



ScHARe

ScHARe Repository Introduction

November 20, 2024

Deborah Duran, PhD • NIMHD
Elif Dede Yildirim, PhD • NIMHD
Mark Aronson, PhD • NIMHD



ScHARe

Science
collaborative for
Health disparities and
Artificial intelligence bias
Reduction

Outline

- 15'** **ScHARe Overview**
- 5'** **Repository Background**
- 5'** **Getting Started**
- 15'** **Uploading your first Data Set**
- 15'** **HANDS ON: Uploading Data**
- 15'** **CDE Mapping and Dataviews**
- 15'** **HANDS ON: Dataviews for CDE Mapping**
- 5'** **Sharing Data**
- 10'** **Data Aggregation and Analysis - Overview**
- 15'** **Conclusion and Q&A**

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

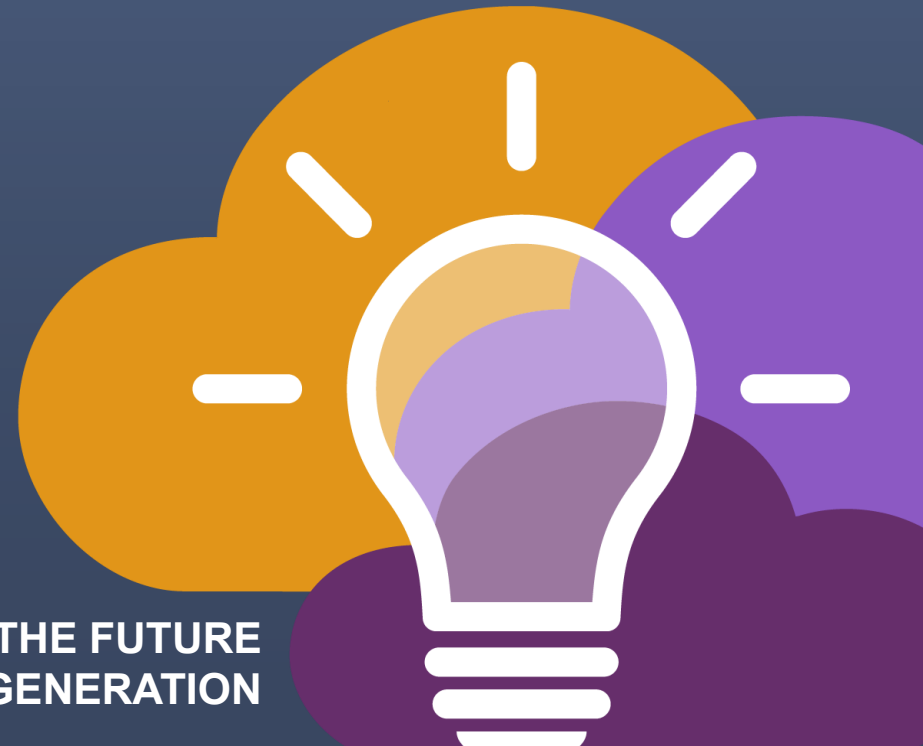
I am interested in (check all that apply):

- ☐ 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

Science
collaborative for
Health disparities and
Artificial intelligence bias
Reduction



Register: nimhd.nih.gov/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 **five 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 health disparity 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
- Offer a **project data repository** centered on core common data elements for enhanced data interoperability and compliance with NIH Data Management and Sharing Policy



ScHARe



Google Platform Terra Interface

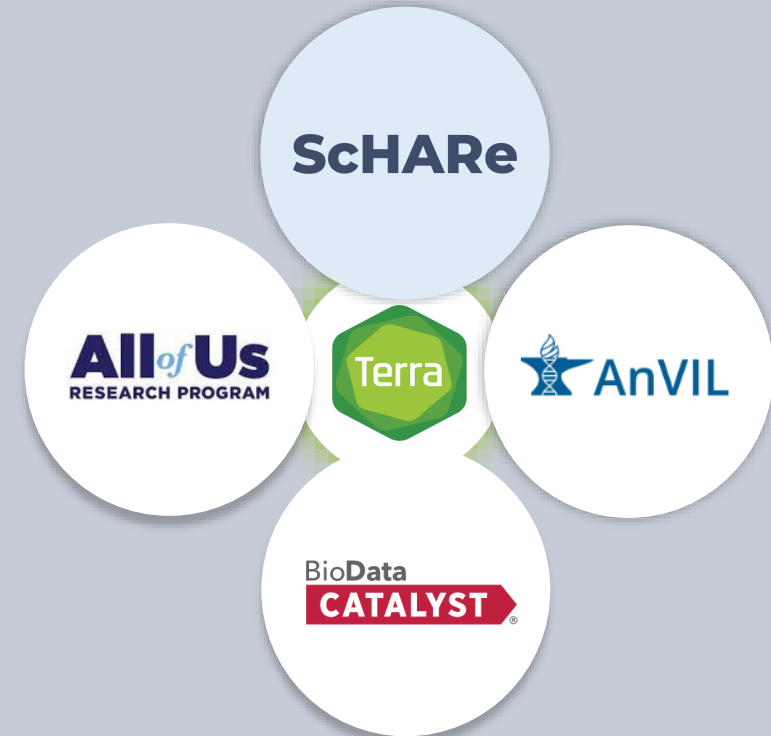
- Secure workspaces
- Data storage
- Computational resources
- Tutorials (how to)
- Copy-and-paste code in Python and R
- Learning Terra on ScHARe prepares you to use other NIH platforms



