



NIH Endorsed
Common
Data Elements

ScHARe

Science Collaborative for Health disparities
and Artificial intelligence bias REduction



National Institute on Minority Health and Health Disparity

Think-a-Thons





NIMHD

Dr. Eliseo Perez-Stable

ODSS

Dr. Susan Gregurick

NIH/OD

Dr. Larry Tabak

NINR

Dr. Shannon Zenk

NINR

Rebecca Hawes
Micheal Steele
John Grason

NIDCR

ORWH

OMH

NIMHD OCPL

Kelli Carrington
Thoko Kachipande
Corinne Baker

BioTeam

STRIDES

Terra

SIDEM

RLA

Broad Institute

CDE Working Group

Deborah Duran
Luca Calzoni
Rebecca Hawes
Micheal Steele
Kelvin Choi
Paula Strassle
Deborah Linares
Crystal Barksdale
Gneisha Dinwiddie
Jennifer Alvidrez
Matthew McAuliffe
Carolina Mendoza-Puccini
Simrann Sidhu
Tu Le



BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION



What is ScHARe?



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....)
- Cut and Paste code in Python and R



Terra recommends using Chrome
Must have a gmail friendly acct

PREPARING FOR AI – RESEARCH AND HEALTH CARE USING BIG DATA

Mapping across cloud platforms with Terra Interface



BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

Data Ecosystem Structure Population Science/SDoH

**FEDERATED
PUBLIC DATA
240+**
(Population Science/
SDoH / Behavioral)

Hosted by Google
& by ScHARe

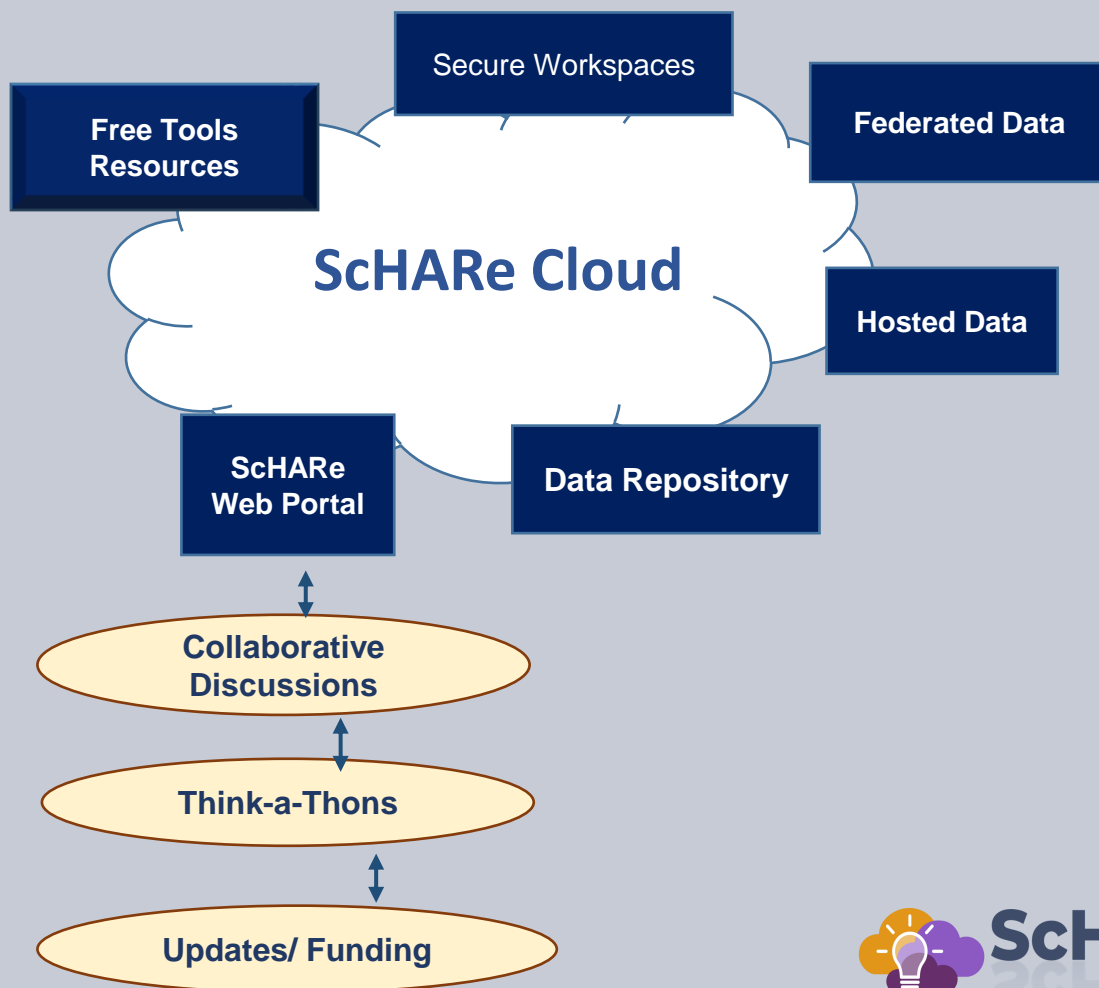
REPOSITORY
CDE FOCUSED

CDEs enhances Data
Interoperability
(Aggregation) by using
semantic standards
and concept codes

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

Components Terra Interface

Intramural & Extramural Resource



ScHARe Data Ecosystem

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

- **Google Cloud Public Datasets:** publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program

Example: *American Community Survey (ACS)*

- **ScHARe Hosted Public Datasets:** publicly accessible, de-identified datasets hosted by ScHARe

Example: *Behavioral Risk Factor Surveillance System (BRFSS)*

- **Funded Datasets on ScHARe:** publicly accessible and controlled-access, funded program/project datasets using Core Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy

Examples: *Jackson Heart Study (JHS); Extramural Grant Data; Intramural Project Data*



OVER 240 DATA SETS CENTRALIZED

The screenshot shows a web application interface for data management. The top navigation bar includes 'WORKSPACES', 'Data', and 'COVID-19 Data & Tools'. Below the navigation, there are tabs for 'DASHBOARD', 'DATA', 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. The main content area displays a table of datasets with columns for 'A_MainTableDatasets.Id', 'Categories', 'Year', 'Data', and 'DataDictionary'. The table lists various datasets, including 'AdjustedGraduationRate_2010-2011' through 'AdjustedGraduationRate_2018-2019', 'BRFSS_PhoneSurvey_2012', and 'BRFSS_PhoneSurvey_2013'. A yellow bracket highlights the 'A_MainTable...' section on the left sidebar.

A_MainTableDatasets.Id	Categories	Year	Data	DataDictionary
AdjustedGraduationRate_2010-2011	Education Access and Quality	2010-2011	acgr-lea-sy2010-11.csv	acgr-sy10-11-public
AdjustedGraduationRate_2011-2012	Education Access and Quality	2011-2012	acgr-lea-sy2011-12.csv	acgr-sy11-12-public
AdjustedGraduationRate_2012-2013	Education Access and Quality	2012-2013	acgr-lea-sy2012-13.csv	acgr-sy12-13-public
AdjustedGraduationRate_2013-2014	Education Access and Quality	2013-2014	acgr-lea-sy2013-14.csv	acgr-sy13-14-public
AdjustedGraduationRate_2014-2015	Education Access and Quality	2014-2015	acgr-release2-lea-sy2014-15.c	acgr-release2-sy201
AdjustedGraduationRate_2015-2016	Education Access and Quality	2015-2016	acgr-lea-sy2015-16.csv	acgr-sy2015-16-pub
AdjustedGraduationRate_2016-2017	Education Access and Quality	2016-2017	acgr-lea-sy2016-17.csv	acgr-sy2016-17-pub
AdjustedGraduationRate_2017-2018	Education Access and Quality	2017-2018	acgr-lea-sy2017-18.csv	acgr-sy2017-18-pub
AdjustedGraduationRate_2018-2019	Education Access and Quality	2018-2019	acgr-lea-sy2018-19-long.csv	acgr-sy2018-19-pub
BRFSS_PhoneSurvey_2012	Health Behaviors	2012	LLCP2012.XPT	CODEBOOK12_ILLCF
BRFSS_PhoneSurvey_2013				

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**

Users will be able to **map and link** across datasets

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)

ScHARe Ecosystem: ScHARe hosted datasets

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

240+ 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**.



www.cdc.gov/about/sdoh/index.html

ScHARe **Ecosystem**: ScHARe hosted datasets

Examples of datasets for each category include:

Education access and quality

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

Examples:

- **EDFacts Data Files** (U.S. Dept. of Education) - Graduation rates and participation/proficiency assessment
- **NHES - National Household Education Surveys Program** (U.S. Dept. of Education) – Educational activities

ScHARe **Ecosystem**: ScHARe hosted datasets

Health care access and quality

Data on health literacy, use of health IT, emergency room waiting times, preventive healthcare, health screenings, treatment of substance use disorders, family planning services, access to a primary care provider and high quality care, access to telehealth and electronic exchange of health information, access to health insurance, adequate oral care, adequate prenatal care, STD prevention measures, etc.

