



ScHARe

Research Think-a-Thons



National Institutes of Health

The logo for ScHARe, featuring the text "ScHARe" in a bold, dark blue, sans-serif font. The letters "S", "H", and "R" are significantly larger than the other letters. The logo is centered within a white circle.

Be a Part of the Future of Knowledge Generation Part II

July 17, 2024

Deborah Duran, PhD • NIMHD

Luca Calzoni, MD MS PhD Cand. • NIMHD



Look deeper with more eyes

“For the first time in history, we have a technology (AI) that is opening our eyes to who we are, is changing us as we speak, and could allow us to play a conscious role in who we want to become.”

Jennifer Aue

IBM Director for AI Transformation
AI professor at the University of Texas

- **Diverse perspectives**
- **Bias mitigation strategies**
- **Research paradigm shift to Big Data**



ScHARe

Science
collaborative for
Health disparities and
Artificial intelligence bias
Reduction

Outline

- 10'** Introduction
- 25'** What is ScHARe?
- 10'** The ScHARe Think-a-Thons
- 35'** Making data AI-ready
- 25'** Ethical and transparent AI
- 15'** Computational strategies: traditional statistics
- 25'** Computational strategies: AI and Machine Learning
- 5'** Resources

Experience poll

Please check your level of experience with the following:

	None	Some	Proficient	Expert
Python	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cloud computing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Terra	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Health disparities research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Health outcomes research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Algorithmic bias mitigation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Interest poll

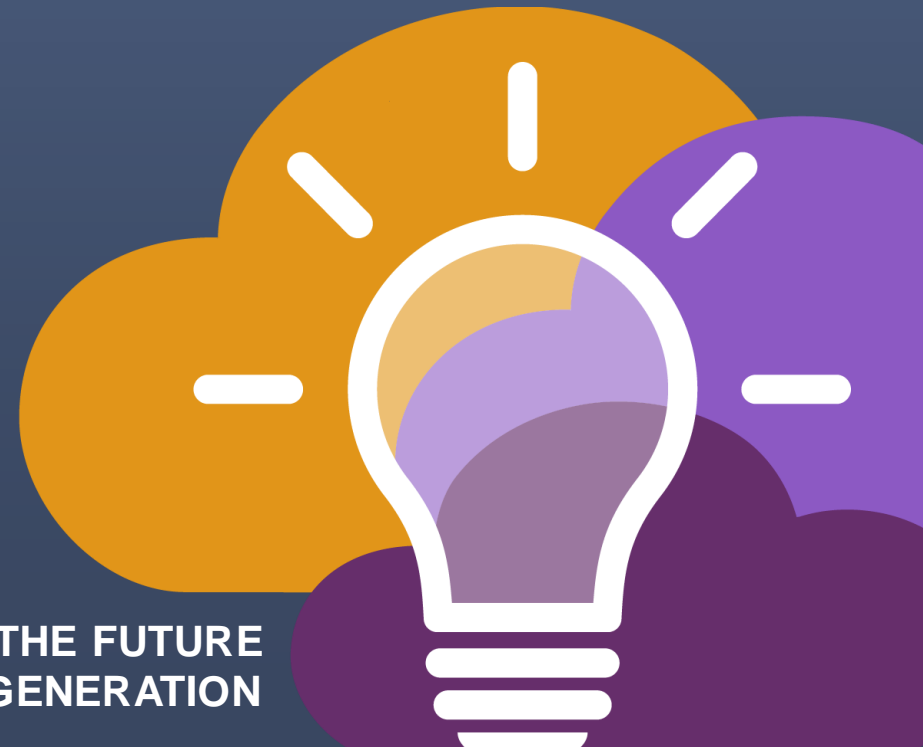
I am interested in (check all that apply):

- Learning about Health Disparities and Health Outcomes research to apply my data science skills
- Conducting my own research using AI/cloud computing and publishing papers
- Connecting with new collaborators to conduct research using AI/cloud computing and publish papers
- Learning to use AI tools and cloud computing to gain new skills for research using Big Data
- Learning cloud computing resources to implement my own cloud
- Developing bias mitigation and ethical AI strategies
- Other

ScHARe

What is ScHARe?

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



ScHARe is a **cloud-based population science data platform** designed to accelerate research in health disparities, health and healthcare delivery outcomes, and artificial intelligence (AI) bias mitigation strategies

ScHARe aims to fill **four critical gaps**:

- Increase participation of **women & underrepresented populations with health disparities** in data science through data science skills training, cross-discipline mentoring, and multi-career level collaborating on research
- Leverage population science, SDoH, and behavioral Big Data and cloud computing tools to foster a **paradigm shift** in healthy disparity, and health and healthcare delivery outcomes research
- **Advance AI bias mitigation and ethical inquiry** by developing innovative strategies and securing diverse perspectives
- Provide a **data science cloud computing resource** for community colleges and low resource minority serving institutions and organizations

ScHARe



nimhd.nih.gov/schare



ScHARe



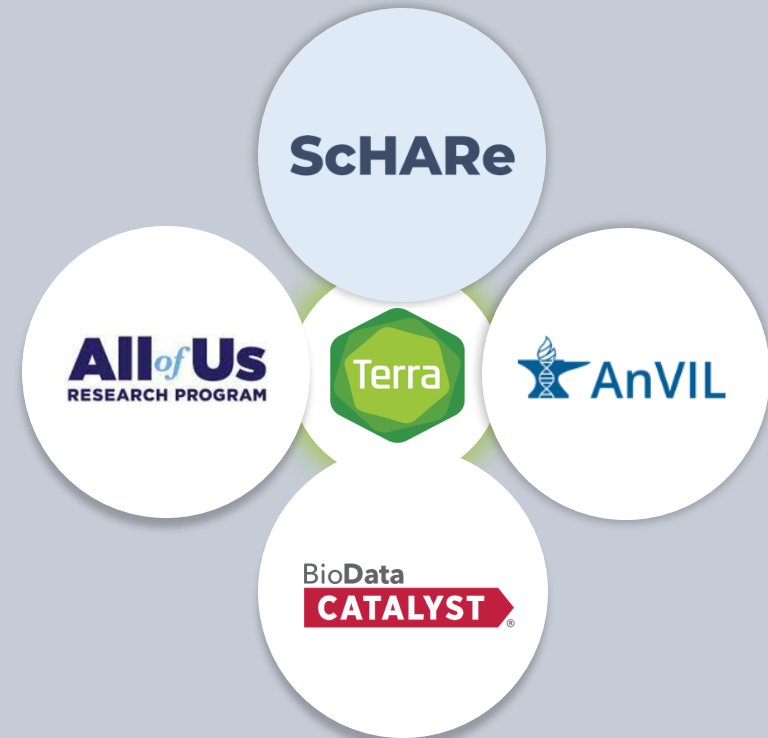
Google Platform Terra Interface

- Secure workspaces
- Data storage
- Computational resources
- Tutorials (how to)
- Copy-and-paste code in Python and R
- Learning Terra on ScHARe prepares you to use other NIH platforms