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

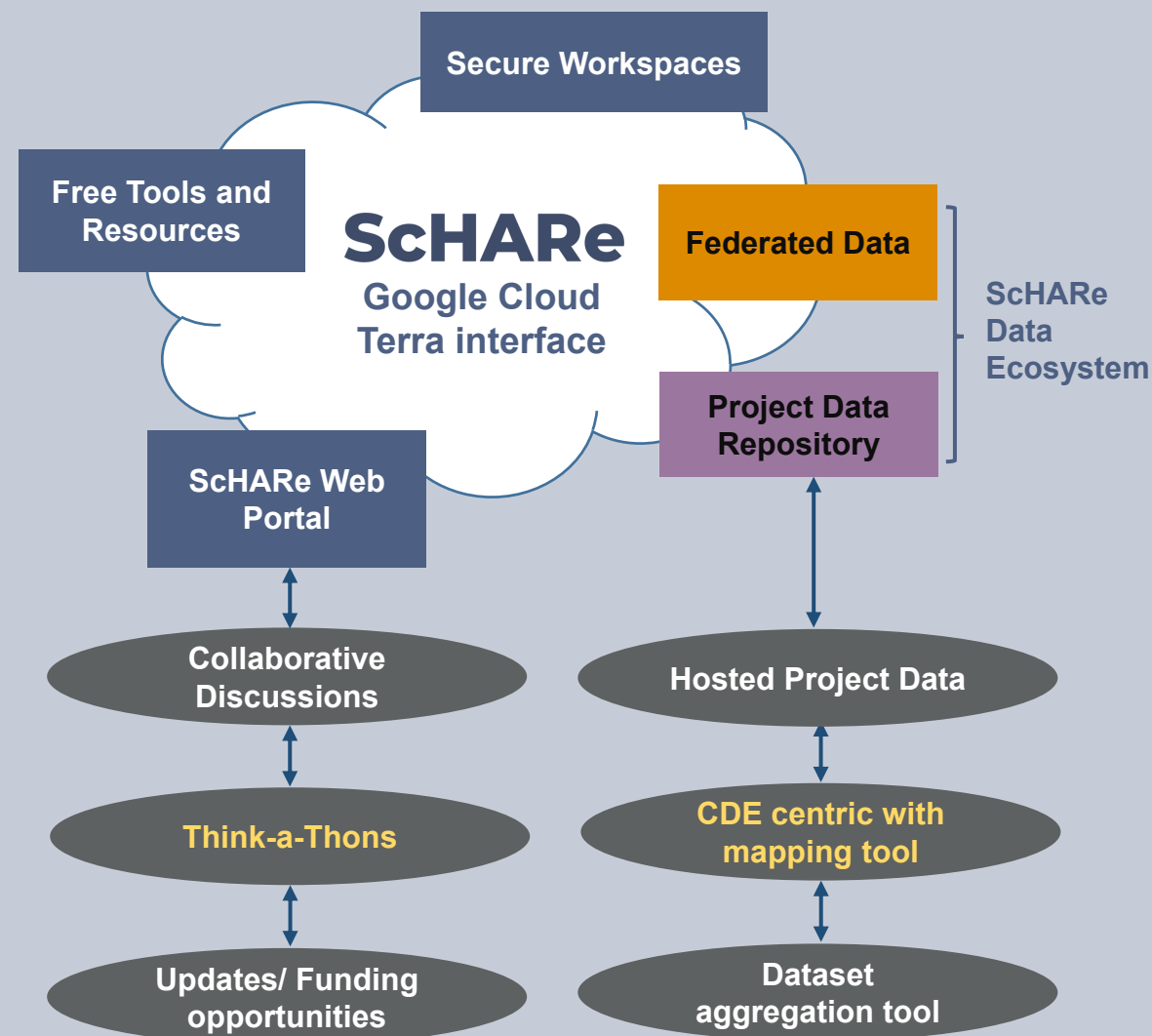


ScHARe Components

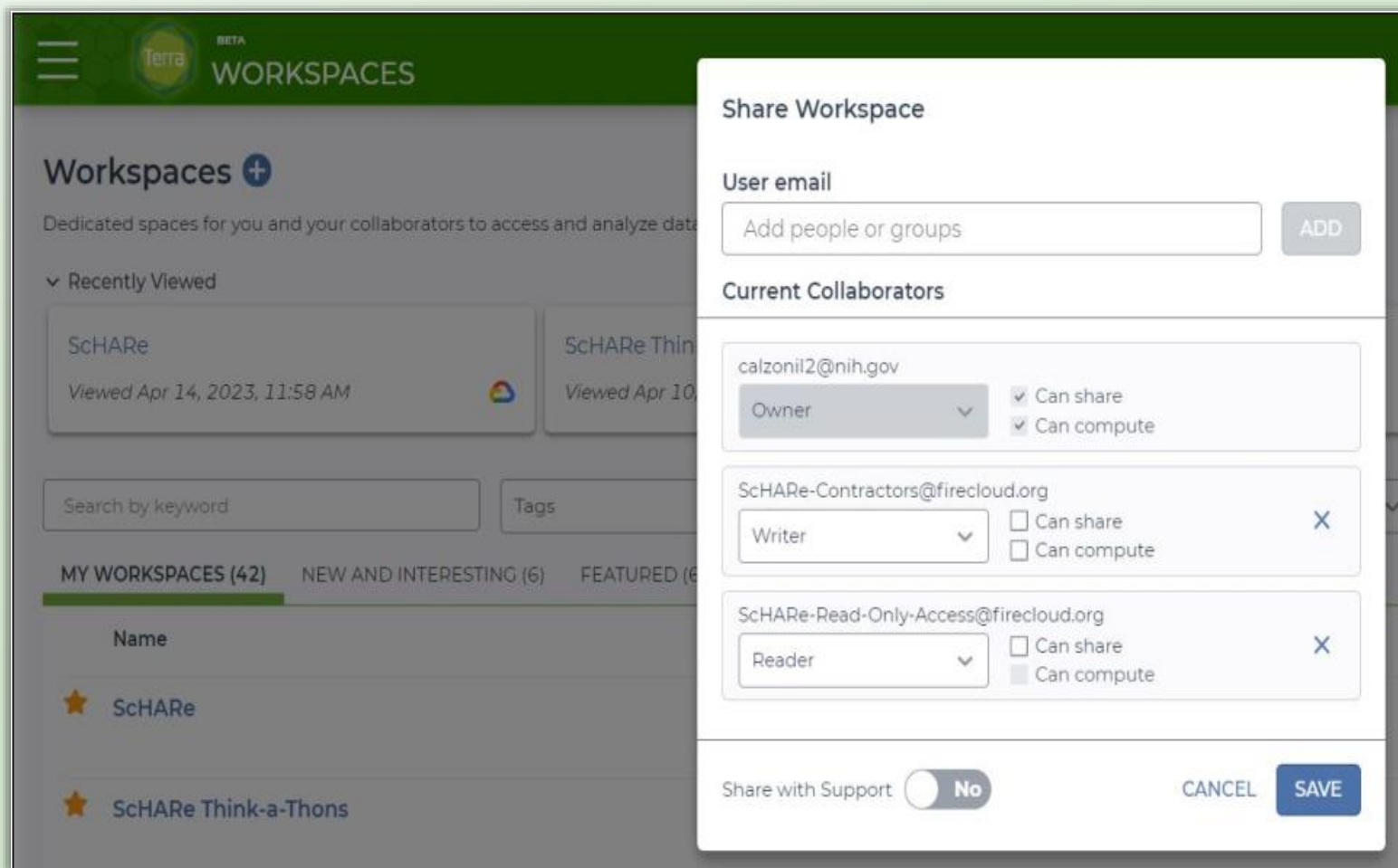
ScHARe co-localizes within the cloud:

