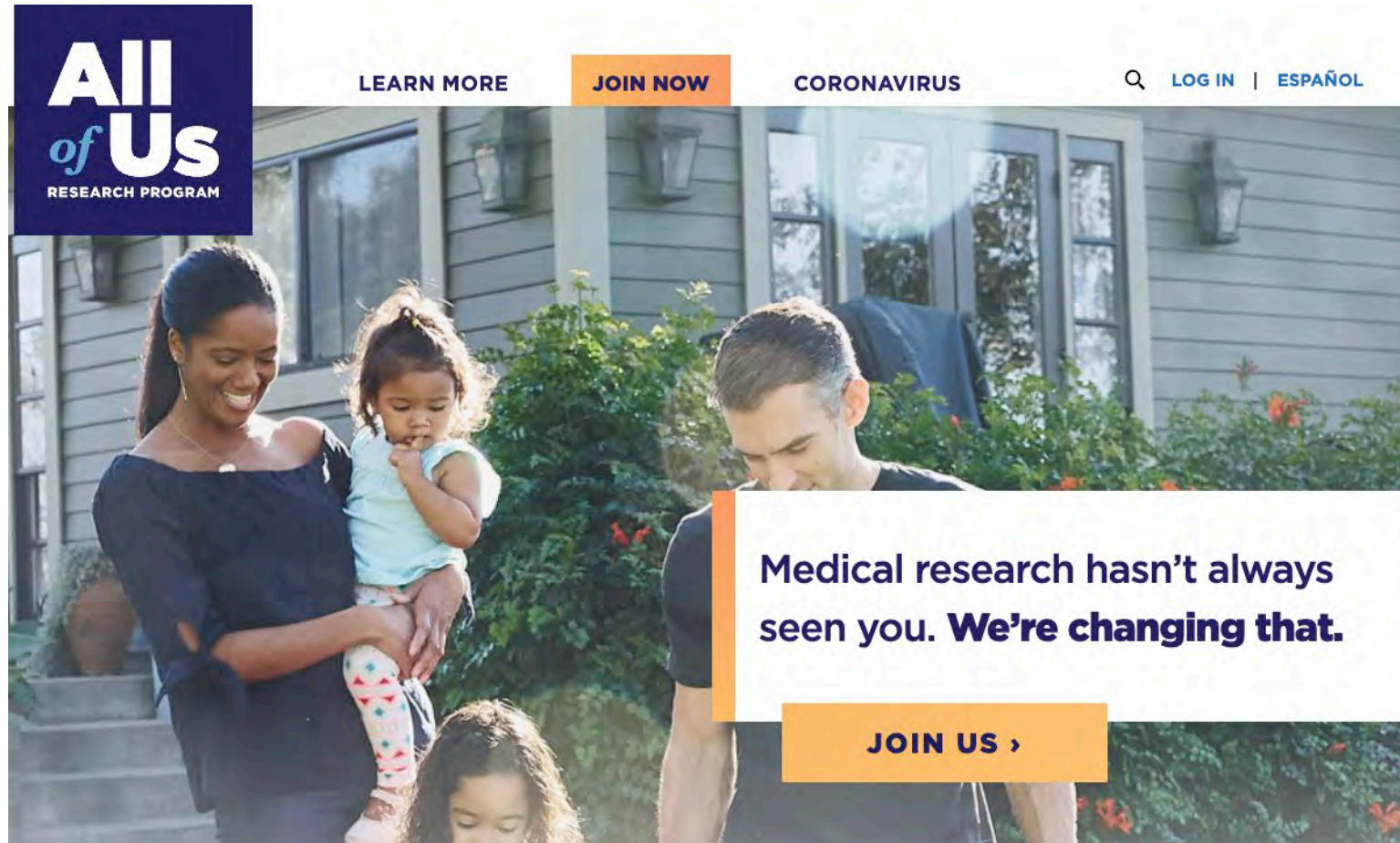


Participants Get
Information



Researchers Share
Discoveries

The Participant side



The image shows the homepage of the All of Us Research Program website. The background is a photograph of a family (a woman, a man, and two children) standing in front of a house. The website's navigation bar is at the top, with the logo on the left and links for 'LEARN MORE', 'JOIN NOW', and 'CORONAVIRUS' in the center. On the right side of the navigation bar, there is a search icon, 'LOG IN', and 'ESPAÑOL'. A large white text box is overlaid on the right side of the image, containing the text: 'Medical research hasn't always seen you. **We're changing that.**' Below this text box is an orange button with the text 'JOIN US >'. The logo on the left is a dark blue square with the text 'All of Us' in white, where 'All' is in a bold sans-serif font, 'of' is in a smaller italicized font, and 'Us' is in a large bold sans-serif font. Below the logo, it says 'RESEARCH PROGRAM' in a smaller white font.