Terra recommends using **Chrome**
Must have a **Gmail** friendly account

PREPARING FOR AI RESEARCH
AND HEALTHCARE USING BIG DATA
Mapping across cloud platforms with
Terra interface for collaborative research



BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

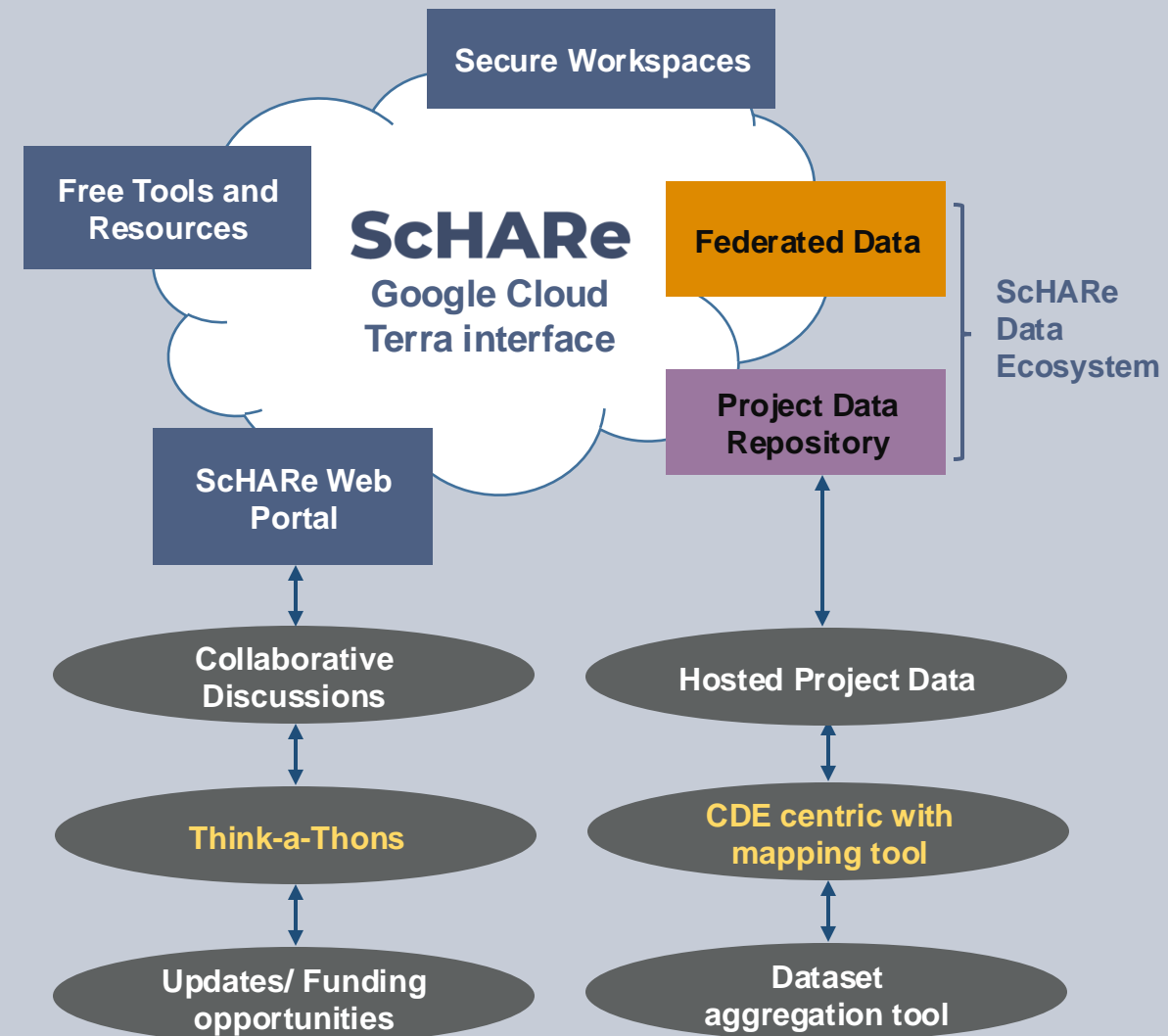


ScHARe Components

ScHARe co-localizes within the cloud:

1. **Datasets** (including social determinants of health and social science data) relevant to minority health, health disparities, and healthcare outcomes research
2. **CDE-focused data repository** to comply with the required hosting and sharing of data from NIMHD-/NINR-funded programs
3. **User-friendly computational capabilities and secure, collaborative workspaces** for students and all career level researchers
4. **Tools for collaboratively evaluating and mitigating biases** associated with datasets and algorithms utilized to inform healthcare and policy decisions (*upcoming*)

Intramural and Extramural Resource



ScHARe Terra interface: secure workspace

The screenshot displays the ScHARe Terra interface with a 'Share Workspace' modal dialog open. The dialog is titled 'Share Workspace' and contains the following elements:

- User email:** A text input field with the placeholder 'Add people or groups' and an 'ADD' button.
- Current Collaborators:** A list of collaborators with their roles and permissions:
 - calzonil2@nih.gov:** Role: Owner (dropdown), Permissions: Can share, Can compute.
 - ScHARe-Contractors@firecloud.org:** Role: Writer (dropdown), Permissions: Can share, Can compute. Includes a close button (X).
 - ScHARe-Read-Only-Access@firecloud.org:** Role: Reader (dropdown), Permissions: Can share, Can compute. Includes a close button (X).
- Share with Support:** A toggle switch currently set to 'No'.
- Buttons:** 'CANCEL' and 'SAVE' buttons at the bottom right.