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



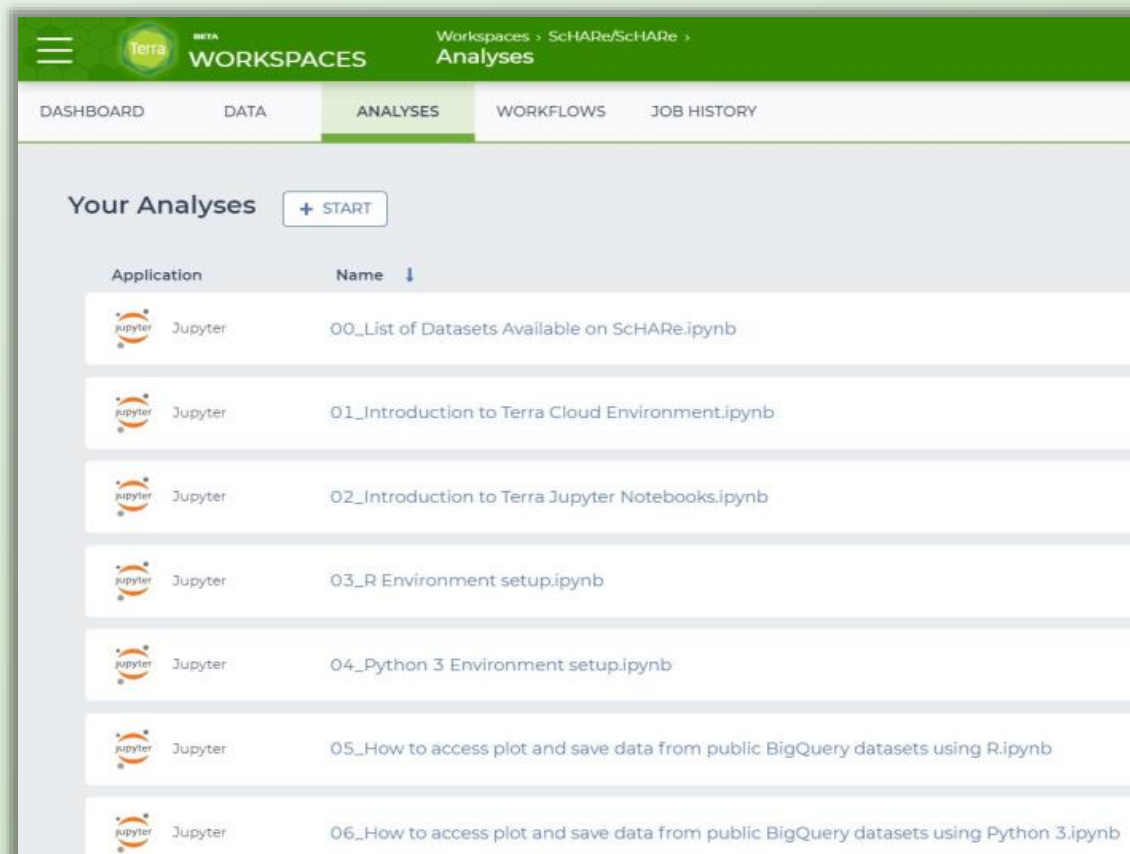
ScHARe Terra interface: secure workspace



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

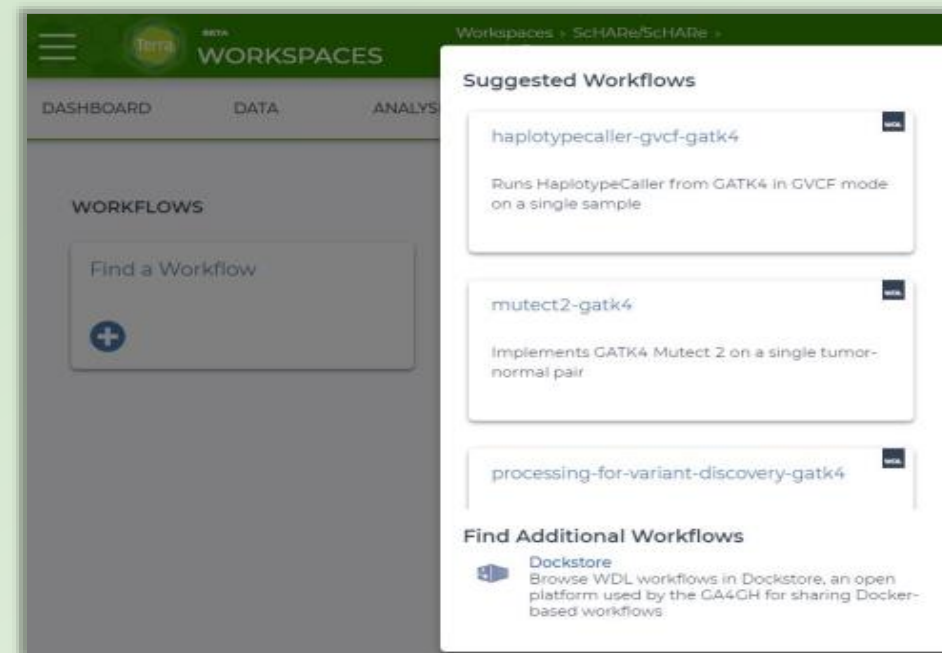
ScHARe Terra interface: analyses

Notebooks for analytics and tutorials



Modular codes

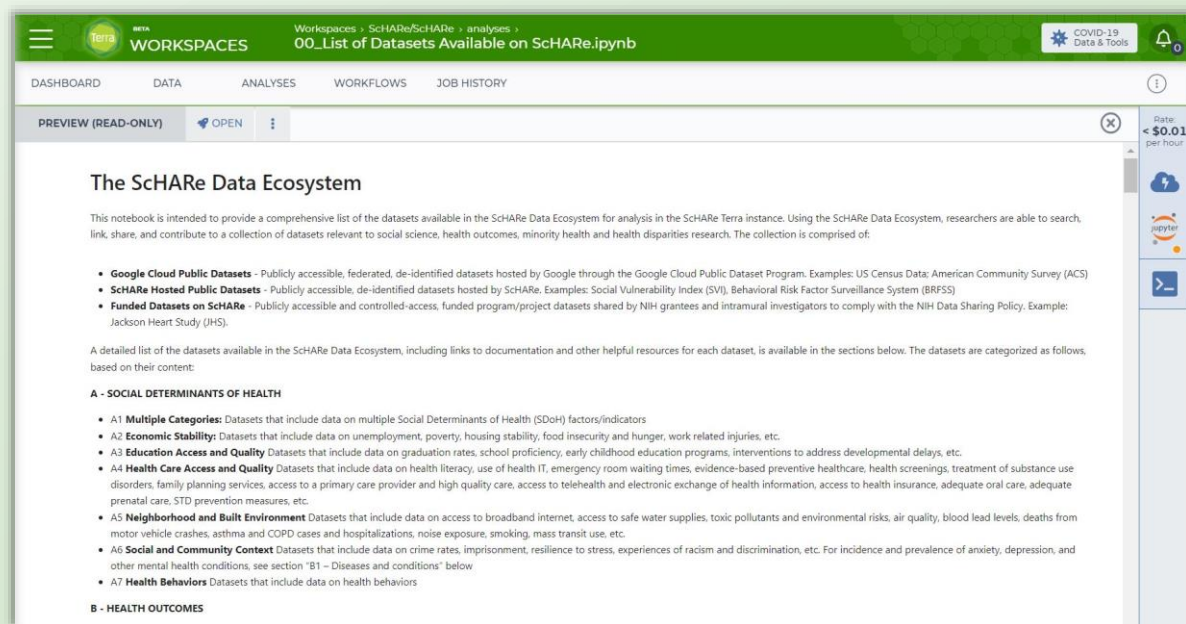
- Easy-to-use copy and paste analytics



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

ScHARe Terra interface: access to datasets

What data?