Example:

- **MEPS - Medical Expenditure Panel Survey** (AHRQ) - Cost and use of healthcare and health insurance coverage
- **Dartmouth Atlas Data** - Selected Primary Care Access and Quality Measures - Measures of primary care utilization, quality of care for diabetes, mammography, leg amputation and preventable hospitalizations

ScHARe **Ecosystem**: ScHARe hosted datasets

Neighborhood and built environment

Data on access to broadband internet, access to safe water supplies, toxic pollutants and environmental risks, air quality, blood lead levels, deaths from motor vehicle crashes, asthma and COPD cases and hospitalizations, noise exposure, smoking, mass transit use, etc.

Examples:

- **National Environmental Public Health Tracking Network** (CDC) - Environmental indicators and health, exposure, and hazard data
- **LATCH - Local Area Transportation Characteristics for Households** (U.S. Dept. of Transportation) – Local transportation characteristics for households

ScHARe **Ecosystem**: ScHARe hosted datasets

Social and community context

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

Example:

- **Hate crime statistics** (FBI) - Data on crimes motivated by bias against race, gender identity, religion, disability, sexual orientation, or ethnicity
- **General Social Survey** (GSS) - Data on a wide range of characteristics, attitudes, and behaviors of Americans.

ScHARe **Ecosystem**: ScHARe hosted datasets

Economic stability

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

Examples:

- **Current Population Survey (CPS) Annual Social and Economic Supplement** (U.S. Bureau of Labor Statistics) - Labor force statistics: annual work activity, income, health insurance, and health
- **Food Access Research Atlas** (U.S. Dept. of Agriculture) – Food access indicators for low-income and other census tracts

ScHARe **Ecosystem**: ScHARe hosted datasets

Health behaviors

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

Examples:

- **BRFSS - Behavioral Risk Factor Surveillance System** (CDC) - State-level data on health-related risk behaviors, chronic health conditions, and use of preventive services
- **YRBSS - Youth Risk Behavior Surveillance System** (CDC) – Health behaviors that contribute to the leading causes of death, disability, and social problems among youth and adults

ScHARe **Ecosystem**: ScHARe hosted datasets

Diseases and conditions

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

Examples:

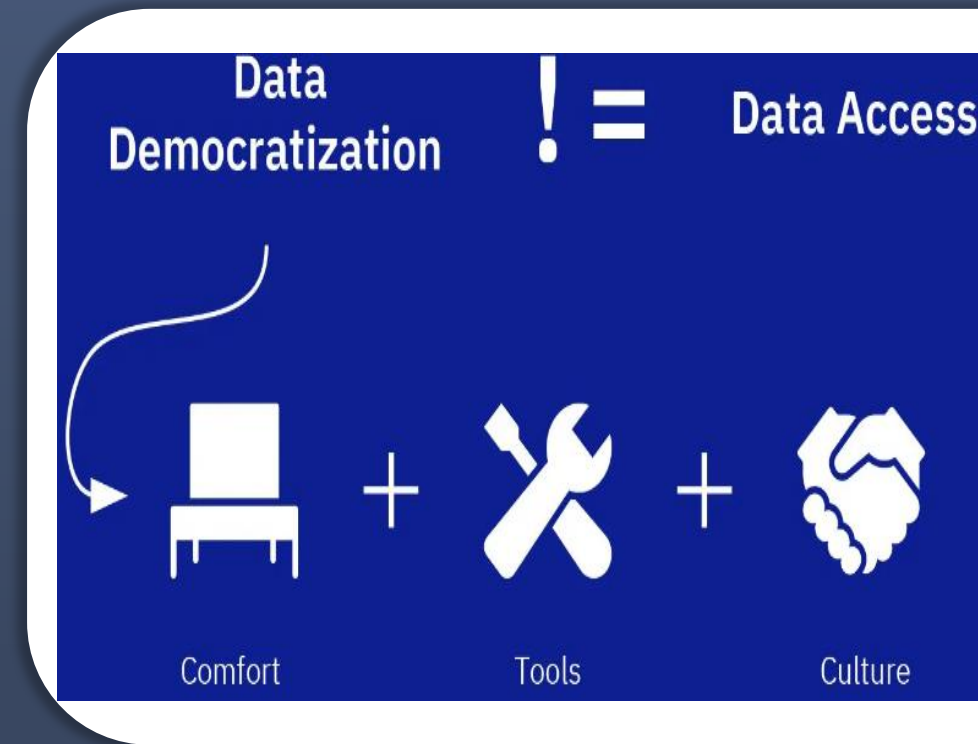
- **U.S. CDI - Chronic Disease Indicators** (CDC) - 124 chronic disease indicators important to public health practice
- **UNOS - United Network of Organ Sharing** (Health Resources and Services Administration) – Organ transplantation: cadaveric and living donor characteristics, survival rates, waiting lists and organ disposition



Data Democratization

Making digital information accessible to the average non-technical user

- Equal access (expert to novice)
- Pool Tooling - Centralized resource (tools, data, documentation, etc)
- Security framework
- Standardize & automate to reduce barriers to success
- Eye on Ethical AI





Terra Interface: Data Sets and Access to Data

Analyses Tab in ScHARe workspace, the notebook 00_List of Datasets Available on ScHARe lists all of the datasets available in the ScHARe Datasets collection

What?

The ScHARe Data Ecosystem

This notebook 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 Cloud 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)
- **Funded Datasets on ScHARe** - Publicly accessible and controlled-access, funded program/project datasets shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy. Example: Jackson Heart Study (JHS).

A detailed list of the datasets available in the ScHARe Data Ecosystem, including links to documentation and other helpful resources for each dataset, is available in the sections below. The datasets are categorized as follows, based on their content:

A - SOCIAL DETERMINANTS OF HEALTH

- **A1 Multiple Categories:** Datasets that include data on multiple Social Determinants of Health (SDoH) factors/indicators
- **A2 Economic Stability:** Datasets that include data on unemployment, poverty, housing stability, food insecurity and hunger, work related injuries, etc.
- **A3 Education Access and Quality:** Datasets that include data on graduation rates, school proficiency, early childhood education programs, interventions to address developmental delays, etc.
- **A4 Health Care Access and Quality:** Datasets that include data on health literacy, use of health IT, emergency room waiting times, evidence-based preventive healthcare, health screenings, treatment of substance use disorders, family planning services, access to a primary care provider and high quality care, access to telehealth and electronic exchange of health information, access to health insurance, adequate oral care, adequate prenatal care, STD prevention measures, etc.
- **A5 Neighborhood and Built Environment:** Datasets that include data on access to broadband internet, access to safe water supplies, toxic pollutants and environmental risks, air quality, blood lead levels, deaths from motor vehicle crashes, asthma and COPD cases and hospitalizations, noise exposure, smoking, mass transit use, etc.
- **A6 Social and Community Context:** Datasets that include data on crime rates, imprisonment, resilience to stress, experiences of racism and discrimination, etc. For incidence and prevalence of anxiety, depression, and other mental health conditions, see section "B1 - Diseases and conditions" below
- **A7 Health Behaviors:** Datasets that include data on health behaviors

B - HEALTH OUTCOMES

Data Tab in ScHARe workspace, search data and keep track of your project data:

- ScHARe workspace, click on the Data tab
- Under Tables, see a list of dataset categories
- Click on a category, to see a list of relevant datasets
- Scroll to the right to learn more about each dataset

Where and Why?

TABLES

Search all tables

- A_MainTableDatasets (118)
- DiseaseAndConditions (1)
- EconomicStability (30)
- EducationAccessAndQuality (47)**
- HealthCareAccessAndQuality (10)
- MultipleCategories (15)
- NeighborhoodAndBuiltEnvironment (10)
- SocialAndCommunityContext (4)

	EducationAccessAndQuality_id	Categories
<input type="checkbox"/>	AdjustedGraduationRate_2010-2011	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2011-2012	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2012-2013	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2013-2014	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2014-2015	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2015-2016	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2016-2017	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2017-2018	Education Access and Quality
<input type="checkbox"/>	AdjustedGraduationRate_2018-2019	Education Access and Quality
<input type="checkbox"/>	ECPP_EarlyChildhoodProgramParticip...	Education Access and Quality
<input type="checkbox"/>	ECPP_EarlyChildhoodProgramParticip...	Education Access and Quality
<input type="checkbox"/>	MathematicsAssessments_LocalEduc...	



