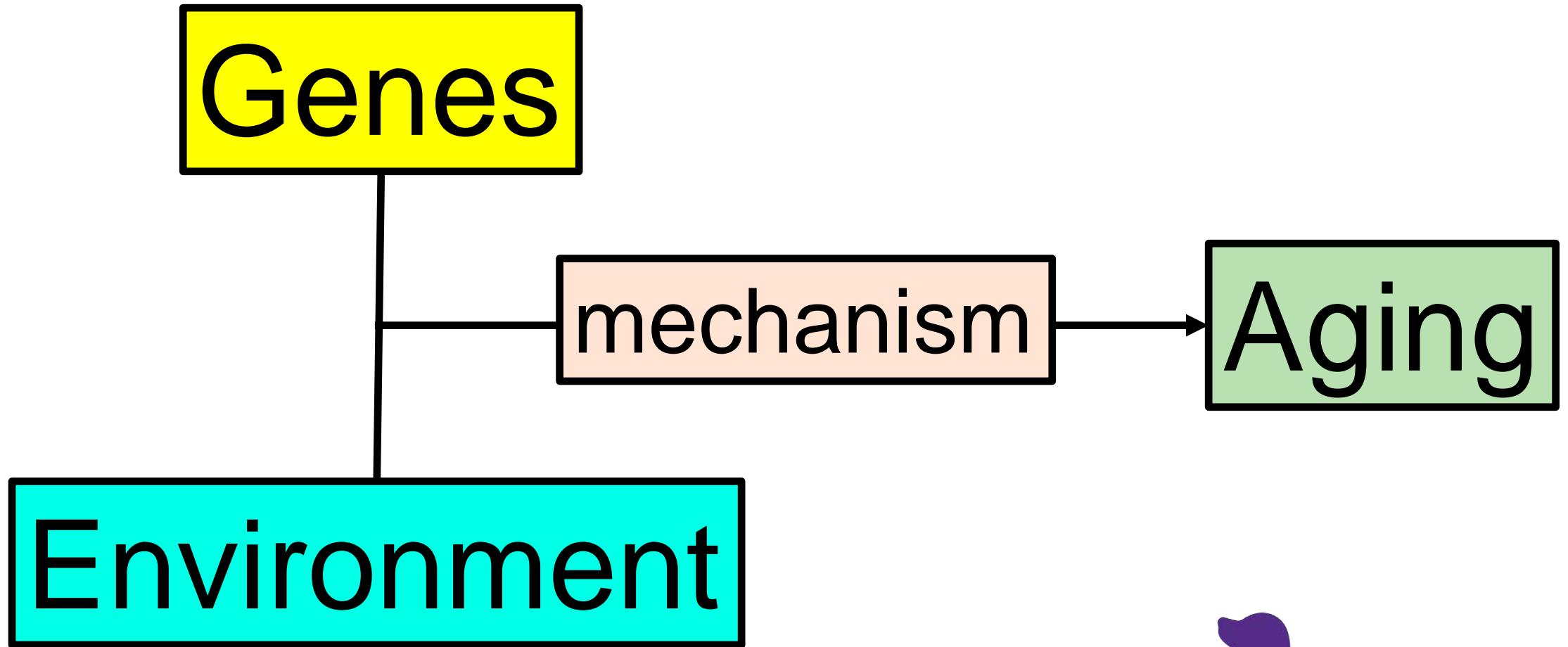


# Development and Use of an AI/ML-Ready Dog Aging Project Dataset

AI Supplements Closeout Meeting  
October 31, 2022

Presenter: Matt Dunbar • [mddunbar@uw.edu](mailto:mddunbar@uw.edu)  
PI: Daniel Promislow





**Dog Aging  
Project**

LONGER, HEALTHIER LIVES.  
TOGETHER.

# DAP is a long-term longitudinal study

U19 Year 5; 5-yr renewal submitted May 2022



## Project 2

## Project 3

## Project 1

**G**

Whole genome sequencing/  
Breed info



**E**

Climate and Air measures  
Sociodemographic variables  
Walkability, density  
Diet, exercise, health care  
Age, size, sex

Clinical chemistry  
Metabolome  
Epigenome  
Microbiome  
Transcriptome  
FACS



**P**

Age-specific disease  
Frailty, Comorbidity  
Mortality  
Fecundity

## Project 4

The TRIAD Trial—Test of Rapamycin in Aging Dogs

# Dog Aging Project dogs

**DAP Pack, 40,000+ dogs,  
Survey data**

**Foundation**  
8500 dogs  
(+DNA, EMRs)

**Precision**  
1000 dogs  
(+ omics, activity)

**Intervention**  
580 dogs  
(+ rapamycin)

Centenarians (~300 dogs)

## Resulting DATA

Baseline and Yearly  
Follow up of Health and  
Life Experience Survey  
+ regular surveys