The screenshot shows the ScHARe Terra interface in the 'Analyzes' tab. A notebook titled '00_List of Datasets Available on ScHARe.ipynb' is open, displaying 'The ScHARe Data Ecosystem' section. The notebook content includes a list of dataset categories and a detailed list of datasets under the 'A - SOCIAL DETERMINANTS OF HEALTH' section.

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:

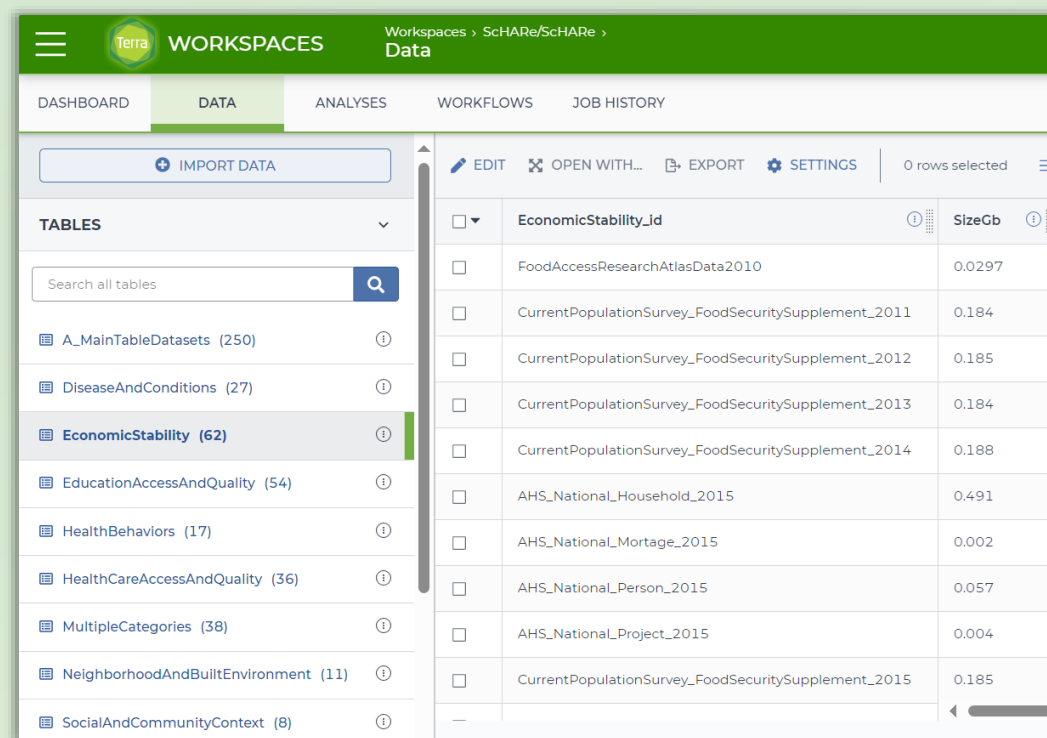
A - SOCIAL DETERMINANTS OF HEALTH

- **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

In the **Analyzes** 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. A table of datasets is displayed, with columns for 'EconomicStability_Id' and 'SizeGb'. The table lists various datasets, including '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', 'CurrentPopulationSurvey_FoodSecuritySupplement_2015', and 'SocialAndCommunityContext'.

EconomicStability_Id	SizeGb
FoodAccessResearchAtlasData2010	0.0297
CurrentPopulationSurvey_FoodSecuritySupplement_2011	0.184
CurrentPopulationSurvey_FoodSecuritySupplement_2012	0.185
CurrentPopulationSurvey_FoodSecuritySupplement_2013	0.184
CurrentPopulationSurvey_FoodSecuritySupplement_2014	0.188
AHS_National_Household_2015	0.491
AHS_National_Mortgage_2015	0.002
AHS_National_Person_2015	0.057
AHS_National_Project_2015	0.004
CurrentPopulationSurvey_FoodSecuritySupplement_2015	0.185
—	—

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:

280+
**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.: <i>Behavioral Risk Factor Surveillance System (BRFSS)</i>
Google	e.g.: <i>American Community Survey (ACS)</i>

