ELSI Workshop June 24, 2019

All of Us
Program Data
Where we are and what is to come

Jennifer Ayala, PhD
All of Us Data and Research Center
Vanderbilt University Medical Center

NIH
National Institutes of Health
Outline

- *All of Us* Scientific Framework
- Current data types
- Data Browser walkthrough
- Future data types
- Data lifecycle
- Researcher Workbench preview
All of Us Scientific Framework = Enable Research That Will:

I. Increase wellness and resilience, and promote healthy living

II. Reduce health disparities and improve health equity in underrepresented in biomedical research (UBR) populations

III. Develop improved risk assessment and prevention strategies to preempt disease

IV. Provide earlier and more accurate diagnosis to decrease illness burden

V. Improve health outcomes and reduce disease/illness burden through improved treatment and development of precision interventions
Current *All of Us* Data Types

electronic health records  
surveys  
bio-samples  
genomics  
physical measurements  
digital health
Participant Provided Information

● Current surveys
  • The Basics (demographics, employment, income, contact information)
  • Overall Health (general health, daily activities, and women’s health)
  • Lifestyle (tobacco, alcohol and recreational drug use)
  • Health Care Access and Utilization (access to and use of health care)
  • Family Health History (medical history of immediate biological family members)
  • Personal Medical History (current and past medical conditions, approximate age of diagnosis)

● Upcoming surveys
  • Mental Health (personality, mental health condition diagnosis, depression, suicidality, adversity/trauma, assault, specific disorders)
  • Social Determinants of Health (optimism, social and emotional support, loneliness, perceived stress, discrimination, social status, food insecurity, neighborhood disorder, religion/spirituality)
  • Diet (food/eating behavior)
Participant Provided Information

- Physical measurements
- Height
- Weight
- BMI
- Waist circumference
- Hip circumference
- Blood pressure
- Heart rate
- Heart rhythm status
- Wheelchair use
- Pregnancy
Electronic Health Records

- **Current**
  - Labs
  - Measurements
  - Conditions
  - Procedures
  - Drugs/medications
  - Visit type

- **Future**
  - Notes
  - Images (CT, MRI, etc.)
Welcome to the *All of Us* Research Hub.

The *All of Us* Research Program, part of the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub will store health data from one million or more diverse participants in the *All of Us* Research Program.

Here in the Research Hub, everyone can learn more about the type of data participants are providing and how approved researchers can use our data and tools to conduct...
Other Data Types

- **Bring Your Own Device (BYOD)**
  - Fitbit (>2500 users, data has been collected but not available to researchers yet)

- **Apple Health Kit (on hold)**

- **External Data**
  - Mortality data (National Death Index)
  - Cancer registry data (NAACCR)
Biospecimens

● **Genomics – *All of Us Array* (in progress)**
  - Captures more than 1.8 million sites
  - Supports association studies using common and rare variations across the genome
  - Includes clinically actionable variation in the genome (ClinVar)
  - Includes variants influencing drug transport, metabolism, elimination

● **Salivary Pilot (in progress)**
  - Direct Volunteers
  - Samples processed at Biobank and stored until ready for genotyping

● **Assays (proposed)**
  - HbA1c
  - CBC
  - Heavy metals
  - Pesticides + creatinine
New Data Types Lifecycle

Requirements Gathering
- Participant facing use cases
- Operations support use cases
- Researcher facing use cases

Participant Portal, HealthPro, Biobank, EHR, Devices, External Data Links

Public Data Browser
Launched May 2019

Workbench eRA Commons
Launch Winter 2019/2020
## All of Us Research Data (expected to grow by launch or shortly after)

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Method</th>
<th>Specification</th>
<th>Expected N at launch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood pressure, heart rate, height, weight, BMI, hip and waist circumference, wheelchair use and pregnancy status</td>
<td>Physical Measurements</td>
<td>Baseline</td>
<td>&gt;160,000</td>
</tr>
<tr>
<td>Sociodemographics, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;200,000</td>
</tr>
<tr>
<td>Lifestyle, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;190,000</td>
</tr>
<tr>
<td>Overall health, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;190,000</td>
</tr>
<tr>
<td>Personal medical history, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>Family medical history, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>Health care access and utilization, self-assessment</td>
<td>PPI Survey</td>
<td>Baseline</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>Electronic Health Record Data (diagnoses, drug exposures, lab measurements, and/or procedures)</td>
<td>EHR</td>
<td>Quarterly</td>
<td>&gt;110,000</td>
</tr>
</tbody>
</table>
Research Dataset Timeline *(incorporation into the CDR)*