<https://www.joinallofus.org/>

The Participant side

What is the *All of Us* Research Program?



ALL OF US IS A RESEARCH PROGRAM FROM THE NATIONAL INSTITUTES OF HEALTH (NIH).

It is seeking one million or more people from across the U.S. By doing so, it hopes to one day help speed up medical research. People who join will share information about their health, habits, and what it's like where they live. By looking for patterns, researchers may learn more about what affects people's health.

LEARN MORE

<https://www.joinallofus.org/>

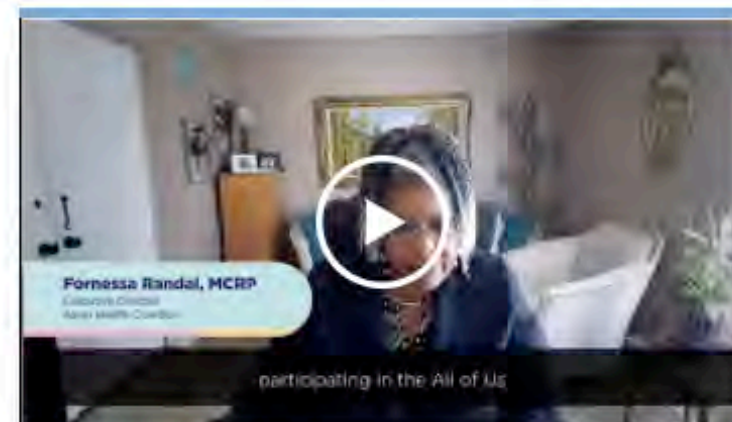
The Researcher side

Opportunities for Researchers

The *All of Us* Research Program is building a dataset to help transform the future of health research by equipping researchers with expansive health data from diverse populations, especially those underrepresented in biomedical research. More than 2,800 health researchers are leveraging this one-of-a-kind dataset to improve understanding of health and disease, identify opportunities to reduce disparities, and enable more precise approaches to care.

On this page:

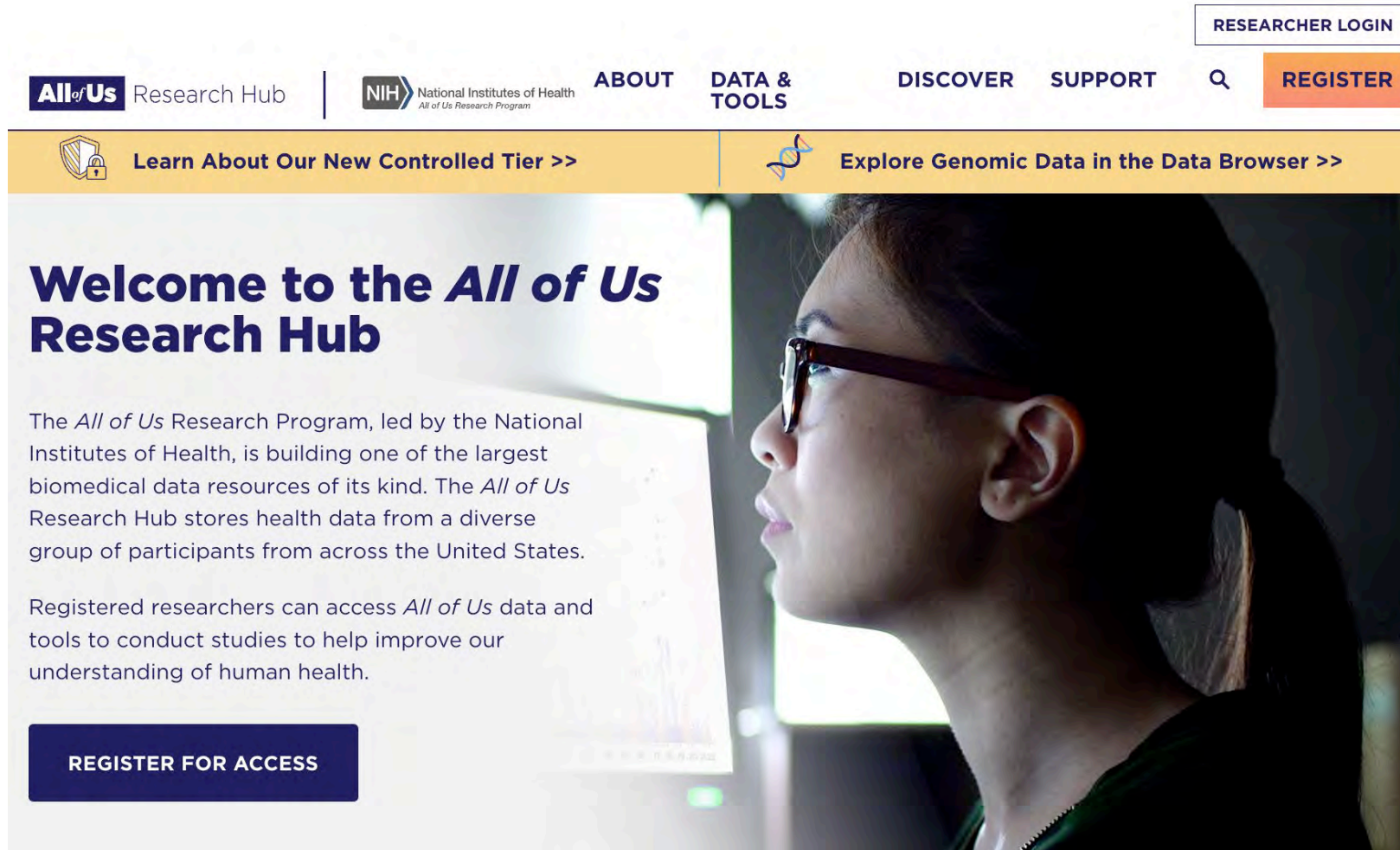
[About the Research Hub](#)
[Benefits for Researchers](#)