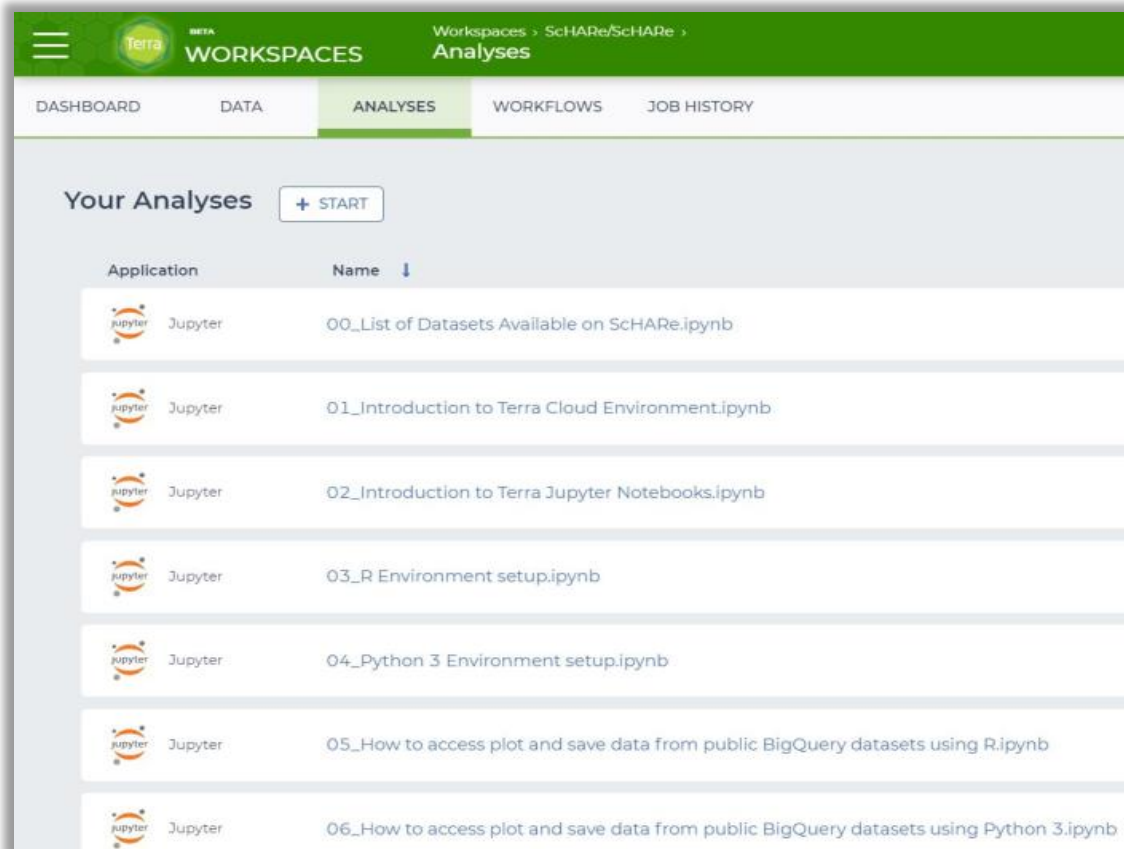
The background interface shows the 'WORKSPACES' section with a search bar and a list of workspaces:

Name
ScHARe
ScHARe Think-a-Thons

- Secure workspace for self or collaborative research
- Assign roles: review or admin
- Host own data and code

ScHARe Terra interface: analyses

Notebooks for analytics and tutorials

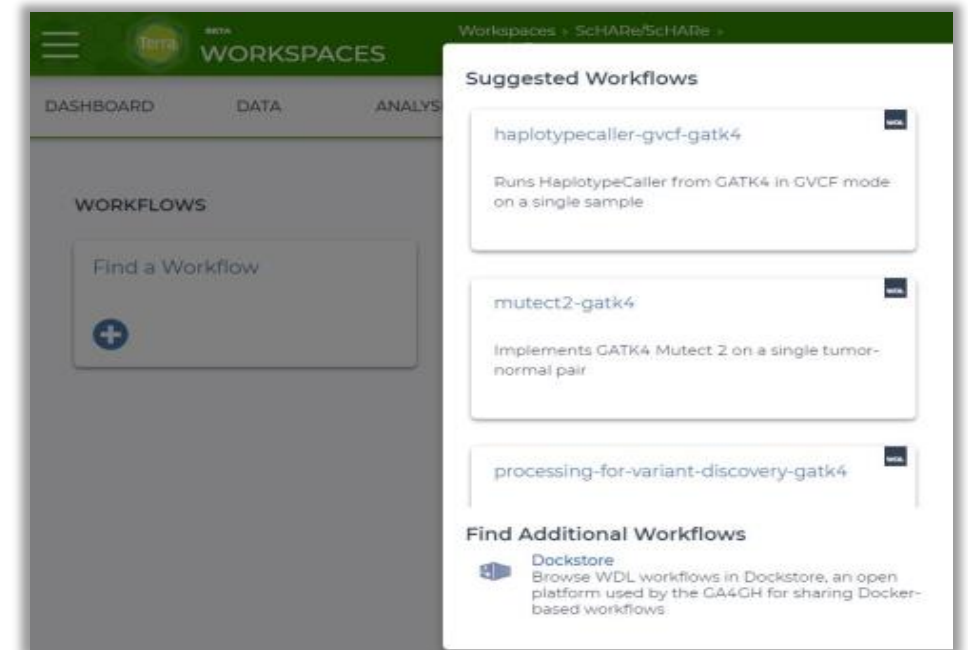


The screenshot shows the Terra interface with the 'ANALYSES' tab selected. The page title is 'Your Analyses' with a '+ START' button. Below is a table of analyses:

Application	Name ↓
Jupyter	00_List of Datasets Available on ScHARe.ipynb
Jupyter	01_Introduction to Terra Cloud Environment.ipynb
Jupyter	02_Introduction to Terra Jupyter Notebooks.ipynb
Jupyter	03_R Environment setup.ipynb
Jupyter	04_Python 3 Environment setup.ipynb
Jupyter	05_How to access plot and save data from public BigQuery datasets using R.ipynb
Jupyter	06_How to access plot and save data from public BigQuery datasets using Python 3.ipynb

Modular codes

- Easy-to-use copy-and-paste analytics



The screenshot shows the Terra interface with the 'WORKFLOWS' tab selected. A 'Suggested Workflows' panel is open, displaying a list of workflows:

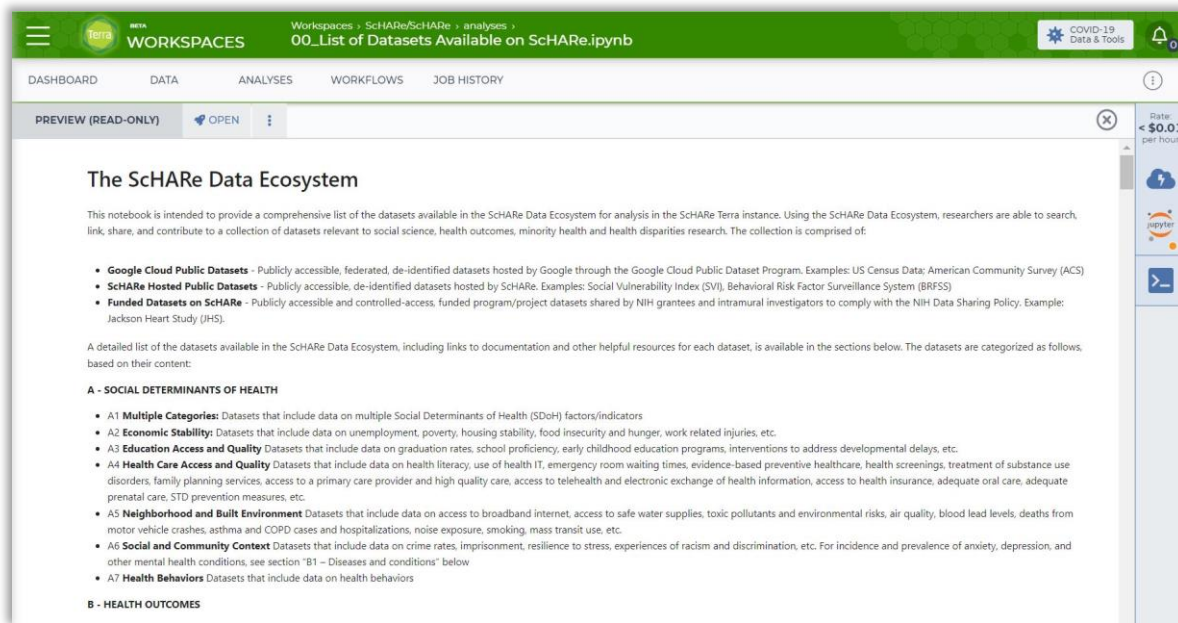
- haplopecaller-gvcf-gatk4
Runs HaplotypeCaller from GATK4 in GVCF mode on a single sample.
- mutect2-gatk4
Implements GATK4 Mutect 2 on a single tumor-normal pair.
- processing-for-variant-discovery-gatk4