**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

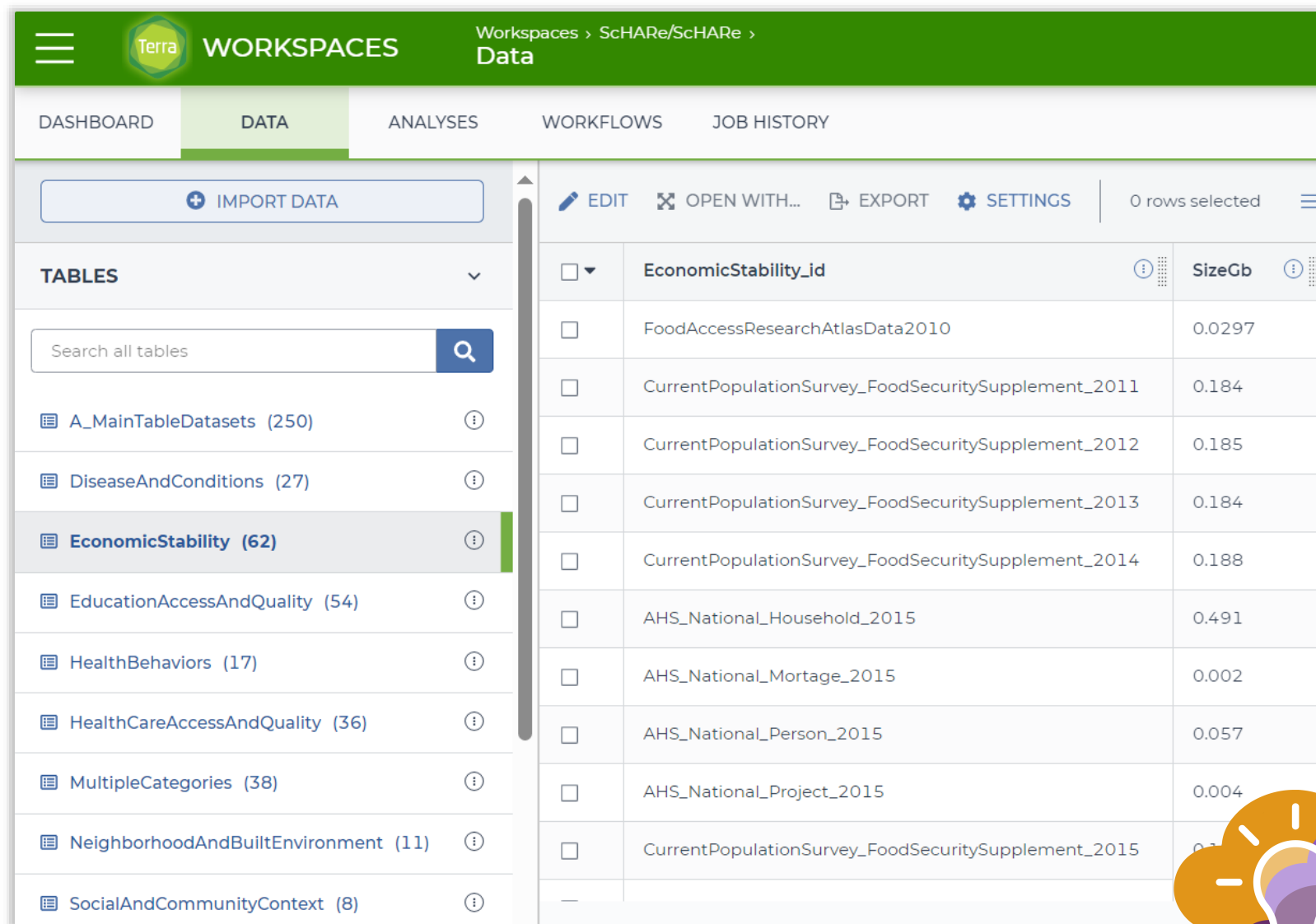
OVER 280 DATA SETS CENTRALIZED

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

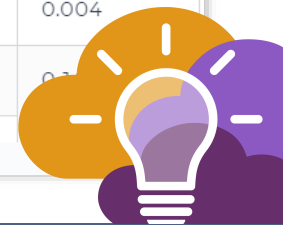
with the addition of:

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



The screenshot displays the Terra WORKSPACES Data interface. The top navigation bar includes a menu icon, the Terra logo, and the text 'WORKSPACES' and 'Data'. Below this, a secondary navigation bar shows tabs for 'DASHBOARD', 'DATA' (selected), 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. The main content area is divided into two panels. The left panel, titled 'TABLES', contains a search bar and a list of dataset categories with their respective counts: 'A_MainTableDatasets (250)', 'DiseaseAndConditions (27)', 'EconomicStability (62)' (highlighted), 'EducationAccessAndQuality (54)', 'HealthBehaviors (17)', 'HealthCareAccessAndQuality (36)', 'MultipleCategories (38)', 'NeighborhoodAndBuiltEnvironment (11)', and 'SocialAndCommunityContext (8)'. The right panel displays a table of datasets, including 'EconomicStability_id', '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'. Each row includes a checkbox, the dataset name, and its size in GB.

	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*:

280+ 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**



https://www.cdc.gov/about/priorities/social-determinants-of-health-at-cdc.html?CDC_AAref_Val=https://www.cdc.gov/about/sdoh/index.html

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

O6_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

**NIH
Endorsed**



- 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

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/

NIMHD Research Framework

		Levels of Influence*			
		Individual	Interpersonal	Community	Societal
Domains of Influence (Over the Lifecourse)	Biological	Biological Vulnerability and Mechanisms	Caregiver–Child Interaction Family Microbiome	Community Illness Exposure Herd Immunity	Sanitation Immunization Pathogen Exposure
	Behavioral	Health Behaviors Coping Strategies	Family Functioning School/Work Functioning	Community Functioning	Policies and Laws
	Physical/Built Environment	Personal Environment	Household Environment School/Work Environment	Community Environment Community Resources	Societal Structure
	Sociocultural Environment	Sociodemographics Limited English Cultural Identity Response to Discrimination	Social Networks Family/Peer Norms Interpersonal Discrimination	Community Norms Local Structural Discrimination	Social Norms Societal Structural Discrimination
	Health Care System	Insurance Coverage Health Literacy Treatment Preferences	Patient–Clinician Relationship Medical Decision-Making	Availability of Services Safety Net Services	Quality of Care Health Care Policies
Health Outcomes		 Individual Health	 Family/ Organizational Health	 Community Health	 Population Health



Project Level CCDEs – 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

NIMHD Research Framework. <https://www.nimhd.nih.gov/about/overview/research-framework/nimhd-framework.html>



NIMHD's Mission: Improve Minority Health

Minority Health:
Distinctive health characteristics and attributes of racial and/or ethnic minority populations who are socially disadvantaged due in part to being subject to racist or discriminatory acts and are underserved in health care.

Minority Health Research

The scientific investigation of singular and combinations of attributes, characteristics, behaviors, biology, and societal and environmental factors that influence the health of minority racial and/or ethnic population(s), including within-group or ethnic sub-populations, with the goals of improving health and preventing disease.

Minority Health Populations

The OMB Directive 15 defines racial and ethnic minority populations as:

- American Indian or Alaska Native
- Asian
- Black or African American
- Hispanic or Latino American
- Middle Eastern or North African
- Native Hawaiian or Pacific Islander



NIMHD's Mission: Reduce Health Disparities

Health Disparity:

A health disparity is a health difference that adversely affects disadvantaged populations in comparison to a reference population, based on one or more health outcomes.

All populations with health disparities are socially disadvantaged due in part to being subject to racist or discriminatory acts and are underserved in health care.

Health Disparity Research

A multi-disciplinary field of study devoted to:

- Gaining greater scientific knowledge about the influence of health determinants.
- Understanding the role of mechanisms.
- Determining how this knowledge is translated into interventions to reduce or eliminate adverse health outcomes.

Populations with Health Disparities

Populations that experience health disparities include:

- Racial and ethnic minority groups
- People with lower socioeconomic status (SES)
- Underserved rural communities
- Sexual and gender minority (SGM) groups
- People with disabilities



Health Disparity Outcomes

Unfair disadvantages that people face in different aspects of life, like education, income, or opportunities **can lead to health disparities**

Some groups of people may experience poorer health outcomes than others as a result

Health Disparity Outcomes

The health outcomes are categorized as:

- 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 clinical outcomes related to the aforementioned.
- Worse outcomes on validated self-reported measures that reflect daily functioning or symptoms from specific conditions.