Terra Interface: Secure workspace

The screenshot displays the Terra Workspaces interface. The main header shows 'Terra BETA WORKSPACES'. Below the header, there's a 'Workspaces +' section with a description: 'Dedicated spaces for you and your collaborators to access and analyze data'. A 'Recently Viewed' section lists two workspaces: 'ScHARe' (viewed Apr 14, 2023, 11:58 AM) and 'ScHARe Think-a-Thons' (viewed Apr 10). A search bar and 'Tags' are also visible. Below this, there are tabs for 'MY WORKSPACES (42)', 'NEW AND INTERESTING (6)', and 'FEATURED (6)'. The 'MY WORKSPACES' tab is active, showing a list of workspaces with columns for 'Name', 'ScHARe', and 'ScHARe Think-a-Thons'. A 'Share Workspace' modal dialog is open in the foreground, showing the following details:

- Share Workspace**
- User email:** A text input field with the placeholder 'Add people or groups' and an 'ADD' button.
- Current Collaborators:**
 - calzonii2@nih.gov:** Role: Owner. Permissions: Can share, Can compute.
 - ScHARe-Contractors@firecloud.org:** Role: Writer. Permissions: Can share, Can compute. Includes a close button (X).
 - ScHARe-Read-Only-Access@firecloud.org:** Role: Reader. Permissions: Can share, Can compute. Includes a close button (X).
- Share with Support:** A toggle switch set to 'No'.
- Buttons:** 'CANCEL' and 'SAVE'.

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



Terra Interface: Notebooks for Analytics & Tutorials

Workflows Modular codes

A notebook integrates code and its output into a single document where you can run code, display the output, and also add explanations, formulas, and charts

Easy to Use--Cut and Paste Analytics

The screenshot shows the Terra interface with the 'ANALYSES' tab selected. The page title is 'Workspaces > SchARE/SchARE > Analyses'. The navigation bar includes 'DASHBOARD', 'DATA', 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. Under 'Your Analyses', there is a '+ START' button and 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

The screenshot shows the Terra interface with the 'WORKFLOWS' tab selected. The page title is 'Workspaces > SchARE/SchARE > Workflows'. The navigation bar includes 'DASHBOARD', 'DATA', and 'ANALYSES'. Under 'WORKFLOWS', there is a 'Find a Workflow' search box and a '+ START' button. A 'Suggested Workflows' panel is open, showing three workflow cards:

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

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



Repository CDE Focused

CDE Validation Tool

Assign CDEs to Uploaded Data Columns

Assign your uploaded item's data columns to their appropriate common data elements.

[More Info >](#)




CDE set selected: **ScHARe**

CDE	UPLOADED DATA COLUMNS	DATA VALID
<input type="checkbox"/> Age	age ◀	✓
<input type="checkbox"/> Age Units	age_units ◀	✓
<input type="checkbox"/> Annual Household Income Range	annual_household_incor	✓
<input type="checkbox"/> Birthplace - Country Outside US		
<input type="checkbox"/> Birthplace - US	birthplace_us ◀	✓
<input type="checkbox"/> Disabilities	disabilities_1 ◀	✗
<input type="checkbox"/> Economic Stability – Social Needs		
<input type="checkbox"/> Economic Stability – Social Needs - Specify Other	economic_stability_soci	✗
<input type="checkbox"/> Education	education ◀	✓

Assign

More Options

Cancel




ScHARe

Repository

CDE Focus

CDE
Project
Matching
Mapping

Item Name	Common Data Element	Data Column
 Sample_ScHARe_data_-_Sheet1.csv	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Age <input type="checkbox"/> Age Units <input type="checkbox"/> Annual Household Income Range <input checked="" type="checkbox"/> Birthplace - Country Outside US <input checked="" type="checkbox"/> Birthplace - US <input checked="" type="checkbox"/> Disabilities <input checked="" type="checkbox"/> Deafness or difficulty hearing <input checked="" type="checkbox"/> Blindness or difficulty seeing <input checked="" type="checkbox"/> Difficulty concentrating,... <input checked="" type="checkbox"/> Difficulty walking or climbing stairs <input checked="" type="checkbox"/> Difficulty dressing or bathing 	<ul style="list-style-type: none"> record_id <input checked="" type="checkbox"/> age (Age) <input checked="" type="checkbox"/> age_units (Age Units) <input checked="" type="checkbox"/> annual_household_income (Annual... <input checked="" type="checkbox"/> birthplace_outside_us (Birthplace -... <input checked="" type="checkbox"/> birthplace_us (Birthplace - US) <input checked="" type="checkbox"/> disabilities_1 (Disabilities - Deafness ... <input checked="" type="checkbox"/> disabilities_2 (Disabilities - Blindness... <input checked="" type="checkbox"/> disabilities_3 (Disabilities - Difficulty... <input checked="" type="checkbox"/> disabilities_4 (Disabilities - Difficulty... <input checked="" type="checkbox"/> disabilities_5 (Disabilities - Difficulty...

Status - Sample_ScHARe_data_-_Sheet1.csv
 data available 15/30 CDEs assigned 0 validation errors

<p><input checked="" type="checkbox"/> Assigned CDEs</p> <ul style="list-style-type: none"> Age Age Units Annual Household Income Range Birthplace - Country Outside US Birthplace - US Disabilities Economic Stability – Social Needs - Specify Other Education 	<p><input checked="" type="checkbox"/> Unassigned CDEs</p> <ul style="list-style-type: none"> Economic Stability – Social Needs Employment Status Employment Status - Specify Gender - Select Other Gender - Specify Health Conditions - Disease Disorders Health Conditions - Disease Disorders - Specify Other Health Conditions - Medication or Other 	<p><input checked="" type="checkbox"/> Validation Errors</p> <p>There are no validation errors.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------



Project Overview

Project CA323887

Abstract
This project collects data from ... and ...

> Links and Documents

> Metadata

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	Sample_ScHARe_data_-_Sheet1...	3 minutes ago	43 KB	⋮

<< Page 1 of 1 >>

Privacy Level

Private

Analysis Readiness

Ready

CDE Compliance - ScHARe

15 / 30 CDEs assigned

*DOIs for project and aggregated data sets



Project Data Table

karl / [Project CA323887](#) / LIVE / Sample_ScHARe_data_-_Sheet1.csv

File [Table](#) Dictionary Meta 43 KB | 10 minutes ago | status:

Item Operations ▾

record_id	age	age_units	annual_househo...	birthplace_outsi...	birthplace_us	disabilities_1	disabilities_2
111111	41	Years	\$25,000-\$34,999		MA	0	0
222222	89	Years	\$35,000-\$49,999		CA	1	0
333333	23	Years	\$10,000-\$24,999		TX	0	0
444444	33	Years	\$50,000-\$74,999		MO	1	0



Repository Tool

Aggregating Data Sets Tool

karl / [Project CA88372](#) / LIVE / test-join-Schare-MHSVI

Advanced [Explorer](#) Table Dictionary Meta 4 KB | 4 hours ago | text/prql | status: Item Operations ▾

Select

Select

record_id
age
annual_household_income
zip_code

Pigeon.Join_select

Join Table	Column	Join Column
<input type="text" value="mh_svi_county-SCHARe"/>	<input type="text" value="zip_code"/>	<input type="text" value="zip_code"/>

+ Select + Filter + Sort + Join + Take + Rename + Map Into Clear All Save Dataview



ScHARe

Repository Tool

Aggregating Data Sets Tool

Pigeon.Join_select

Join Table: mh_svi_county-ScHARe Column: zip_code Join Column: zip_code

+ Select + Filter + Sort + Join + Take + Rename + Map Into **Clear All** **Save Dataview**

Results > ✓ Data available ✓ 0 parsing errors ✓ 0 validation errors

Transformation Preview

record_id	age	annual_househo...	zip_code	OID_	ST	STATE	ST_ABBR	COUNTY	FIPS	LOCATION	E_T
333333	23	\$10,000-\$24,999	29330	2358	45	SOUTH CAROLINA	SC	Spartanburg	45083	Spartanburg County, South Carolina	302
666666	45	\$50,000-\$74,999	98374	2980	53	WASHINGTON	WA	Pierce	53053	Pierce County, Washington	859
123450	65	\$35,000-\$49,999	72937	177	5	ARKANSAS	AR	Sebastian	5131	Sebastian County, Arkansas	127
234569	21	\$50,000-\$74,999	23153	2857	51	VIRGINIA	VA	Goochland	51075	Goochland County, Virginia	224
345670	23	\$10,000-\$24,999	29330	2358	45	SOUTH CAROLINA	SC	Spartanburg	45083	Spartanburg County, South Carolina	302