Conducting research with the *All of Us* Researcher Workbench (4:21)

<https://allofus.nih.gov/get-involved/opportunities-researchers>

The Researcher side



The screenshot shows the top navigation bar of the All of Us Research Hub website. On the right side of the navigation bar, there is a "RESEARCHER LOGIN" button. Below it, there are links for "ABOUT", "DATA & TOOLS", "DISCOVER", "SUPPORT", and a search icon. A prominent orange "REGISTER" button is also visible. Below the navigation bar, there are two yellow banners: "Learn About Our New Controlled Tier >>" and "Explore Genomic Data in the Data Browser >>". The main content area features a large heading "Welcome to the All of Us Research Hub" and a paragraph describing the program. A dark blue "REGISTER FOR ACCESS" button is located at the bottom left of the main content area. The background of the main content area is a photograph of a woman wearing glasses, looking at a computer screen.

All of Us Research Hub | **NIH** National Institutes of Health
All of Us Research Program

ABOUT **DATA & TOOLS** **DISCOVER** **SUPPORT** **REGISTER**

RESEARCHER LOGIN

Learn About Our New Controlled Tier >> **Explore Genomic Data in the Data Browser >>**

Welcome to the *All of Us* Research Hub

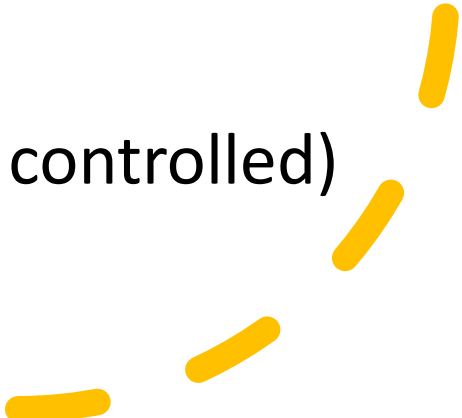
The *All of Us* Research Program, led by the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub stores health data from a diverse group of participants from across the United States.

Registered researchers can access *All of Us* data and tools to conduct studies to help improve our understanding of human health.

REGISTER FOR ACCESS

<https://www.researchallofus.org>

What is All of Us (Research Hub)?

- Centralized, secure, cloud-based platform
 - Allows researchers across a wide range of institutions and at all career levels to execute rapid, hypothesis-driven research
 - Any US-based academic, non-profit, or healthcare organization can enter and utilize the All of Us Research Hub
 - The All of Us data is curated into research datasets
 - Tiered access (public, registered, controlled)
- 

Why multiple tiers?

Public Tier

- Helps exposure
- Increase popularity
- Early assessment of usefulness of data
- However... many limitations (multiples of 20, etc.)