Project Level CCDEs – Research Area Focus

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

- ☐ Minority health
- ☐ Health Disparity (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 clinical outcomes related to the aforementioned
 - ☐ Worse outcomes on validated self-reported measures that reflect daily functioning or symptoms from specific conditions
- ☐ Other Health Outcomes / Health Delivery or care

Duran DG, Pérez-Stable EJ. Novel Approaches to Advance Minority Health and Health Disparities Research. Am J Public Health. 2019 Jan; 109(S1):S8-S10. doi: 10.2105/AJPH.2018.304931. PMID: 30699017; PMCID: PMC6356124. ADAPTED with Other health outcomes delivery/care





ScHARe

ScHARe Repository Introduction

November 20, 2024

Deborah Duran, PhD • NIMHD
Elif Dede Yildirim, PhD • NIMHD
Mark Aronson, PhD • NIMHD





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 | ▪ Common Data Elements |
| ▪ Terra | ▪ AI readiness |
| ▪ Social Determinants of Health analytics | ▪ 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 Archive

February

Artificial Intelligence and Cloud Computing 101

-
-
-

Recent Think-a-thons

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

August

An Introduction to Python for Data Science

September

Data Management and Analysis in Python

October

ScHARe Data Repository

- 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

Findable: *providing metadata, documentation, and clear identifiers*

Accessible: *wide audience*

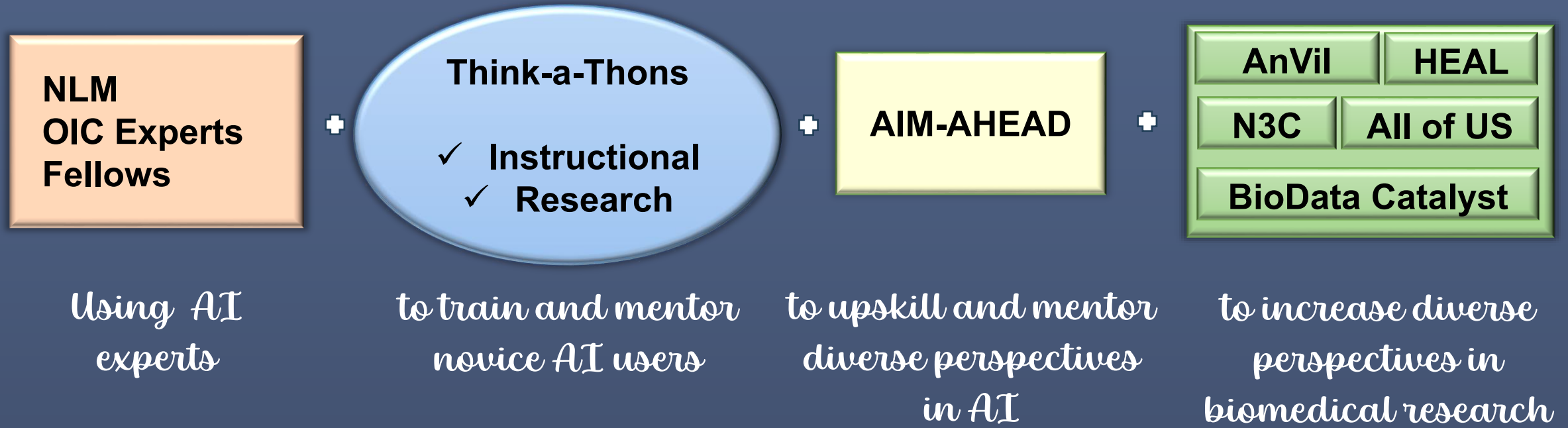
Interoperable: *standardized formats and APIs enable seamless integration*

Reusable: *clear documentation, licensing, reduce redundancy*

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



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

Slido Poll

Which statement is most true for Common Data Elements:

- a) Questions that are forced upon investigators to use
- b) Questions that are asked the same way for everyone involved in a study
- c) Questions that are universally semantically defined and can be used similarly by all
- d) Questions that require a series of questions and a set of responses only relevant to the study