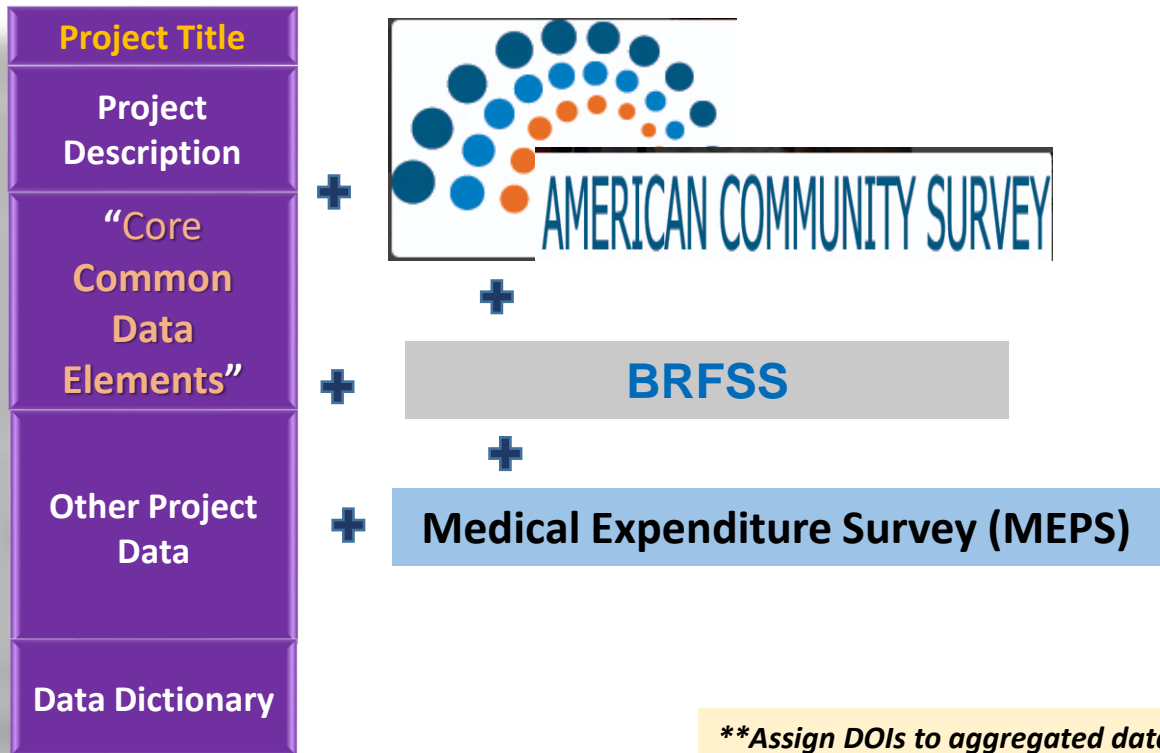


ScHARe

PREPARING FOR CLOUD COMPUTING AI – HEALTH DISPARITY AND CARE RESEARCH USING BIG DATA

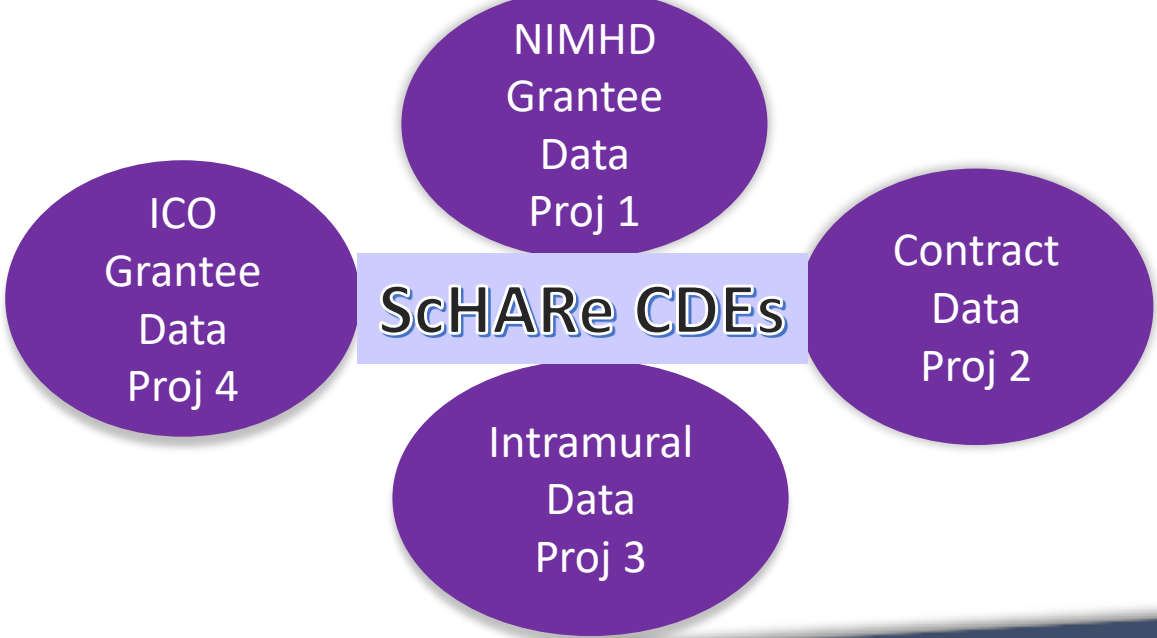
AGGRAGATING DATA SETS TOOL: Variables & CDEs

CDE Mapping Project & Federated Data



****Assign DOIs to aggregated data sets**

CDE Mapping across Program Projects



BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

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

Other Cloud Platforms
AnVil, BDC, All of Us

DATA MAPPING

ACROSS DATASETS AND PLATFORMS
BASED ON CDES

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

URI approach for data interoperability



#1 Collaboration: Sharing Data Across Platforms



- Population Science/SDoH Data Sets
- Repository of Health Disparity, Minority Health, Health Care Delivery Project Data using CDEs



- Electronic Health Records
- Synthetic Practice Data

**USING BIG DATA TO BE A PART OF THE FUTURE OF
KNOWLEDGE GENERATION IN HEALTH DISPARITY RESEARCH**



ScHARe CDE Adoption:

Making Data Interoperable (URI Approach)
Concept Code Mapping (Data Harmonization)

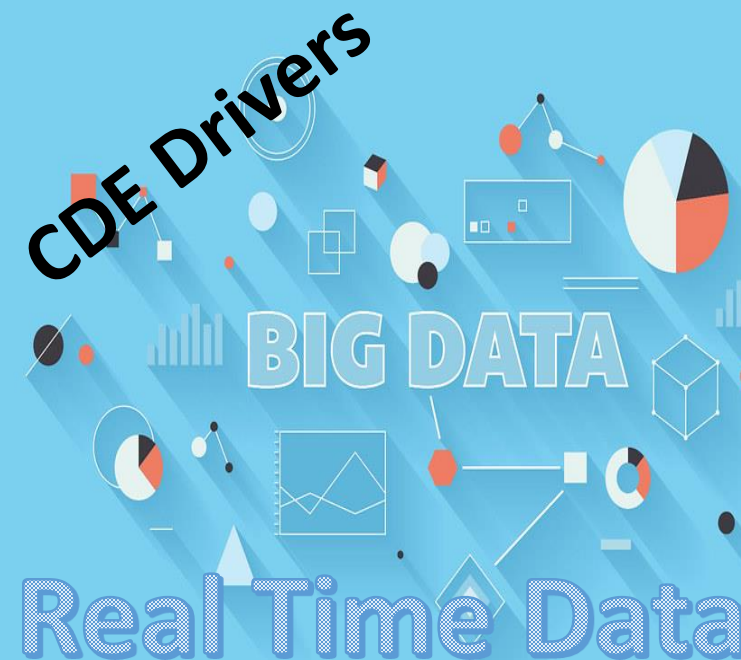




Adopted CDEs to advance Health Disparity and Health Care Delivery Research:

- Standardize data for people & computers (human and machine readable)
- Enable data sharing across studies (data interoperability)
- Enhance data interpretation & analysis (semantically defined and standardized coded)
- Simplify collaborations
- Reduces project start-up & results time