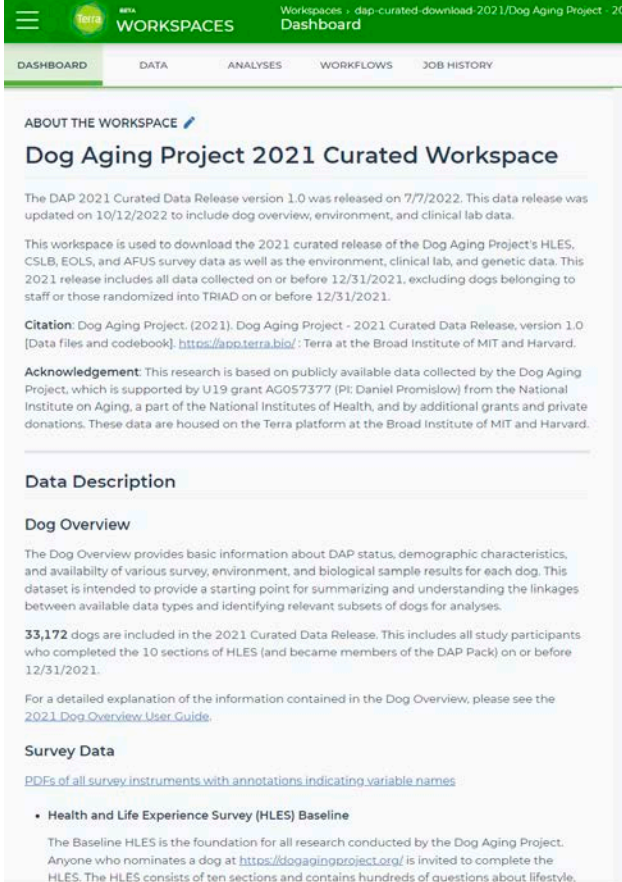
Annual vet check; EMR  
Demography  
Home and local  
environment  
Behavior

Tissue/blood samples  
Activity data  
Necropsy/Pathology

Echocardiograms  
Cognition

# Supplement Specific Aims

1. Develop and implement workflows to ensure that DAP data are AI/ML ready
2. Develop a series of working groups, workshops, and webinars to conduct AI/ML research and train others in applying AI/ML approaches to DAP data



The screenshot shows the Terra Workspaces interface. At the top, there's a green header with the Terra logo and 'WORKSPACES' text. Below the header, a navigation bar includes 'DASHBOARD', 'DATA', 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. The main content area is titled 'ABOUT THE WORKSPACE' and 'Dog Aging Project 2021 Curated Workspace'. It contains several paragraphs of text: a release notice for version 1.0, a description of the workspace's purpose, a citation, and an acknowledgement. Below this is a 'Data Description' section with a 'Dog Overview' subsection, followed by a 'Survey Data' section with a bullet point for 'Health and Life Experience Survey (HLES) Baseline'.

**WORKSPACES** Dashboard

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

ABOUT THE WORKSPACE

### Dog Aging Project 2021 Curated Workspace

The DAP 2021 Curated Data Release version 1.0 was released on 7/7/2022. This data release was updated on 10/12/2022 to include dog overview, environment, and clinical lab data.

This workspace is used to download the 2021 curated release of the Dog Aging Project's HLES, CSLB, EOLS, and AFUS survey data as well as the environment, clinical lab, and genetic data. This 2021 release includes all data collected on or before 12/31/2021, excluding dogs belonging to staff or those randomized into TRIAD on or before 12/31/2021.

**Citation:** Dog Aging Project. (2021). Dog Aging Project - 2021 Curated Data Release, version 1.0 [Data files and codebook]. <https://app.terra.bio/>; Terra at the Broad Institute of MIT and Harvard.

**Acknowledgement:** This research is based on publicly available data collected by the Dog Aging Project, which is supported by U19 grant AG057377 (PI: Daniel Promislow) from the National Institute on Aging, a part of the National Institutes of Health, and by additional grants and private donations. These data are housed on the Terra platform at the Broad Institute of MIT and Harvard.

#### Data Description

##### Dog Overview

The Dog Overview provides basic information about DAP status, demographic characteristics, and availability of various survey, environment, and biological sample results for each dog. This dataset is intended to provide a starting point for summarizing and understanding the linkages between available data types and identifying relevant subsets of dogs for analyses.

**33,172** dogs are included in the 2021 Curated Data Release. This includes all study participants who completed the 10 sections of HLES (and became members of the DAP Pack) on or before 12/31/2021.

For a detailed explanation of the information contained in the Dog Overview, please see the [2021 Dog Overview User Guide](#).

##### Survey Data

[PDFs of all survey instruments with annotations indicating variable names](#)

- **Health and Life Experience Survey (HLES) Baseline**

The Baseline HLES is the foundation for all research conducted by the Dog Aging Project. Anyone who nominates a dog at <https://dogagingproject.org/> is invited to complete the HLES. The HLES consists of ten sections and contains hundreds of questions about lifestyle.

# AI/ML Accomplishments and Plans

- AI/ML Ready Data:
  - Automating data ingestion, cleaning, and structuring. Developing detailed data documentation and metadata.
- AI/ML Research and Training:
  - Data Core and Systems Biology Project are working closely with an ML specialist at Microsoft to develop ML approaches for data capture, including EMR scraping and ML analysis approaches for building predictive biomarkers of aging.
  - Ongoing work creating training modules with our amazing partners at the University of Washington's eScience institute

# Thank you!