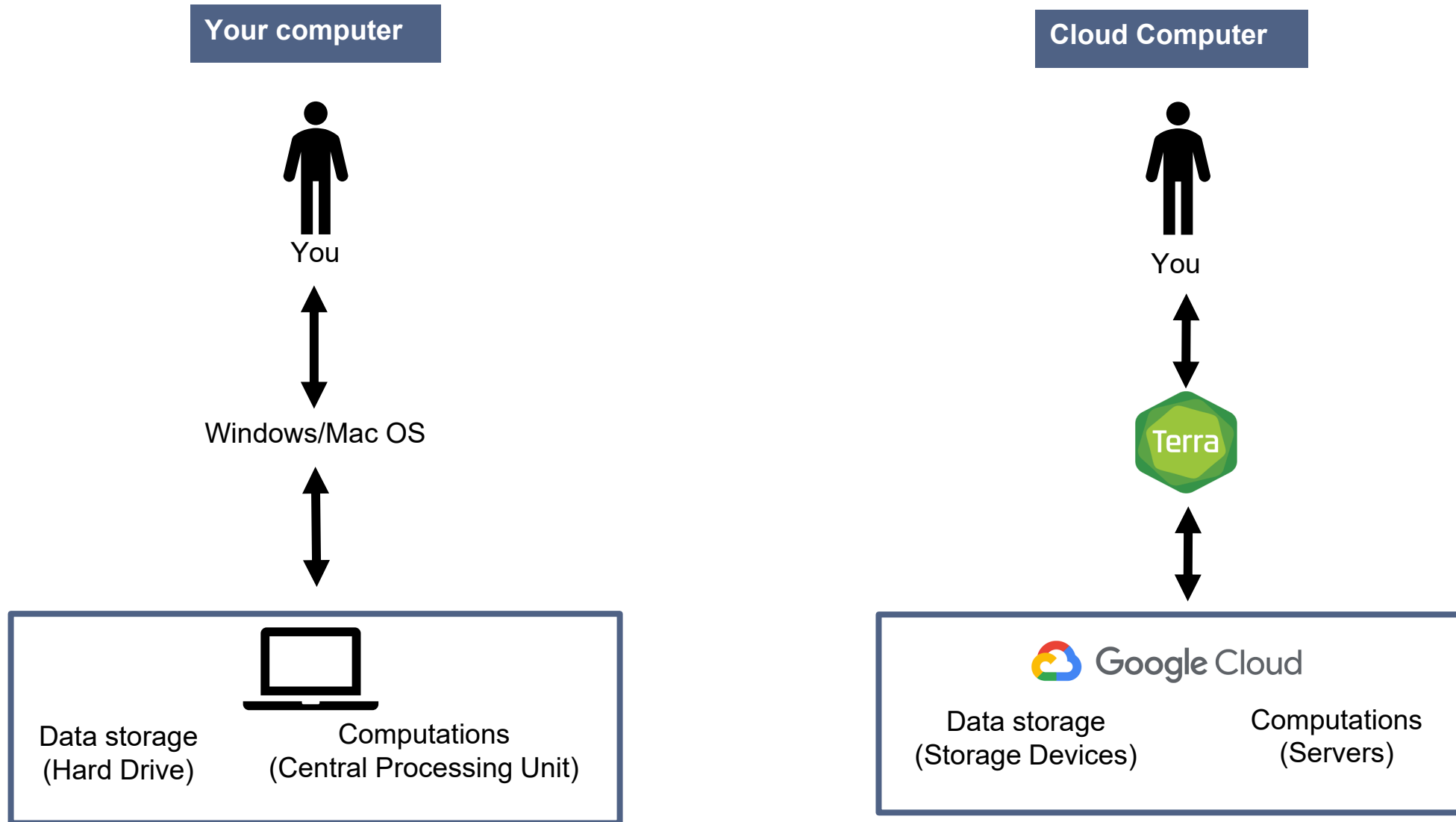
ScHARe

Getting Started on
Terra

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



ScHARe provides an easy portal to cloud computing



Today's mission

Getting Set Up on Terra

Register for ScHARe

A tour of the ScHARe Workspace

Your first steps

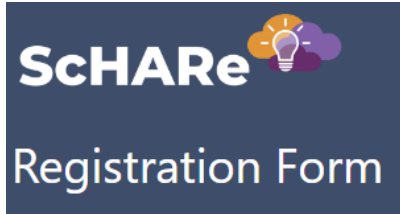
Create a Terra Account

Link a billing account

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account

Link a billing account

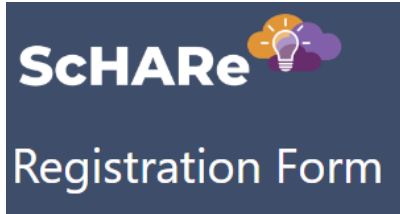
A tour of the ScHARe Workspace

Your first steps

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account

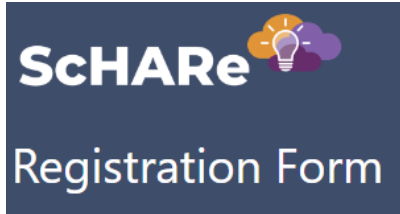
A tour of the ScHARe Workspace

Your first steps

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account



A tour of the ScHARe Workspace

Your first steps

Slido Poll

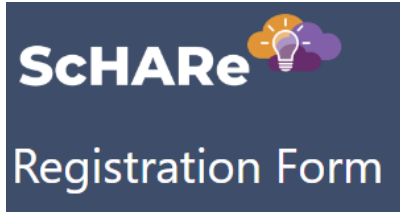
Which of the following do you NOT need to do before running your own data analysis on Terra?

- A. Link a Google Cloud Billing account to your Terra account
- B. Use an .edu email account to create a Terra account
- C. Use a Google account to register for SCHARE
- D. Create a Terra account

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account



A tour of the ScHARe Workspace

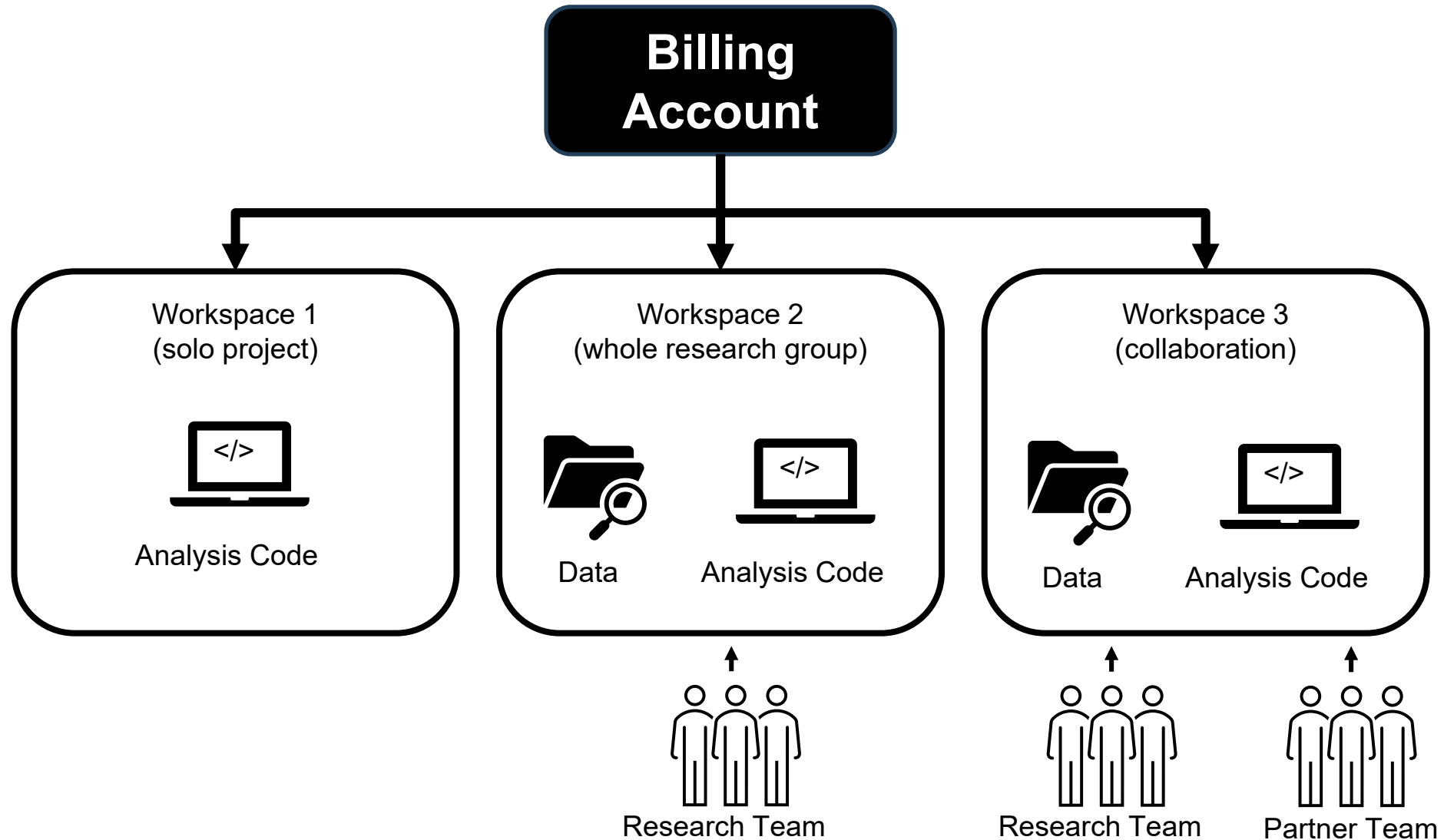
Terra Terminology

ScHARe Workspace: Data Tab

ScHARe Workspace: Analyses Tab

Your first steps

An intro to Terra Terminology



What costs money on Terra

Scenario 1: Analysis Code Only (pulling in data live)