BIG DATA AND AI: REQUIRES NEW APPROACHES FOR COLLECTION, MANAGEMENT, ANALYSIS



Covid revealed the need to have real time data



ScHARe “CORE” CDE Development

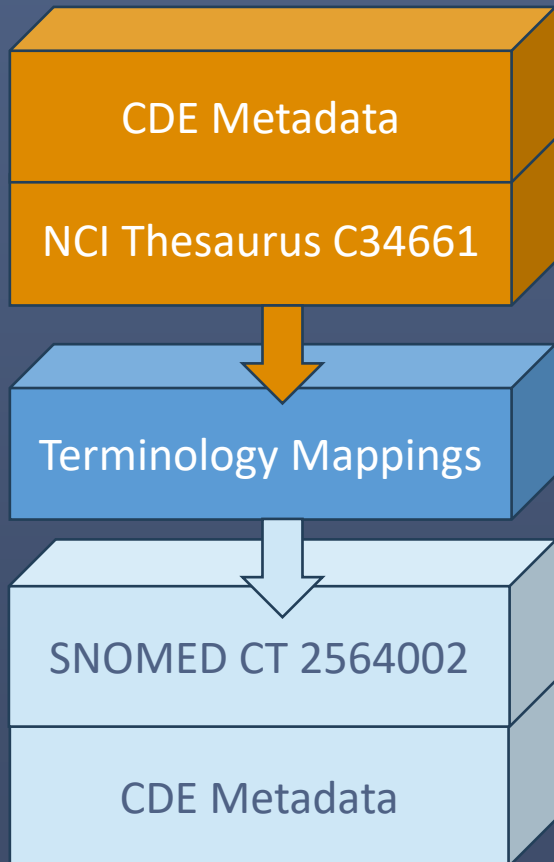
Core Set:

- Few critical questions required from all studies/sites
- Minimal burden
- Allows for questions to be asked in any way, but reported in a standardized format
- Allows for any number of other questions to be collected as collector chooses

Criteria:

- PhenX Toolkit first
- Validated source
- Adaptation of a validated source
- Generate new CDEs to fill gap areas (SDoH screening tool)

Importance of Concept Code Mapping and Interoperability (Uniform Resource Identifier (URI))



- CDE unique **CONCEPT CODES** represent data semantics
 - Human readable
 - Machine readable format
- **Mapping** enables interoperability even if the same standard terminology was not used in another CDE
- CDE Metadata enables searching for concept codes across CDEs to compare data

CDES: Words Precisely Defined-Shared Meaning

Why are Concept Definitions Important? Words can be **SEMANTICALLY AMBIGUOUS!**

- Context is important in conveying meaning when using CDEs
 - Words have different meanings depending on words around it and context.
 - Words have different cultural meanings and implications
- Some examples:
 - **Seizure:** uncontrolled electrical activity between brain cells / spiritual experience?
 - **Agent:** chemical compound or government employee?
 - **Alcohol:** disinfecting or drinking?
 - **Colon:** sentence punctuation or biological organ?
 - **Mole:** animal, blemish, unit of measure, or spy?
 - **Probe:** examination, investigation, or instrument?

Words can mean different things in different contexts



Questions become CDEs When Defined and Coded

Education

What is the highest level of education you have completed?

Shared Semantics and Concept Code:

An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools. **C17953**

URI approach in data repository uses codes to harmonize data rather than semantics (words).

Human Readable w Shared Meaning

Codes facilitate Machine Readable

Human &
Machine
Readable



Questions become CDEs When Defined and Coded

Education

What is the highest level of education you have completed?

Shared Semantics and Concept Code:

An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools. **C1795**

URI approach in data repository uses codes to harmonize data rather than semantics (words).

Human Readable w Shared Meaning

Machine-Readable format—excel spreadsheet: codes increase interoperability and use of pipes to separate concepts & codes

Permissible Value (PV) Labels	PV Definitions	PV Concept Identifiers
No formal Schooling	Indicates that a person has never attended an educational program or formal schooling.	C67122
Primary/Grade/Elementary School (approximately grades 1st through 5th)	Indicates that 5th grade potentially is the highest level of educational achievement.	C67127
Middle School/Lower Secondary Education (approximately grades 6th through 8th)	Indicates that 8th grade potentially is the highest level of educational achievement.	C67130

Some Concept Coding Systems One NOT Better Than the Other

General use....

LOINC Laboratory and Clinical Research

ULMS (CUI) Biomedical

FHIR Electronic Health Records

*NCIt Cancer

*ScHARe used NCIt because it has several population concepts

How a Survey Question Became a CDE

Please select the racial category or categories with which you most closely identify. *(select all that apply)*

- American Indian or Alaska Native
- Asian or Asian American
- Black or African American
- Hispanic or Latino
- Native Hawaiian or Other Pacific Islander
- Middle Eastern or North African (in current reporting tables will be reported as white)
- White

Survey Questions become CDEs when they are:

- **semantically defined by a standardized coding system for shared meaning**
- **in a format that is human and machine readable for ease of reuse**

Making of a CDE from a Protocol/Question

Need a standardized defined concept and related code.

Source: NCI Thesaurus

Race/Ethnicity Self-Identification

A textual description of a person's race. **C17049** | The ethnicity of a person. **C16564** | An individual's perspective or subjective interpretation of an event or information. **C74528**

American Indian or Alaska Native |
Asian or Asian American |
Black or African American |
Hispanic, Latino, or Spanish |
Native Hawaiian or Other Pacific Islander |
Middle Eastern or North African |
White

URI approach in data repository uses codes to harmonize data rather than semantics. This improves data interoperability.

Making of a CDE from a Protocol/Question

- A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment. (OMB) **C41259** |
- A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (OMB) **C41260** |
- A person having origins in any of the Black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American". (OMB) **C16352** |
- A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin" can be used in addition to "Hispanic or Latino". (OMB) **C17459** |
- A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. (OMB) **C41219** |
- Denotes a person having origins in the region of southwest Asia, between the India subcontinent and Europe, including Kuwait, Turkey, Lebanon, Israel, Iraq, Iran, Jordan, Saudi Arabia, lands east of Pakistan or the other countries of the Arabian Peninsula. Also includes people of Jewish ethnicity including Sephardic and Ashkenazic. **C77820** :
- Denotes a person whose ancestry is in any of the countries of the northern part of the African continent: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia, and Western Sahara. **C126529** |
- A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. (OMB) **C41261**

Making of a CDE from a Protocol/Question

Need a standardized defined concept and related code. Source: NCI Thesaurus

Code Mapping

NCIT

Loinc

UMLS CUI

American Indian or Alaska Native	C41259	LA10608-0	C0282204
Asian or Asian American	C41260	LA6156-9	C0003988
Black or African American	C16352	LA10610-6	C0085756
Hispanic, Latino, or Spanish	C17459	LA6214-6	C0086409
Native Hawaiian or Other Pacific Islander	C41219	LA10611-4	C1513907
Middle Eastern or North African	C43866	Mena no loinc	C1553353
White	C41261	LA4457-3	C0043157

Data Harmonization: Matched Same CDE from different projects



Income (Project 1)

Collected this way

Less than \$10,000 |
\$10,000-\$24,999 |
\$25,000-\$34,999 |
\$35,000-\$49,999 |
\$50,000-\$74,999 |
\$75,000-\$99,999 |
\$100,000-\$149,999 |
\$150,000-\$199,999 |
\$200,000 or more

Core CDE

Reported this way

Less than \$10,000 |
\$10,000-\$24,999 |
\$25,000-\$34,999 |
\$35,000-\$49,999 |
\$50,000-\$74,999 |
\$75,000-\$99,999 |
\$100,000-\$149,999 |
\$150,000-\$199,999 |
\$200,000 or more

Income (Project 2)

Collected this way

Less than \$10,000 |
\$10,000-\$24,999 |
\$25,000-\$34,999 |
\$35,000-\$49,999 |
\$50,000-\$74,999 |
\$75,000-\$99,999 |
\$100,000-\$149,999 |
\$150,000-\$199,999 |
\$200,000 or more

Data Harmonization: Mappable Different Approach with Conceptually the Same CDE

Income (Project 1)

Collected this way

Less than \$24,999 |
\$25,000-\$49,999 |
\$50,000-\$74,999 |
\$75,000-\$99,999 |
\$100,000-\$199,999 |
\$200,000 or more

Core CDE

Reported this way

Less than \$10,000 |
\$10,000-\$24,999 |
\$25,000-\$34,999 |
\$35,000-\$49,999 |
\$50,000-\$74,999 |
\$75,000-\$99,999 |
\$100,000-\$149,999 |
\$150,000-\$199,999 |
\$200,000 or more

Income (Project 2)

Collected this way

Less than \$10,000 |
\$10,000-\$19,999 |
\$20,000-\$29,999 |
\$30,000-\$39,999 |
\$40,000-\$49,999 |
\$50,000-\$59,999 |
\$60,000-\$69,999 |
\$70,000-\$79,999 |
\$80,000-\$89,999 |
\$90,000-\$99,999 |

.....