Below the list is a 'Find Additional Workflows' section with a 'Dockstore' link: 'Browse WDL workflows in Dockstore, an open platform used by the GA4GH for sharing Docker-based workflows.'

- Modular codes developed for reuse
- **Adding SAS**

ScHARe Terra interface: access to datasets

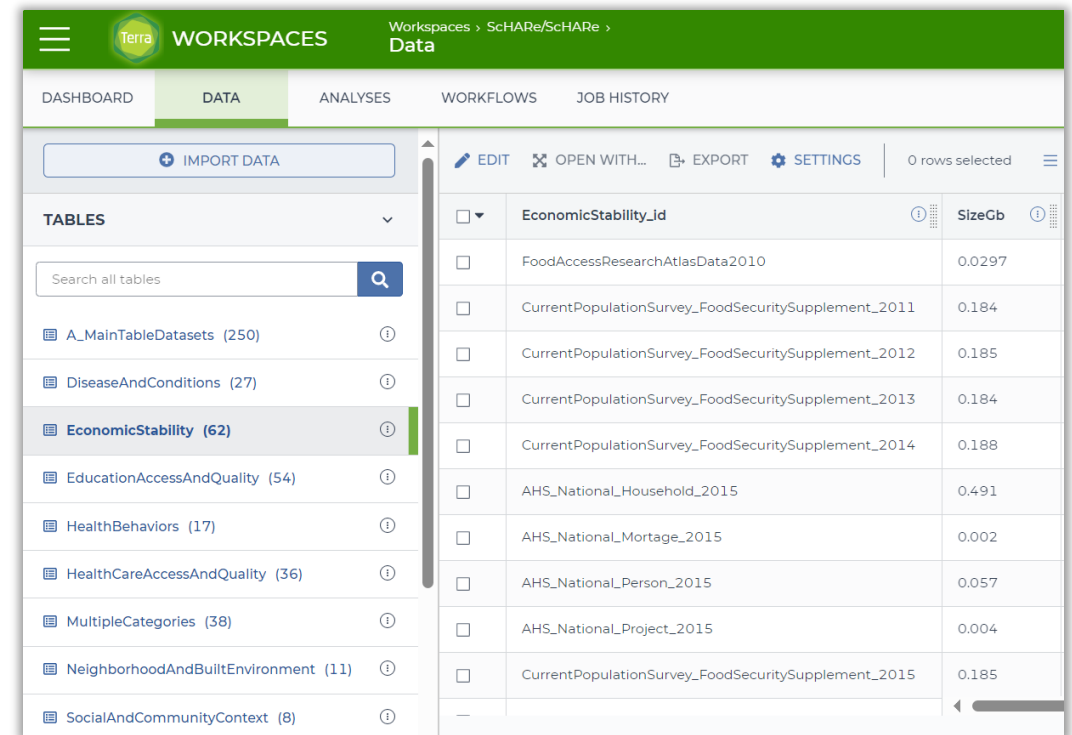
What data?



The screenshot shows the ScHARe Terra interface in the 'Analyses' tab. The notebook '00_List of Datasets Available on ScHARe.ipynb' is open, displaying a page titled 'The ScHARe Data Ecosystem'. The page contains a detailed list of datasets categorized into 'SOCIAL DETERMINANTS OF HEALTH' and 'HEALTH OUTCOMES'. The 'SOCIAL DETERMINANTS OF HEALTH' section includes categories like 'Multiple Categories', 'Economic Stability', 'Health Care Access and Quality', 'Neighborhood and Built Environment', 'Social and Community Context', and 'Health Behaviors'. The 'HEALTH OUTCOMES' section is also visible at the bottom.

In the **Analyses tab**, the notebook **00_List of Datasets Available on ScHARe** lists all datasets

Where?



The screenshot shows the ScHARe Terra interface in the 'Data' tab. The 'Data' tab is active, displaying a table of datasets. The table has columns for 'Tables' and 'SizeGb'. The 'EconomicStability' table is highlighted, showing 62 datasets. The table lists various datasets such as 'FoodAccessResearchAtlasData2010', 'CurrentPopulationSurvey_FoodSecuritySupplement_2011', 'CurrentPopulationSurvey_FoodSecuritySupplement_2012', 'CurrentPopulationSurvey_FoodSecuritySupplement_2013', 'CurrentPopulationSurvey_FoodSecuritySupplement_2014', 'AHS_National_Household_2015', 'AHS_National_Mortgage_2015', 'AHS_National_Person_2015', 'AHS_National_Project_2015', and 'CurrentPopulationSurvey_FoodSecuritySupplement_2015'.