Registered Tier

- Enforces control
- Offers access to a wider range of data
- Greater granularity

Controlled Tier

- Contains genomic data in the form of whole genome sequencing (WGS) and genotype arrays
- Genotype arrays previously suppressed demographic data fields from EHRs, surveys and unshifted dates of events

Data Snapshots

Data Snapshots

Data Snapshots showcase the scale and diversity of the *All of Us* Research Program participant cohort. The snapshots provide participant demographics, geographic distribution, and more. We update the snapshots daily.



527,000+

Participants



322,000+

Electronic Health Records



390,000+

Biosamples Received

VIEW MORE DATA SNAPSHOTS

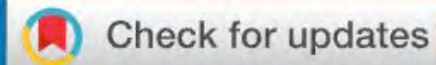
DESCRIPTOR | VOLUME 3, ISSUE 8, 100570, AUGUST 12, 2022

The *All of Us* Research Program: Data quality, utility, and diversity

[Andrea H. Ramirez](#)  ³²  • [Lina Sulieman](#) • [David J. Schlueter](#) • ... [Anthony Philippakis](#) •

[Dan M. Roden](#)   • [the All of Us Research Program](#) • [Show all authors](#) • [Show footnotes](#)

Open Access • DOI: <https://doi.org/10.1016/j.patter.2022.100570> •



<https://doi.org/10.1016/j.patter.2022.100570>

Data browser

- The Data Browser provides interactive views of the publicly-available All of Us Research Program participant data.
 - **Electronic Health Record (EHR) data** are derived from reports by health care providers.
 - **Genomic data** are derived from biosamples provided by participants.
 - **Physical measurements** are taken at the time of participant enrollment.
 - **Data from survey responses** and **wearables data** are collected from participants on an ongoing basis.

In order to protect participant privacy, personal identifiers are removed, aggregate data are rounded to counts of 20, and only summary demographic information is included.


Using the Data Browser

[PUBLIC DATA USE STATEMENT](#)


Search Across Data Types i

x


Data includes 372,380 participants and is current as of 6/6/2022.



FAQs



Introductory Videos



User Guide

EHR Domains

<p>Conditions i</p> <p>12</p> <p>matching medical concepts</p> <p>227,740 participants in this domain</p> <p>View Conditions</p>	<p>Procedures i</p> <p>4</p> <p>matching medical concepts</p> <p>221,860 participants in this domain</p> <p>View Procedures</p>
---	--

Using the Data Browser

Survey Questions

Personal Medical History



1

matching survey questions

142,100 participants in this domain

This survey includes information about past medical history, including medical conditions and approximate age of diagnosis.

[View Complete Survey](#)

Family Health History



1

matching survey questions

145,620 participants in this domain

Survey includes information about the medical history of a participant's immediate biological family members.

[View Complete Survey](#)

Using the Data Browser

Survey Questions

Personal Medical History i

1

matching survey questions

142,100 participants in this domain

This survey includes information about past medical history, including medical conditions and approximate age of diagnosis.

[View Complete Survey](#)

Family Health History i

1

matching survey questions

145,620 participants in this domain

Survey includes information about the medical history of a participant's immediate biological family members.

[View Complete Survey](#)

Using the Data Browser

🔍 sleep apnea ✕

< Back to main search

142,100 1

Participants completed this survey matching of **465** questions available

Download Survey as PDF

Has a doctor or health care provider ever told you that you have or had any of the following respiratory conditions? (select all that apply)

[See Answers >](#)

Using the Data Browser

Has a doctor or health care provider ever told you that you have or had any of the following respiratory conditions? (select all that apply)

[See Answers](#) ▾







Answer	Concept Code ⓘ	Participant Count ⓘ	% Answered out of 142100	
Respiratory Conditions: No Lung Condition	1384615	85,920	60.46%	
Respiratory Conditions: Asthma	43528439	27,240	19.17%	>
Respiratory Conditions: Sleep Apnea	43529837	22,960	16.16%	>
Respiratory Conditions: Other Lung Condition	1384564	6,960	4.90%	>
Skip	903096	6,260	4.41%	
Respiratory Conditions: Chronic Lung	1384567	3,780	2.66%	>
Did not answer	0	≤ 20	0.01%	

Using the Data Browser

Respiratory Conditions: Sleep Apnea	43529837	22,960	16.16%	▼
↳ About how old were you when you were first told you had Sleep Apnea?				
Answer	Concept Code 	Participant Count 	% Answered out of 22960	
How Old Were You Sleep Apnea: Adult	1385003	19,100	83.19%	
How Old Were You Sleep Apnea: Older Adult	1385493	2,980	12.98%	
How Old Were You Sleep Apnea: Elderly	1384967	360	1.57%	
How Old Were You Sleep Apnea: Adolescent	1385492	280	1.22%	
How Old Were You Sleep Apnea: Child	1385417	160	0.70%	
Skip	903096	120	0.52%	







Using the Data Browser

↳ Are you still seeing a doctor or health care provider for Sleep Apnea?