\$200,000 or more

**Mapped using algorithms





Adopted CDEs prospective approach to General Data Ecosystem Repository

- Standardize data for people & computers (human and machine readable)
- Enable data sharing across studies (data interoperability)
- Enhance data interpretation & analysis
- Simplify collaborations
- Reduces project start-up & results time

BIG DATA AND AI: REQUIRES NEW APPROACHES FOR COLLECTION, MANAGEMENT, ANALYSIS

CDE Benefits

Real Time Data

BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION



ScHARe NIH Endorsed CDEs in Survey Format





ScHARe CDEs Labels

For FUNDED PROJECT DATA – Common Data Elements Centralized for Interoperability and Data Sharing

- 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 (Associated Medications/Treatments)

**NIMHD Framework

**Health Disparity Outcomes



(** project level CDE)

NIH CDE Repository: <https://cde.nlm.nih.gov/home>

Cross-walked with PhenX SDoH

NIH-endorsed CDEs have been reviewed and approved by an expert panel, and meet established criteria. They are designated with a gold ribbon. 🏆

1. Age

What is the person's age? (collapse data over 89 yrs old / 2 yrs and under, report in months-does not exclude asking full birthdate)

_____ years months

Project 5 Covid-19 Age <https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true>

2. Birthplace |

Where were you born?

- In the United States, including U.S. Territories (Puerto Rico, Guam, U.S. Virgin Islands, American Samoa and Northern Mariana Islands) (**Select from Drop Down-not doable on word doc**)
- Outside the United States (**Select from Drop Down-ISSO categories-not doable on word doc**)

PhenX – Birthplace <https://www.phenxtoolkit.org/protocols/view/10201> ADAPTED-Territoires with US; instead of seperate

Source for PhenX : American Community Survey (ACS), 2008

3. ZIP code (caveat collapse zip codes w less than 10)

What is your current postal ZIP code? _____

Project 5 Covid-19 Address Postal Code https://cde.nlm.nih.gov/deView?tinyId=w_BHatIMoA

4. Self-Identification (This question's intent is to get at bare minimum of identification, which will be determined by the changes proposed by OMB. Study can collect details of Race and Ethnicity as preferred. This does not supplant other required R/E reporting. Awaiting OMB.)

Please select the racial category or categories with which you most closely identify. (select all that apply)

- American Indian or Alaska Native
- Asian or Asian American
- Black or African American
- Hispanic or Latino
- Native Hawaiian or Other Pacific Islander
- Middle Eastern or North African (in current reporting tables will be reported as white)
- White

Potential
New
Approach

SCHARE working group preference based on potential classifications in 2030 census <https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic#:~:text=And%20under%20that%20combined%20question%2C%20the%20list%20of,federal%20agencies%20collect%20data%20on%20race%20and%20ethnicity.>

Potential
New
Approach

5. Sex

What was your sex assigned at birth, on your original birth certificate?

- Female
- Male
- Intersex
- None of these describe me
- Prefer not to answer

PhenX Protocol - Biological Sex Assigned at Birth <https://www.phenxtoolkit.org/protocols/view/11601>

All of Us Research Program, Participant Provided Information (PPI), 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation

<https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health>

and All of Us

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

6. Gender

What is your current gender? [Select only one]

- Man
- Woman
- Non-Binary
- Transgender
- None of these describe me-I would like to consider additional options

Are any of these a closer description to your gender identity?

- Trans man/Transgender Man/FTM
- Trans woman/Transgender Woman/MTF
- Genderqueer
- Genderfluid
- Gender variant
- Questioning or unsure of your gender identity
- None of these describe me, and I want to specify _____
- Prefer not to answer

Potential New Approach

PhenX Protocol - Gender Identity <https://www.phenxtoolkit.org/protocols/view/11801>

All of Us Research Program, Participant Provided Information (PPI), 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation

<https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health>

Adapted: Non Binary added

7. Sexual orientation

Which of the following best represents how you think of yourself? [Select only one]

- Lesbian
- Gay
- Straight, that is, not gay or lesbian, etc.
- Bisexual

If none of the above represents you, are any of these a closer description of how you think of yourself (drop down)

- Queer
 - Polysexual, omnisexual, sapiosexual or pansexual
 - Asexual
 - Two-spirit
 - Have not figured out or are in the process of figuring out your sexuality
 - Mostly straight, but sometimes attracted to people of your own sex
 - Do not think of yourself as having sexuality
 - Do not use labels to identity yourself
 - Don't know the answer
 - No, I mean something else (optional free text) _____
- Prefer not to answer

Phen X Sexual Orientation Protocol. <https://www.phenxtoolkit.org/protocols/view/11701?origin=subcollection>

All of Us Research Program Participant Provided Information (PPI) Version: December 17, 2018

National Academies Sciences, Engineering, Medicine report: Measuring Sex, Gender Identity, and Sexual Orientation

<https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health>

Potential
New
Approach

8. Marital status

What is your current marital status?

- Married
- Living as married or living with a romantic partner
- Divorced
- Widowed
- Separated
- Single, never been married-not living with romantic partner
- Prefer not to answer

Hints 5 Cycle 4 (2020) <https://hints.cancer.gov/view-questions-topics/question-details.aspx?qid=593> (BRFSS Questionnaire (2001), Section 13: Demographics modified)

All of Us

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

9. Education

What is the highest level of education you have completed?

- No formal schooling
- Primary/Grade/Elementary School (approximately grades 1st through 5th)
- Middle School/Lower Secondary Education (approximately grades 6th through 8th)
- Secondary/High School/Upper Secondary (grades 9th through 11th) without a high school diploma
- General Educational Diploma (GED)
- Secondary/High School/Upper Secondary (grades 9th through 12th) with a high school diploma
- Occupational/Technical/Vocational Programs/Short Cycle Tertiary Education - Associate's Degree (approximately 2 years)
- College/University/Bachelor's Degree/Equivalent Tertiary Education (approximately 3-5 years)
- Graduate/post-graduate degree/professional degree/ (JD, PhD, MD, EdD, Eng, Master's Degree, etc.)

Unique -
Potentially
Mappable

International Standard Classification of Education (ISCED)

<https://datatopics.worldbank.org/education/wRsc/classification#:~:text=The%20International%20Standard%20Classification%20of,revise%20in%201997%20and%202011>

and

USA standards of Education <https://nces.ed.gov/programs/digest/d01/fig1.asp>.

10. Annual household income range

What is your annual household income from all sources within family, not including roommates?

- Less than \$10,000
- \$10,000-\$24,999
- \$25,000-\$34,999
- \$35,000-\$49,999
- \$50,000-\$74,999
- \$75,000-\$99,999
- \$100,000-\$149,999
- \$150,000-\$199,999
- \$200,000 or more

Potentially Mappable

All of Us - Basic Information Survey https://allofus.nih.gov/sites/default/files/aou_ppi_basics_version.pdf

BRFSS = Behavioral Risk Factor Surveillance System (CDC)

11. Household family size

Approximately how many individuals (adult and children) does your household family income support?

- _____

Project 5 Covid-19 Shared Living Space Number of Individuals <https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true>

12. English proficiency

We are interested in your own opinion of how well you speak English. Would you say you speak English:

- Very well
- Well
- Not well
- Not at all
- Refused
- Don't Know

PhenX Toolkit - English Proficiency <https://www.phenxtoolkit.org/protocols/view/270201>

Regents of the University of California. (2019). CHIS 2018 Adult Questionnaire, question number "QA18_G8" is represented in this protocol as question 1. Retrieved from <http://healthpolicy.ucla.edu/chis/design/Pages/questionnairesEnglish.aspx>

13. Disabilities

Do you have a disability or have serious difficulty with any of the following? Select all that apply.

- Deafness or difficulty hearing
- Blindness or difficulty seeing
- Difficulty concentrating, remembering, and deciding
- Difficulty walking or climbing stairs
- Difficulty dressing or bathing
- Difficulty doing errands alone
- Not disabled

CDC Standard Disability Questions <https://www.cdc.gov/ncbddd/disabilityandhealth/datasets.html> (format adapted)

14. Health insurance

Are you currently covered by any of the following types of health insurance or health coverage plans?

- Insurance through a current or former employer or union (of yours or another family member's). This would include COBRA coverage.
- Insurance purchased directly from an insurance company (by you or another family member). This would include coverage purchased through an exchange or marketplace
- Medicare, for people 65 and older, or people with certain disabilities.
- Medicaid, Medical Assistance (MA), the Children's Health Insurance Program (CHIP), or any kind of state or government-sponsored assistance plan based on income or a disability.
- TRICARE or other military health care, including VA health care.
- Indian Health Service
- Any other type of health insurance. coverage or health coverage plan
- Uninsured

15. Employment status

We would like to know about what you do: are you working now, looking for work, retired, keeping house, a student, or what?