In the **Data tab**, data tables help access data

ScHARe Ecosystem structure

Researchers can access, link, analyze, and export a **wealth of SDoH and population science related datasets** within and across platforms relevant to research about health disparities, health care delivery, health outcomes and bias mitigation, including:

250+
FEDERATED
PUBLIC
DATASETS

Public datasets

Publicly accessible, federated, de-identified datasets hosted by ScHARe or hosted by Google through the Google Cloud Public Dataset Program

ScHARe e.g.: *Behavioral Risk Factor Surveillance System (BRFSS)*
Google e.g.: *American Community Survey (ACS)*

CDE
FOCUSED
REPOSITORY

Funded datasets

Publicly accessible and controlled-access, funded program/project datasets using Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy

e.g.: *Jackson Heart Study (JHS)*
Extramural Grant Data
Intramural Project Data

Innovative Approach:
CDE Concept Codes
Uniform Resource Identifier (**URI**)

ScHARe Ecosystem

OVER 260 DATA SETS CENTRALIZED

Datasets are categorized by content based on the CDC **Social Determinants of Health** categories:

1. Economic Stability
2. Education Access and Quality
3. Health Care Access and Quality
4. Neighborhood and Built Environment
5. Social and Community Context

with the addition of:

- **Health Behaviors**
- **Diseases and Conditions**

The screenshot shows the Terra WORKSPACES Data interface. The top navigation bar includes 'DASHBOARD', 'DATA', 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. The 'DATA' tab is active, displaying an 'IMPORT DATA' button and a search bar for tables. A list of tables is shown on the left, with 'EconomicStability (62)' highlighted. The main table on the right lists datasets with columns for 'EconomicStability_id' and 'SizeGb'. The datasets listed include 'FoodAccessResearchAtlasData2010', 'CurrentPopulationSurvey_FoodSecuritySupplement_2011', 'CurrentPopulationSurvey_FoodSecuritySupplement_2012', 'CurrentPopulationSurvey_FoodSecuritySupplement_2013', 'CurrentPopulationSurvey_FoodSecuritySupplement_2014', 'AHS_National_Household_2015', 'AHS_National_Mortgage_2015', 'AHS_National_Person_2015', 'AHS_National_Project_2015', and 'CurrentPopulationSurvey_FoodSecuritySupplement_2015'.

	EconomicStability_id	SizeGb
<input type="checkbox"/>	FoodAccessResearchAtlasData2010	0.0297
<input type="checkbox"/>	CurrentPopulationSurvey_FoodSecuritySupplement_2011	0.184
<input type="checkbox"/>	CurrentPopulationSurvey_FoodSecuritySupplement_2012	0.185
<input type="checkbox"/>	CurrentPopulationSurvey_FoodSecuritySupplement_2013	0.184
<input type="checkbox"/>	CurrentPopulationSurvey_FoodSecuritySupplement_2014	0.188
<input type="checkbox"/>	AHS_National_Household_2015	0.491
<input type="checkbox"/>	AHS_National_Mortgage_2015	0.002
<input type="checkbox"/>	AHS_National_Person_2015	0.057
<input type="checkbox"/>	AHS_National_Project_2015	0.004
<input type="checkbox"/>	CurrentPopulationSurvey_FoodSecuritySupplement_2015	0.1



ScHARe Ecosystem: ScHARe hosted datasets

Organized based on the **CDC SDoH categories**, with the addition of *Health Behaviors and Diseases and Conditions*:

260+ datasets

- What are the Social Determinants of Health?

Social determinants of health (SDoH) are the **nonmedical factors that influence health outcomes**

They are the **conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life**



ScHARe Ecosystem: ScHARe hosted datasets

Education access and quality

Data on graduation rates, school proficiency, early childhood education programs, interventions to address developmental delays, etc.

Health care access and quality

Data on health literacy, use of health IT, preventive healthcare, access to health insurance, etc.

Neighborhood and built environment

Data on access to safe water supplies, toxic pollutants and environmental risks, air quality, blood lead levels, noise exposure, smoking, mass transit use, etc.

Social and community context

Data on crime rates, imprisonment, resilience to stress, experiences of racism and discrimination, etc.

Economic stability

Data on unemployment, poverty, housing stability, food insecurity and hunger, work related injuries, etc.

* Health behaviors

Data on health-related practices that can directly affect health outcomes.

* Diseases and conditions

Data on incidence and prevalence of specific diseases and health conditions.



** Not Social Determinants of Health*

ScHARe Ecosystem: Google hosted datasets

Examples of interesting datasets include:

- **American Community Survey** (U.S. Census Bureau)
- **US Census Data** (U.S. Census Bureau)
- **Area Deprivation Index** (BroadStreet)
- **GDP and Income by County** (Bureau of Economic Analysis)
- **US Inflation and Unemployment** (U.S. Bureau of Labor Statistics)
- **Quarterly Census of Employment and Wages** (U.S. Bureau of Labor Statistics)
- **Point-in-Time Homelessness Count** (U.S. Dept. of Housing and Urban Development)
- **Low Income Housing Tax Credit Program** (U.S. Dept. of Housing and Urban Development)
- **US Residential Real Estate Data** (House Canary)
- **Center for Medicare and Medicaid Services - Dual Enrollment** (U.S. Dept. of Health & Human Services)
- **Medicare** (U.S. Dept. of Health & Human Services)
- **Health Professional Shortage Areas** (U.S. Dept. of Health & Human Services)
- **CDC Births Data Summary** (Centers for Disease Control)
- **COVID-19 Data Repository by CSSE at JHU** (Johns Hopkins University)
- **COVID-19 Mobility Impact** (Geotab)
- **COVID-19 Open Data** (Google BigQuery Public Datasets Program)
- **COVID-19 Vaccination Access** (Google BigQuery Public Datasets Program)

How to access Google hosted datasets

Big Query

The Google public datasets are **available for access on Terra using BigQuery**

- BigQuery is the Google Cloud storage solution for structured data
- It is easy to use, works with large amounts of data and offers fast data retrieval and analysis
- **Our instructional notebooks in the Analyses tab** provide code and instructions on using Big Query to access Google datasets

```
Jupyter 06_How to access plot and save data from public BigQuery datasets using Python 3.ipynb
```

The following Python code will read a BigQuery table into a Pandas dataframe.

From <https://cloud.google.com/community/tutorials/bigquery-ibis>

ibis is a Python library for doing data analysis. It offers a Pandas-like environment for executing data analysis composable, and familiar replacement for SQL.

```
In [9]: # Connect to the dataset
conn = ibis.bigquery.connect(dataset_id='bigquery-public-data.broadstreet_adi')
```

```
In [10]: # Read table
ADI_table_2 = conn.table('area_deprivation_index_by_census_block_group')
ADI_table_2
```

```
Out[10]: BigQueryTable[table]
name: bigquery-public-data.broadstreet_adi.area_deprivation_index_by_census_block_group
schema:
  geo_id : string
  state_fips_code : string
  county_fips_code : string
  block_group_fips_code : string
  description : string
  county_name : string
  state_name : string
  state : string
  year : int64
  area_deprivation_index_percent : float64
```

ScHARe



The ScHARe Data Ecosystem

This document is intended to provide a comprehensive list of the datasets available in the ScHARe Data Ecosystem for analysis in the ScHARe Terra instance. Using the ScHARe Data Ecosystem, researchers are able to search, link, share, and contribute to a collection of datasets relevant to social science, health outcomes, minority health and health disparities research.