Answer	Concept Code 	Participant Count 	% Answered out of 22960	
Sleep Apnea Currently: Yes	43530234	14,920	64.98%	
Sleep Apnea Currently: No	43529540	7,900	34.41%	
Skip	903096	120	0.52%	

Using the Data Browser

↳ Are you currently prescribed medications and/or receiving treatment for Sleep Apnea?

Answer	Concept Code 	Participant Count 	% Answered out of 22960	
Rx Meds for Sleep Apnea: Yes	43530212	14,640	63.76%	
Rx Meds for Sleep Apnea: No	43529517	8,160	35.54%	
Skip	903096	140	0.61%	

Using the Data Browser

Survey Questions

Personal Medical History 

1

matching survey questions

142,100 participants in this domain

This survey includes information about past medical history, including medical conditions and approximate age of diagnosis.

[View Complete Survey](#)

Family Health History 

1

matching survey questions

145,620 participants in this domain

Survey includes information about the medical history of a participant's immediate biological family members.

[View Complete Survey](#)

Using the Data Browser

[< Back to main search](#)

145,620 1

 [Download Survey as PDF](#)

Participants completed this survey

matching of **104** questions available

Has anyone in your family ever been diagnosed with the following lung conditions? Think only of the people you are related to by blood. Select all that apply.

[See Answers >](#)










Using the Data Browser

Has anyone in your family ever been diagnosed with the following lung conditions? Think only of the people you are related to by blood. Select all that apply.

[See Answers](#) ▾

Answer	Concept Code ⓘ	Participant Count ⓘ	% Answered out of 145620	
Respiratory Condition: Asthma	43528438	27,840	19.12%	>
Respiratory Condition: Chronic Lung Disease	43528553	27,620	18.97%	>
Respiratory Condition: Sleep Apnea	43529836	21,740	14.93%	>
None	903095	12,660	8.69%	▮▮▮
Dont Know	903087	3,980	2.73%	▮▮▮
Skip	903096	1,400	0.96%	▮▮▮
Prefer Not To Answer	903079	60	0.04%	▮▮▮
Did not answer	0	≤ 20	0.02%	

Using the Data Browser

Respiratory Condition: Sleep Apnea43529836		21,740	14.93%	▼
↳ Who in your family has had sleep apnea? Select all that apply.				
Answer	Concept Code 	Participant Count 	% Answered out of 21740	
Father Respiratory Condition: Sleep Apnea	43529833	8,780	40.39%	
Sibling Respiratory Condition: Sleep Apnea	43529838	7,500	34.50%	
Mother Respiratory Condition: Sleep Apnea	43529835	5,700	26.22%	
Grandparent Respiratory Condition: Sleep Apnea	43529834	3,120	14.35%	
Son Respiratory Condition: Sleep Apnea	43529839	1,580	7.27%	
Daughter Respiratory Condition: Sleep Apnea	43529832	880	4.05%	
Skip	903096	240	1.10%	



Researcher Workbench

- The workbench is a cloud-based platform where registered researchers can access Registered and Controlled Tier data.
- It has powerful tools which support data analysis and collaboration.