- Working now or paid sick leave/parental leave/family leave/administrative leave
- Only temporarily laid off, or unpaid sick leave/parental leave/family leave/administrative leave
- Looking for work, unemployed
- Retired
- Disabled, permanently or temporarily
- Raising children full-time, full-time caregiver, or keeping house
- Student
- Other/specify: _____

PhenX - Current Employment Status <https://www.phenxtoolkit.org/protocols/view/11301> (Adapted-used parental instead of maternal, and family leave added with paid/unpaid)

16. Usual place of health care

Is there a place that you **USUALLY** go to when you are sick or need advice about your health? Select all that apply.

- A doctor's office or community health center, including Indian Health Service, or hospital-based clinics
- Walk-in clinic, urgent care center, or retail clinic in a pharmacy or grocery store
- Emergency room
- A VA Medical Center or VA outpatient clinic
- Some other place
- Does not go to one place most often
- Don't know

PhenX Protocol Access to Health Services Ques #5 <https://www.phenxtoolkit.org/protocols/view/270101> (adapted with hospital-based clinics)

Project 5 Covid-19 Usual Place of Health Care Type <https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true> (adapted with hospital-based clinics)

17. Economic Stability – Social Needs

In the past year, have you or any family members you live with been unable to get any of the following when it was really needed? Select all that apply.

- Childcare
- Clothing
- Food
- Housing
- Internet/Broadband
- Phone (e.g., mobile or landline)
- Transportation (e.g., private or public)
- Utilities (e.g., gas, electric, propane, natural gas, etc.)
- Medicine or any health care (medical, dental, mental health, vision)
- Other/specify: _____

High Level SDoH
Assessment

Source: Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) tool (Adapted-internet, housing, transportation added to question #14) Housing and transportation is included in survey. <https://prapare.org/wp-content/uploads/2023/01/PRAPARE-English.pdf>

U.S. Census Bureau, 2015 and 2016 American Community Survey – Internet/Broadband
<https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf>

18. Self-reported health

Would you say your health in general is excellent, very good, good, fair, or poor?

- Excellent
- Very good
- Good
- Fair
- Poor

Patient-Reported Outcomes Measurement Information (PROMIS)

https://www.healthmeasures.net/index.php?option=com_instruments&task=downloadComponentFile&file=PROMIS%20Scale%20v1.2%20-%20Global%20Health%20Physical%20a%2009062016.pdf

19. Health conditions and medications or other Treatments

Has a health care provider told you that you have any one or more of the following conditions? Select all that apply currently. Check the second box if you are taking medications or receiving some other treatment for the condition.

- Cancer
- Coronary heart disease
- Heart failure
- High blood pressure/hypertension
- Stroke
- Thrombotic disorders
- High cholesterol
- Diabetes (type I)
- Diabetes (type II)
- Obesity
- Hepatitis
- Other chronic liver disease

Addresses
Co-Morbidities

- Asthma
- Other chronic respiratory disease (e.g., COPD, emphysema)
- Chronic kidney disease
- Psychological and/or psychiatric disease or disorder (e.g., anxiety, depression, bipolar disorder)
- Alzheimer's disease
- Dementia
- Epilepsy
- Multiple sclerosis
- Other chronic neurological condition (e.g., Parkinson's disease, migraine)
- Immunodepression
- HIV/AIDS
- Autoimmune condition (e.g., rheumatoid arthritis, systemic lupus erythematosus, vasculitis)
- Chronic musculoskeletal condition (e.g., back pain, osteoarthritis, osteoporosis)

- Sickle cell disease
- Sleep disorder (e.g., insomnia, sleep apnea, narcolepsy)
- Solid organ transplant
- Smoking
- Other substance use disorder (e.g., drugs and/or alcohol dependence)
- Long Covid (also known as long-haul COVID, long-term effects of COVID, chronic COVID, post-acute COVID-19, and PASC - post-acute sequelae of SARS-CoV-2)
- Chronic fatigue
- Dental diseases and conditions (e.g., caries, periodontal disease, oral and pharyngeal cancer)
- Eye diseases and conditions (e.g., cataract, glaucoma, amblyopia, myopia and other refractive errors, age-related macular degeneration, diabetic retinopathy, ocular trauma, uveitis, keratoconus)
- Other chronic disease/specify:
- None of the above

Project 5 Covid-19 Comorbidity or Underlying conditions

<https://cde.nlm.nih.gov/cde/search?q=PROJECT%205&nihEndorsed=true> (Adapted for Medications-Added Chronic musculoskeletal conditions, High Cholesterol, Sleep Disorders and Stroke)

20. Minority Health and Health disparities research content area

Which of the following content areas of research is this study addressing, if any? Select all that apply.

- Minority health study focused on a one race or ethnic population and not addressing a health disparity.
- Health Disparity Outcome (select the focus area)
 - Higher incidence and/or prevalence of disease, including earlier onset or more aggressive progression of disease
 - Premature or excessive mortality from specific health conditions
 - Greater global burden of disease, such as Disability Adjusted Life Years (DALY), as measured by population health metrics
 - Poorer health behaviors and/or clinical outcomes using established measures
 - Worse outcomes on validated self-reported measures that reflect daily functioning or symptoms from specific conditions
- Other Health Outcomes / Healthcare Delivery

Duran D, Perez-Stable, E. Novel Approaches to Advance Minority Health and Health Disparities; Am J Public Health. 2019, Jan;109(S1):S8-S10. doi:10.2105/AJPH. 2019.304952. PMID: 30699026; PMCID:PMC6356133. ADAPTED with Other health outcomes delivery/care

21. NIMHD Framework

What NIMHD Research framework levels and domains of influence is your study targeting? (Select all that apply)

Levels of Influence

- Individual
- Interpersonal
- Community
- Societal

Domains of Influence

- Biological
- Behavioral
- Physical/Built Environments
- Sociocultural Environment
- Health Care Systems and Clinical Care



What are Think-a-Thons?





Think-a-Thons (TaT)

- Monthly sessions (2 1/2 hours)
- Instructional/interactive
- Designed for new and experienced users
- Research & analytic teams to:
 - Conduct health disparities, health outcomes, bias mitigation research
 - Analyze/create tools for bias mitigation
- Publications from research team collaboration
- Networking
- Mentoring and coaching
- Focus:

Types:

- ✓ Instructional / Tutorial
- ✓ Collaborative Research Teams
- ✓ Bias mitigation

ScHARe

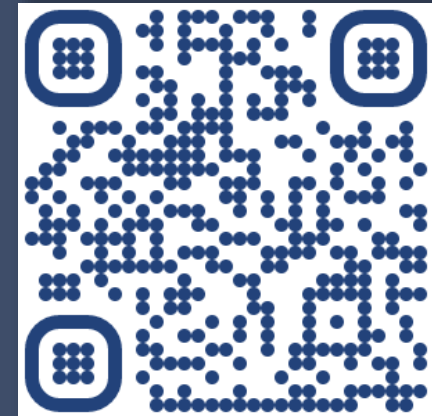
Think-a-Thon

Artificial Intelligence and
Cloud Computing Basics

Terra: Datasets and
Analytics



Register:



bit.ly/think-a-thons

Generational Career & Discipline Exchange



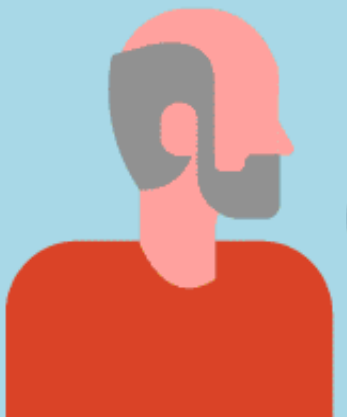
Re-SKILL

Up-SKILLING



LLING

Re-SKILLING



Up-SKIL

Think-a-Thon Instructional Tutorials

Web: bit.ly/think-a-thons

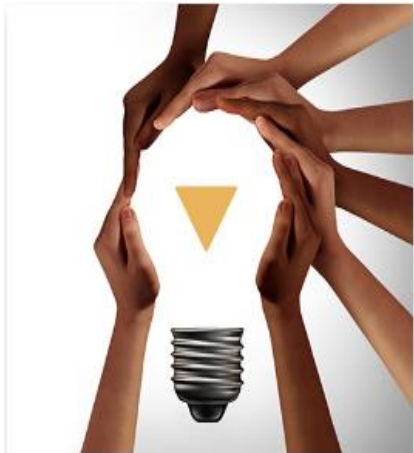


February	Artificial Intelligence and Cloud Computing 101
March	ScHARe 1 – Accounts and Workspaces
April	ScHARe 2 – Terra Datasets
May	ScHARe 3 – Terra Google-hosted Datasets
June	ScHARe 4 – Terra ScHARe-hosted Datasets
July	An Introduction to Python for Data Science – Part 1
August	An Introduction to 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	Preparing for AI 4: Overview Prep for AI Summary with Transparency, Privacy, Ethics

ScHARe for Educators (Community Colleges & Low Resource MSIs)

ScHARe for American Indian / Alaska Native Researchers

ScHARe for Non-Biomedical Researcher Coders and Programmers to conduct Research



The monthly **SchARE Think-a-Thons** scheduled or archived below are designed so participants reach one of these goals (as noted with each session):

- **Goal 1:** Achieve a **better understanding of both the fields and the terminology** used to describe the AI/cloud computing infrastructure, components and processes.
- **Goal 2:** **Develop research questions and projects relevant to AI and cloud computing** that leverage the cutting-edge technology and data/computing resources now available to health disparities researchers (including the ones at their disposal on the SchARE platform).