The collection is comprised of:

- **Google-hosted Public Datasets** - Publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program. Examples: US Census Data; American Community Survey (ACS)
- **ScHARe-hosted Public Datasets** - Publicly accessible, de-identified datasets hosted by ScHARe. Examples: Social Vulnerability Index (SVI), Behavioral Risk Factor Surveillance System (BRFSS)
- **ScHARe-hosted Project Datasets** - Publicly accessible and controlled-access, funded program/project datasets shared by NIH grantees and intramural investigators to comply with the

ScHARe
Datasets
PDF list



Scan me

bit.ly/ScHARe-datasets

CDE benefits:

- Faster start-up for project
- Better data aggregation across projects
- Shared meaning
- Concept-focused to allow questions/answers variations
- Coding enables an URI approach for better data interoperability

A **Common Data Element (CDE)** is a standardized, precisely defined question, paired with a set of allowable responses, used systematically across different sites, studies, or clinical trials to ensure consistent data collection

Because Researchers use CDEs...

they can more quickly share data and get results faster, which ultimately can help make a **meaningful difference to our nation's health.**



For more information about how CDEs accelerate research discoveries, visit: cde.nlm.nih.gov/resources

ScHARe Core CDEs

PhenX Toolkit

- Age
- Birthplace
- Zip Code
- Race and Ethnicity
- Sex
- Gender
- Sexual Orientation
- Marital Status
- Education
- Annual Household Income
- Household Size

- English Proficiency
- Disabilities
- Health Insurance
- Employment Status
- Usual Place of Health Care
- Financial Security / Social Needs
- Self-Reported Health
- Health Conditions (and Associated Medications/Treatments)

- **NIMHD Framework***
- **Health Disparity Outcomes***

* Project Level CDEs

NIH Endorsed



ScHARe has developed **Common Data Elements** to ensure consistent data collection across studies, facilitate interoperability, and link data from different sources

NIH CDE Repository:

cde.nlm.nih.gov/home

PhenX Toolkit:

www.nimhd.nih.gov/resources/phenx/

COMMON DATA ELEMENTS

NLM CDE Repository
Coded NIMHD Common Data Elements

- Labels
- Questions
- Permissible Values

A
T
O

Common Data Elements + Data

Data Access
Based On PII Levels and User Needs:

- Public
- Data Use Agreement
- Private

DATA UPLOAD

Acquired Google and ScHARe Hosted Datasets

Overview

Data Dictionaries

Data Updates

ScHARe REPOSITORY

Project and Key Acquired Datasets

Overview

Description and Links to Overview Material

4-Privacy Levels

COMMON DATA ELEMENTS

Data

Metadata

Data Dictionaries

Analysis Ready

RAS Single Sign-on

DATA MAPPING, DOWNLOAD AND EXPORT

DATA MAPPING
ACROSS DATASETS AND PLATFORMS
BASED ON CDES

Other Cloud Platforms
AnVil, BDC, All of Us

EXAMPLE: CDE linked
ACS NIMHD Project BioData Catalyst
Aggregated Data Set

CDE Linked Project Data

Data Download in a Variety of Formats
CSV, TSV, XLSX

Data Export to Terra for Analysis
Workspaces

Visualizations Tools
Shiny



ScHARe Repository

PUBLICLY AVAILABLE FALL 2024

The screenshot shows the 'Create New Collection' form in the ScHARe Repository. The form is titled 'Create New Collection' and is located in the center of the page. It has a dark blue header with the 'Pigeon' logo and navigation links for 'About', 'Docs', 'Community', and 'Collections'. A search bar is also present in the header. The form itself is white and contains the following fields:

- NAME:** A single-line text input field.
- DESCRIPTION:** A large multi-line text area.
- METADATA:** A section with a help icon (i) containing two input fields labeled 'key' and 'value', and a plus sign (+) button to add more metadata.
- Submit:** A button at the bottom of the form.

The left sidebar of the application is dark blue and contains navigation options: 'Recent', 'My Collections', and 'Starred', each with a right-pointing arrow.

- Host your project data in a **safe space** with privacy levels, secure workspaces, collaboration platform
- **CDE centric**
- **Focus:** Social Science, SDoH, Health Disparities, Health Outcomes Research
- Comply with **NIH Data Management and Data Sharing Policy**
- **Link your data** with others and federated data

The screenshot shows the ScHARe Repository interface. At the top, there is a navigation bar with 'About', 'Resources', and 'Data' buttons, a search bar, and a user profile 'AB'. Below this is a sidebar with 'Create a Collection' and 'Most Recent' collections. The main content area is titled 'pigeon@localhost / Collection Path' and features a 'CDE Configuration' section. This section includes a 'Choose a data standard' dropdown menu set to 'ScHARe', 'Save' and 'Cancel' buttons, and a table mapping data elements to file columns. Below the table is a 'Status' section showing '7/22 CDEs assigned' and '0 validation errors', with a list of data elements categorized by status.

Home Page

About Resources Data search AB

+ Create a Collection

Most Recent

- Example Collection 1
- Mouseover Collection
- Example Collection 2

Your Collections

- My Collection 1
- My Collection 2
- My Collection 3

pigeon@localhost / Collection Path Admin Star 10.1k

CDE Configuration

Assign your data elements to relevant data standards like ScHARe at scale to enable more powerful analysis. Hold tab when selecting to assign multiple files or columns at once.

Choose a data standard
ScHARe

Save Cancel

File	Common Data Element	Column Name	Data Type
file2.csv	Sex	Client Age	integer
exampleTab.xlsx	Age	Smoker	
	Education Level	College	

Status data available 7/22 CDEs assigned 0 validation errors

✓ Address Age Education Health Insurance Orientation Sex Zipcode

✗ Annual Income Birthplace Disabilities Disease Disorders Education Employment English Proficiency Household Size Marital Status Medical Treatment Self-Reported Health Social Needs Usual Place of Care