All dates and order of new data types are tentative and subject to change.
Coming This Winter: Powerful Tools for Collaborative, Reproducible Science
Custom Tools for Exploring and Selecting Subsets of AoU Data

Data
The Data Tab is the gateway to all Workbench tools and All of Us Research data that will help you complete your research project. Here, users can build a cohort of participants, select concept sets of interest, and build analysis-ready tables from the raw data.

Cohorts
A "Cohort" is a group of participants that researchers are interested in. The cohort builder allows you to create and review cohorts and annotate participants in a researcher's study group.

Concept Sets
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." Users can search for and save collections of concepts from a particular domain as a "concept set" and then use concept sets and cohorts to create a dataset, which can be used for analysis.

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

Recent data
- Diabetes Meds
  - Research purpose: Treat or prevent diabetes mellitus... Last changed: 3/13/2019
- All AoU Participants
  - Research purpose: Identifies participants based on demographics, last changed: 3/26/2019
- Diabetes Case 1
  - Research purpose: Identifies participants with diabetes mellitus... Last changed: 3/22/2019
- Diabetes Case 2
  - Research purpose: Identifies participants with diabetes mellitus... Last changed: 3/21/2019

Data Type | Name | Owner | Last Changed
--- | --- | --- | ---
Cohort | Diabetes Control Group | Keri Wol | 3/13/2019
Concept Set | Demographics | Keri Wol | 3/1/2019
Measurements | Physical Measurements | Karthik N | 2/19/2019
Measurements | Blood Pressure | Helen Sullivan | 1/21/2019
Measurements | Vitals | Keri Wol | 12/24/2018
Intuitive, Advanced Tool for Building Cohorts of Participants
**Search Interface for Identifying Medical Concepts of Interest**

Concept Sets
Nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat ut wisi enim ad.

- **Search Term:** Metformin
- **Standard Concepts Only:** Selected

Learn more about OMOP domains

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Drugs</th>
<th>Measurements</th>
<th>Procedures</th>
<th>Demographics</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>4,504</td>
<td></td>
<td>2,981</td>
<td>3,549</td>
<td>4,285</td>
<td>7,500</td>
</tr>
</tbody>
</table>

Showing top 100 of 4,504 Drugs

- **Sulfonylureas:** 12,874
- **Biguanides:** 51,045
- **Glucoselowering drugs excl. insulins:** 21,012
- **Insulin and analogues:** 54,765
- **Drugs used in diabetes:** 102,456
- **Count:**

---

**ADD TO SET**
Point and Click Interface for Creating Analysis-ready Datasets

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.

Select Cohorts
- All Acu. Participants
- Diabetes Case 1
- Diabetes Case 2
- Diabetes Control Group

Select Concept Sets
- Demographics
- Physical Measurements
- Measurements
- Vitals
- Drugs

Values
- Race
- Ethnicity
- Drugs
- Drug.Concept.ID
- Drug.Concept.Name
- Drug.Concept.Code
- Drug.Vocabulary.ID
- Drug.Exposure.Start.Date

Preview Dataset
A visualization of your data table based on the variable and value you selected above.

Demographics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>201235</td>
<td>Acetohexamide</td>
<td>5645415</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12346</td>
<td>00451</td>
<td>Chlorpropamide</td>
<td>5458936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12347</td>
<td>117456</td>
<td>Gliclazide</td>
<td>6802710</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Jupyter Notebooks for Powerful, Interactive Data Analysis
Tools and Services for Research Support

- **Help Desk**
  - General & technical support
  - Community forums
  - Knowledge base

- **Training Materials**
  - Tool user guides
  - Instructional videos
  - Data model tutorials
  - Workspaces w/ example analyses
  - Reusable “code snippets”

- **Data Dictionary**
  - PDF as part of detailed data documentation
  - Integrated directly within Workbench tools
Thank you