Upcoming Think-a-Thons

Past Think-a-Thons

See FAQs



SchARE

Think - a - Thons

PAST Think-a-Thons Posted

November 15, 2023 2.5 hours View video: [Preparing for AI 2: An Introduction to FAIR Data and AI-ready Datasets](#) [View slides \(PDF, 4 MB\)](#)

Toward Goal 1:
How to prepare an AI-ready dataset using gold standard data management principles, including:

- Making datasets findable, accessible, interoperable, and reusable (FAIR)
- Using transparent data documentation to foster data re-use
- Ensuring that selected data addresses expected outcomes and drives meaningful AI insights
- Handling missing data through strategies, proxies, and synthetic data

<https://www.nimhd.nih.gov/resources/schare/think-a-thons.html>

Think-a-Thon Schedule

Think-a-Thons are held on the third Wednesday of each month. [Accommodations information](#) | [Think-a-Thon recordings](#)

Date	Time	Topic	Register
March 20, 2024	2:00 – 4:30 p.m. ET	<p>Preparing for AI-driven Research on SchARE: A Comprehensive Review and Brainstorming Session – Part 2</p> <p>Toward Goal 2:</p> <p>Prepares participants for SchARE research collaborations by covering:</p> <ul style="list-style-type: none"> • Choosing computational strategies (AI, ML, statistics) • An overview of Python data science libraries • The significance of testing and monitoring in algorithm development • The role of open science in ensuring reproducible and transparent AI-based research <p>For researchers and students at all levels who want to collaborate on SchARE to develop innovative and publishable research projects</p>	<p>Register</p> <p><i>Registration closes at 12:00 p.m. ET on the day of the event.</i></p>



Upcoming



Think-a-Thons (TaT)

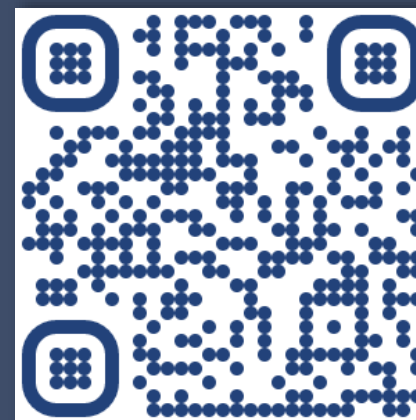
Research Teams

<p>Title: Data Science Projects 1 – Health Disparities and Individual SDoH</p> <p>Description: Exploring the impact of individual Social Determinants of Health on health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.</p>
<p>Title: Data Science Projects 2 - Health Disparities and Structural SDoH</p> <p>Description: Assessing the impact of structural Social Determinants of Health on health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.</p>
<p>Title: Data Science Projects 3 – Health Outcomes</p> <p>Description: Investigating the influence of non-clinical factors on disparities in health care delivery: a hands-on session for researchers and students at all levels interested in collaborating on ScHARe to develop innovative research questions and projects leading to publications.</p>

- Foster a research paradigm shift to use Big Data
- Promote use of Dark Data
- Generational Career & Discipline Exchange

- Multi-career (students to sr. investigators)
- Multi-discipline (data scientist & researchers)
- Feature Datasets with Guest Expert Leads
- Secure experts in topic area, analytics, data sources etc. to provide guidance
- Generate research idea - decide potential design, datasets & analytics
- Select co-leads to coordinate completion outside of TaT
- Publications

Register:



bit.ly/think-a-thons



Research Think - a - Thons



Expectation of the Research Project.

- The launch of the project will occur during the Think-a-Thon.
 - Pre-Assigned Co-Leads: Data Science Expert and a Health Disparity/Health Care Delivery Expert
 - There will be 4 sessions: 2 python, 1 R and 1 Statistic defined research collaborative
 - Volunteers who want to participate in health disparity/health care delivery research will select one of the 4 sessions based upon the analytics expected to be used
 - In the breakouts, the group decides the research topic and data sets to be used.
- The co-leads will assign tasks to the participants for the next **three months** to complete the project in preparation for publication. There will be meetings other than Think-a-Thons to:
 - review progress of tasks
 - help/teach others what each participant is contributing
 - assessing what else needs to be completed



Goal: Experience Conducting Ethical AI

TRANSPARENCY:

- Def:
 - Public Perception & understanding of how AI works
 - Comprehend the algorithmic views and decisions taken based on them
- Technical Documentation for duplication / re-use
- Tools:
 - Data Dictionary
 - Health Sheet (Data Sheet)
- Model Cards (capabilities & purpose of algorithms are openly and clearly communicated to relevant stakeholders)
- Documentation of methodologies
- Doesn't disclose intellectual property

FAIRNESS:

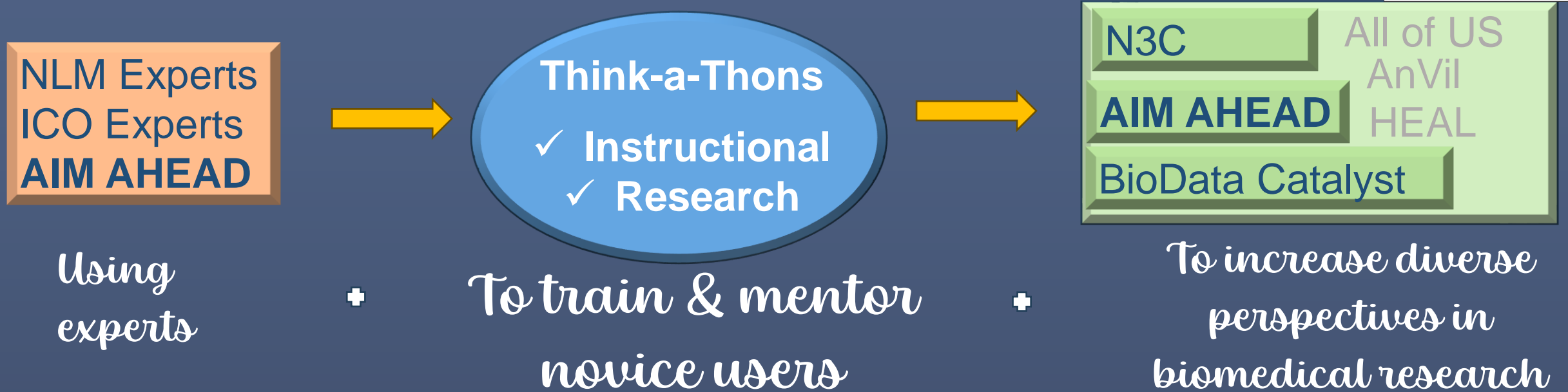
- **Findable:** providing metadata, documentation, clear identifiers
- **Accessible:** wide audience
- **Interoperable:** standardized formats and APIs enable seamless integration.
- **Reusable:** clear documentation, licensing, reduce redundancy

Metadata and data should be easy to find for both humans and computers

Ensure that data represents relevant

Data Democratization

#2 Collaboration: Think-a-Thons Training/Mentoring Pipeline



Goal: “UpSkilling” and “ReSkilling”

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

Target Audience:

- ✓ Underrepresented populations (women, race/ethnic) users not trained in data science
- ✓ Data scientist with no or little research experience.
- ✓ Resource & Tool for Community Colleges and Low Resource MSIs and Organizations



SchARE

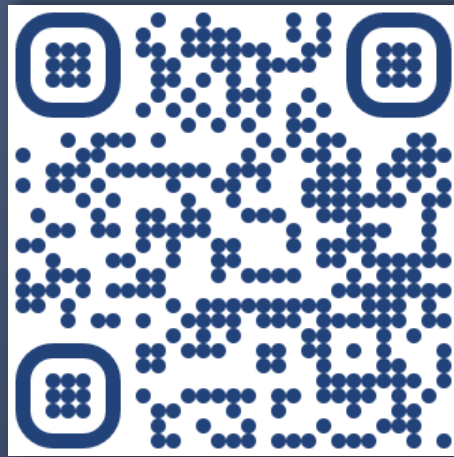
Thank You

Next Think-a-Thons:



bit.ly/think-a-thons

Register for SchARE:



bit.ly/join-schare

 schare@mail.nih.gov

durande@mail.nih.gov or luca.calzoni@nih.gov