Map project CDEs or variables to ScHARe-PhenX CDEs

ScHARe Repository

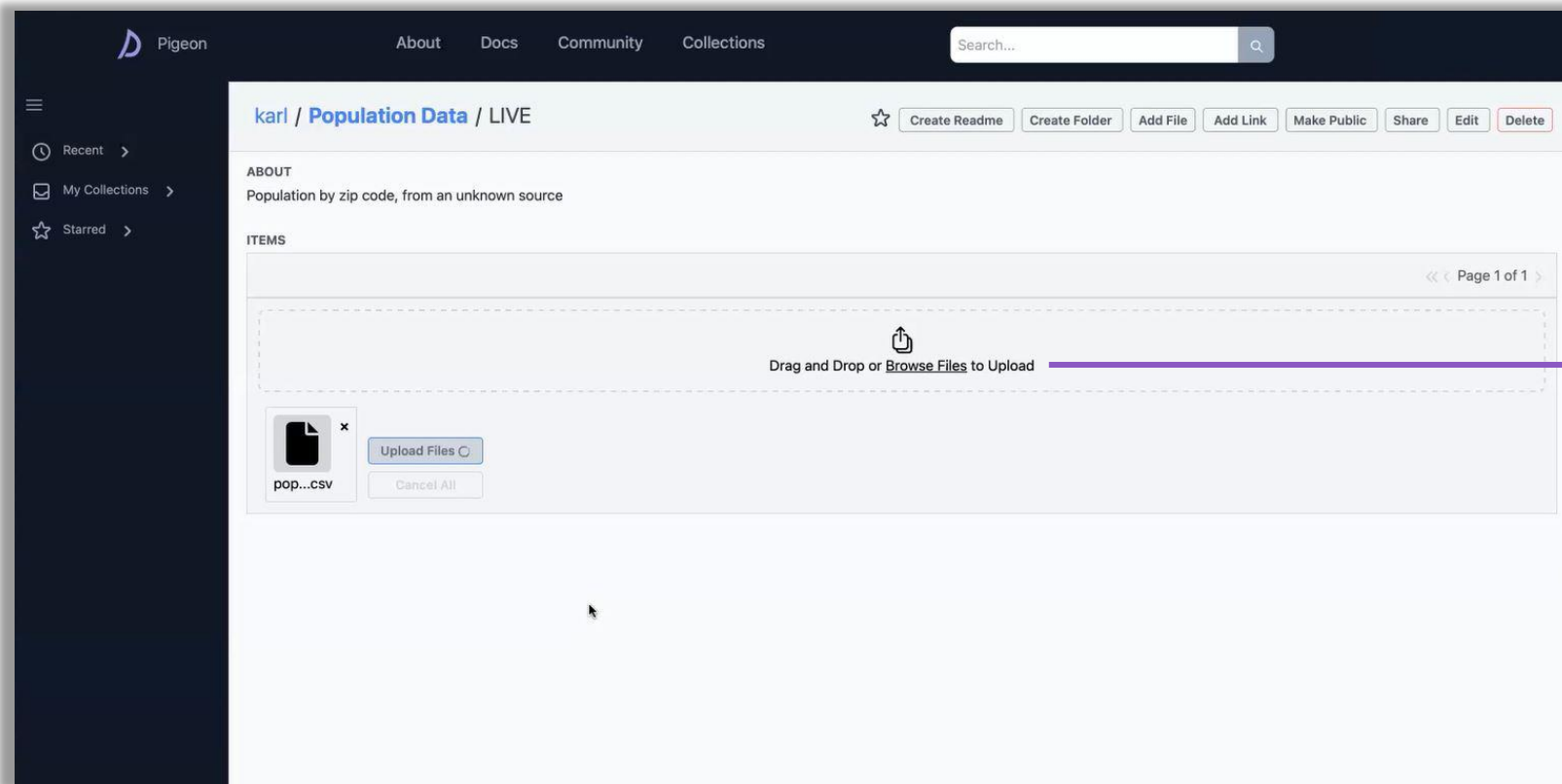
PUBLICLY AVAILABLE FALL 2024

The screenshot displays the ScHARe Repository web interface. At the top, there is a navigation bar with 'About', 'Resources', and 'Data' buttons, a search bar, and a user profile icon labeled 'AB'. The main content area shows a collection page for 'pigeon@localhost / Collection Path'. The collection is titled 'Big_Test Collection' and has a description: 'Description text and stuff. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, ullamco laboris nisi ut commodo consequat.' The collection has 10.1k stars and is set to 'Restricted Access' (Privacy Level) and 'Ready' (Analysis Readiness). It also shows '7/22 CDEs present in this collection' under 'ScHARe CDE Compliance'. The left sidebar contains a 'Create a Collection' button and lists 'Most Recent' and 'Your Collections'.

Shows number of project CDEs that match or can map to ScHARe-PhenX CDEs

ScHARe Repository

PUBLICLY AVAILABLE FALL 2024



Aggregate datasets with drag-and-drop features

ScHARe Repository

PUBLICLY AVAILABLE FALL 2024

The screenshot displays the ScHARe Repository interface for configuring a data parser. The main content area is titled 'karl / Population Data / LIVE / population_by_zip_2010.csv'. The 'Parser Type' is set to 'csv'. The 'Columns' table lists the following columns:

Column Name	Type	Value
minimum_age	Integer	
maximum_age	Integer	
gender	String	Gender fMCdaD9i:0001
zipcode	String	nhcde:7kijL9i3sx
geo_id	String	

The 'Results' section shows: Data available (checked), 0 parsing errors (checked), and 5 validation errors (unchecked). The 'Table Preview' section shows the following data:

population	minimum_age	maximum_age	gender	zipcode	geo_id
50	30	34	female	61747	8600000US61747
5	85		male	64120	8600000US64120
1389	30	34	male	95117	8600000US95117
231	60	61	female	74074	8600000US74074
56	0	4	female	58042	8600000US58042

View
aggregated
dataset



ScHARe

Research Think-a-Thons

- Novice **training webinars** for data science, cloud computing and research using Big Data
- **Target:** underrepresented populations, women, racial/ethnic and sexual gender minorities, rural and poor populations

Generational career & discipline exchange



Think-a-Thons

Goals:

- Upskill underrepresented populations in data science and cloud computing
- Foster a research paradigm shift to use Big Data in health disparities/health outcomes research
- Promote use of Dark Data



1. TUTORIAL AND TARGETED THINK-A-THONS

- Monthly sessions (2 1/2 hours)
- Instructional/interactive
- Designed for new/experienced users
- Networking
- Mentoring and coaching
- Topics include:
 - Data Science 101
 - Terra
 - Social Determinants of Health analytics
 - Common Data Elements
 - AI readiness
 - Ethical and transparent AI
 - Bias mitigation



2. RESEARCH THINK-A-THONS

- Multi-career (students to senior investigators)
- Multi-discipline (data scientists and researchers)
- Featured datasets with guest experts leads
- Guest experts in topic areas, analytics, data sources etc. to provide guidance
- Generate research idea - decide design, datasets and analytics
- Learn Ethical AI
- Publications

Register:
bit.ly/think-a-thons



Think-a-Thon tutorials

bit.ly/think-a-thons

SPECIAL EVENTS

February

Artificial Intelligence and Cloud Computing 101

March

ScHARe 1 – Accounts and Workspaces

April

ScHARe 2 – Terra Datasets

May

ScHARe 3 – Terra Google-hosted Datasets

June

ScHARe 4 – Terra ScHARe-hosted Datasets

July

An Introduction to Python for Data Science – Part 1

August

An Introduction to Python for Data Science – Part 2

September

ScHARe 5: A Review of the ScHARe Platform and Data Ecosystem

October

Preparing for AI 1: Common Data Elements and Data Aggregation

November

Preparing for AI 2: An Introduction to FAIR Data and AI-ready Datasets

January

Preparing for AI 3: Computational Data Science Strategies 101

February/March

Preparing for AI 4: Overview Prep for AI Summary with Transparency, Privacy, Ethics