Electronic Health Records



Biosamples and Bioassays



Surveys



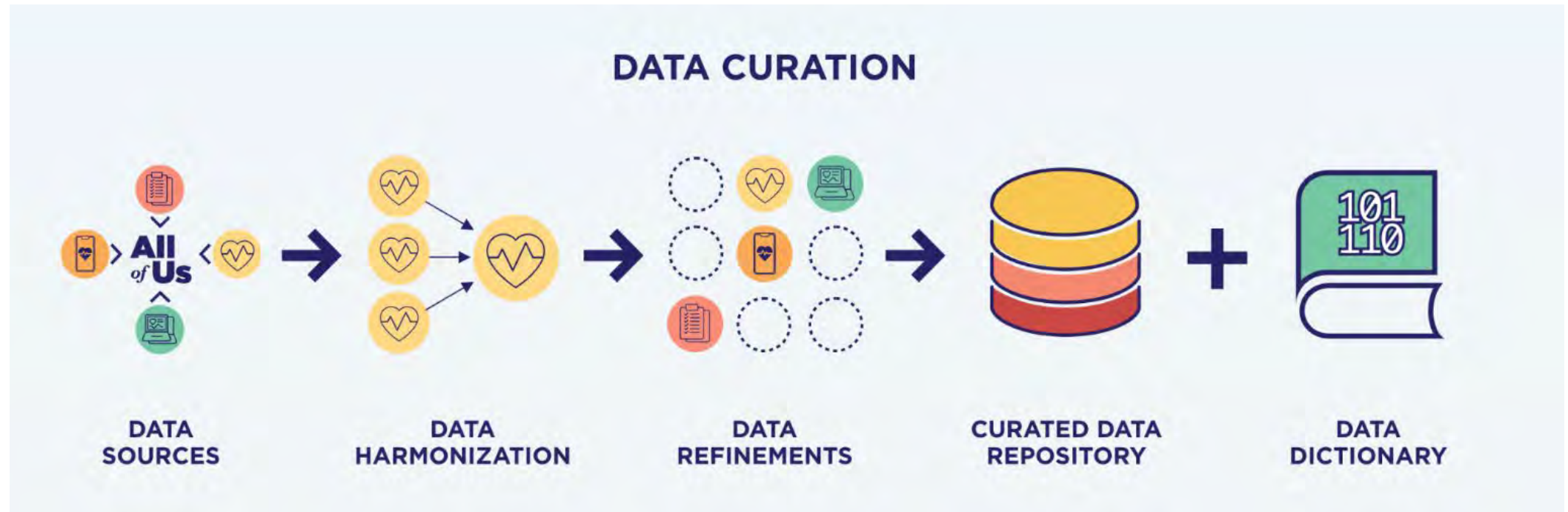
Physical Measurements



Mobile Health Data (Wearable Devices)

Data Sources

Data Methods



Researcher Workbench

•The workbench includes:

- **Workspaces** which are a collaborative environment, where registered researchers can access, store and analyze data for specific research purposes. This allows them to organize research projects and join forces in order to tackle more challenging tasks.
- **Notebooks** are where researchers with R or Python experience can perform high-powered analysis within the All of Us datasets using the integrated, cloud-based Jupyter Notebook environment.
- **Dataset builder** which allows researchers to search and save collections of health information about cohorts called concept sets. It is extremely beneficial when performing pre-populated analysis and dataset reviews.
- **Cohort Builder** is a custom point-and-click tool allowing researchers to create, review and annotate groups of participant data or cohorts within the dataset. The primary focus of the cohort builder is ultimately cohort creation.
- **Workbench User Support Hub** contains a variety of resources, for example, training materials, that demonstrate how to use the tools and navigate the data within the Researcher Workbench. This is used for learning, support and guides.

Home > Discover > Research Projects Directory

Research Projects Directory

3,058 ACTIVE PROJECTS

This information was updated 11/23/2022

The Research Projects Directory includes information about all projects that currently exist in the Researcher Workbench to help provide transparency about how the Workbench is being used. Each project specifies whether [Registered Tier](#) or [Controlled Tier](#) data are used.

Note: Researcher Workbench users provide information about their research projects independently. Views expressed in the Research Projects Directory belong to the relevant users and do not necessarily represent those of the All of Us Research Program. Information in the Research Projects Directory is also cross-posted on [Allofus.nih.gov](https://allofus.nih.gov) in compliance with the 21st Century Cures Act.

Search By: Project title:

2 projects have '**fau**' in the **project title**

Transdisciplinary research



All of Us
RESEARCHER WORKBENCH

DATA ANALYSIS ABOUT

Cohorts +
A cohort is a group of participants based on specific criteria.

All of Us Participants → Your Cohort

- Participant ID 1
- Participant ID 2
- Participant ID 3

Datasets +
A dataset is a table containing data about a cohort that can be exported for analysis.

Your Cohort + Data About Your Cohort = Your Dataset

ID 1	Med 1	Labs 1
ID 2	Med 2	Labs 2
ID 3	Med 3	Labs 3

NSF-Funded Traineeship
in Data Science Technologies and Applications
Cohort 2 Starts Fall 2022

Reminder: Please complete Qualtrics Evaluation

Thank you