April

Research Teams – SDoH and Health Disparities

May

Be a Part of the Future of Knowledge Generation 1: AI/Cloud Computing Basics and CDEs

July

Be a Part of the Future of Knowledge Generation 2: AI-Ready Datasets and Computations

- ScHARe for **Educators** (Community Colleges and low-resource MSIs)
- ScHARe for **American Indian/Alaska Native Researchers**
- ScHARe for **Coders and Programmers** to conduct research



Experience conducting ethical AI

Transparency

Public perception and understanding of how AI works

- **Technical documentation for duplication/re-use**
- **Tools:**
 - Data dictionary
 - Health sheet (Data sheet)
 - Model cards (capabilities and purpose of algorithms are openly and clearly communicated to relevant stakeholders)

Fairness

F indable: *providing metadata, documentation, and clear identifiers*

A ccessible: *wide audience*

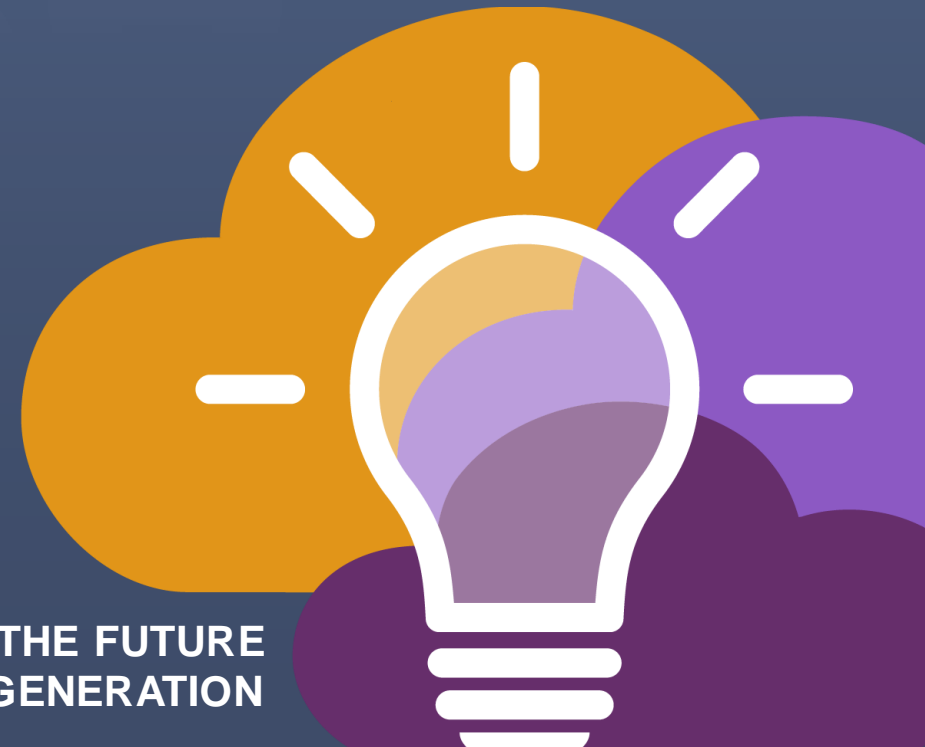
I nteroperable: *standardized formats and APIs enable seamless integration*

R eusable: *clear documentation, licensing, reduce redundancy*

- Metadata and data should be **easy to find** for both humans and computers
- Ensure that **data represents** relevant populations

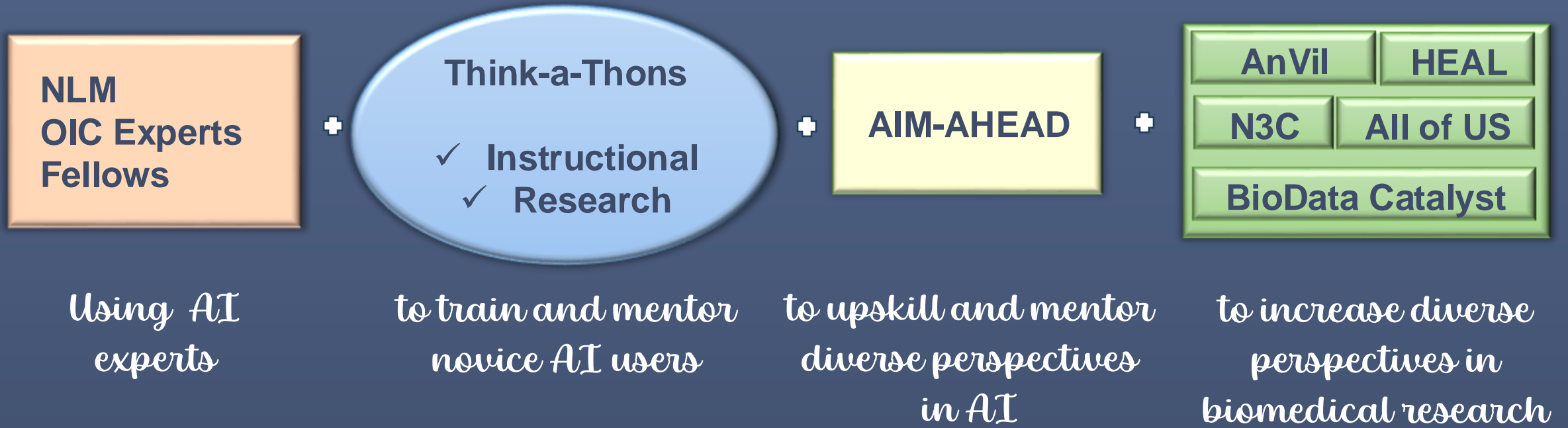
SCHARe

Training
pipeline



BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

Think-a-Thons training/mentoring pipeline



Goal: “Upskilling”

- ✓ Data science specialists into health disparities and health outcomes research
- ✓ Health disparities/outcomes researchers into using big data and cloud computing

Target Audience:

- ✓ Underrepresented populations (women, race/ethnic) users not trained in data science
- ✓ Data scientists with no or little research experience
- ✓ Resource and tool for Community Colleges and low-resource MSIs and organizations

Join AIM-AHEAD Connect



- AIM-AHEAD's community, networking, mentoring, and career development platform
- Virtual space to engage with the entire AIM-AHEAD Consortium and build community!
- Custom tools available to the AIM-AHEAD Coordinating Center:
 - Connect with experts, learners, stakeholders, etc.
 - Mentoring, Q&A, video calls, groups, funding & jobs board, etc.
 - SignUp: Event registration & information solicitation
 - Surveys: Request feedback on various activities
 - HelpDesk: Respond to topic-specific questions
 - Programs: Collaborative space, exclusive content, and mentor matching

**Scan QR Code
to Join
AIM-AHEAD Connect**