Not actively running analysis

**Billing
Account**



Workspace 1
(solo project)



Analysis Code

No Cost

Actively running analysis

**Billing
Account**



Workspace 1
(solo project)



Analysis Code

Incurring Cost

Scenario 2: Analysis Code and Data (storing data in the workspace)

Not actively running analysis

**Billing
Account**



Workspace 2
(whole research group)



Data



Analysis Code

Incurring Cost

Actively running analysis

**Billing
Account**



Workspace 2
(whole research group)



Data



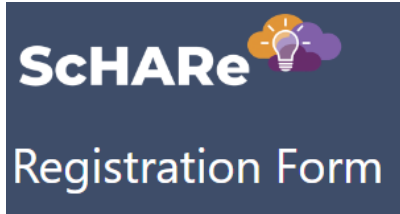
Analysis Code

Incurring Cost

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account



A tour of the ScHARe Workspace

Terra Terminology



ScHARe Workspace: Data Tab

ScHARe Workspace: Analyses Tab

Your first steps

Find public datasets on the ScHARe Data Tab

Datasets hosted on SCHARE Workspace are:

- public data
- contain no personal identifiable information (PII)

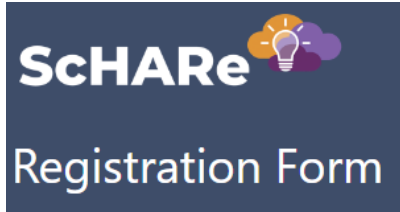
The screenshot displays the Terra WORKSPACES interface for the ScHARe Data Tab. The main table lists various datasets, including 'A_MainTable...', 'DiseaseAr...', 'EconomicStabili...', 'EducationAcces...', 'HealthBehaviors', 'HealthCareAcce...', 'MultipleCategor...', 'NeighborhoodA...', 'SocialAndComm...', and 'AdjustedGradu...'. The table columns are: A_MainT..., Categories, Data, DataDictionary, and FileFormat. The 'A_MainTable...' dataset is highlighted, showing 307 rows. The 'Categories' column for all datasets is 'Health Care Access and Quality'. The 'Data' column lists various data files, such as '20182021LongitudinalDataFileDat...', '2021FoodSecurityData.xlsx', '2021FullYearConsolidatedData.xlsx', '2021JobsFileData.xlsx', '2021MedicalConditionsData.xlsx', '2021PersonRoundPlanPublicUseD...', '2022FoodSecurityData.xlsx', '2022FullCharacteristicsData.xlsx', '2022FullYearConsolidatedData.xlsx', '2022JobsData.xlsx', '2022MedicalConditionsData.xlsx', '2022PersonRoundPlanPublicUseD...', and 'acqr-lea-sv2010-11.csv'. The 'DataDictionary' column lists corresponding codebooks, such as '20182021LongitudinalDataFileCod...', '2021FoodSecurityCodebook.pdf', '2021FullYearConsolidatedDataCod...', '2021JobsFileCodebook.pdf', '2021MedicalConditionsCodebook.p...', '2021PersonRoundPlanPublicUseco...', '2022FoodSecurityCodebook.pdf', '2022FullCharacteristicsCodebook.p...', '2022FullYearConsolidatedDataCod...', '2022JobsCodebook.pdf', '2022MedicalConditionsCodebook.p...', and '2022PersonRoundPlanPublicUseC...'. The 'FileFormat' column lists the file formats, such as 'XLSX' and 'CSV'. The interface also includes a sidebar with 'TABLES' and 'REFERENCE DATA' sections, a search bar, and a right sidebar with search and pricing options.

A_MainT...	Categories	Data	DataDictionary	FileFormat
20182021Lon...	Health Care Access and Quality	20182021LongitudinalDataFileDat...	20182021LongitudinalDataFileCod...	XLSX
2021FoodSec...	Health Care Access and Quality	2021FoodSecurityData.xlsx	2021FoodSecurityCodebook.pdf	XLSX
2021FullYearC...	Health Care Access and Quality	2021FullYearConsolidatedData.xlsx	2021FullYearConsolidatedDataCod...	XLSX
2021JobsFileD...	Health Care Access and Quality	2021JobsFileData.xlsx	2021JobsFileCodebook.pdf	XLSX
2021MedicalC...	Health Care Access and Quality	2021MedicalConditionsData.xlsx	2021MedicalConditionsCodebook.p...	XLSX
2021PersonRo...	Health Care Access and Quality	2021PersonRoundPlanPublicUseD...	2021PersonRoundPlanPublicUseco...	XLSX
2022FoodSec...	Health Care Access and Quality	2022FoodSecurityData.xlsx	2022FoodSecurityCodebook.pdf	XLSX
2022FullChara...	Health Care Access and Quality	2022FullCharacteristicsData.xlsx	2022FullCharacteristicsCodebook.p...	XLSX
2022FullYearC...	Health Care Access and Quality	2022FullYearConsolidatedData.xlsx	2022FullYearConsolidatedDataCod...	XLSX
2022JobsData	Health Care Access and Quality	2022JobsData.xlsx	2022JobsCodebook.pdf	XLSX
2022MedicalC...	Health Care Access and Quality	2022MedicalConditionsData.xlsx	2022MedicalConditionsCodebook.p...	XLSX
2022PersonRo...	Health Care Access and Quality	2022PersonRoundPlanPublicUseD...	2022PersonRoundPlanPublicUseC...	XLSX
AdjustedGradu...	Education Access and Quality	acqr-lea-sv2010-11.csv	acqr-sv10-11-public-file-document...	CSV

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account

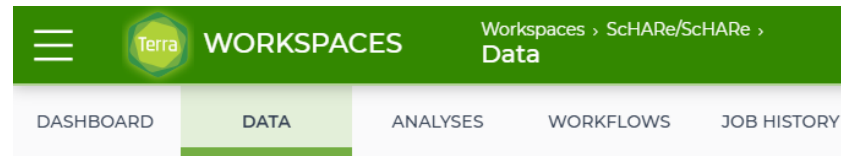


A tour of the ScHARe Workspace

Terra Terminology

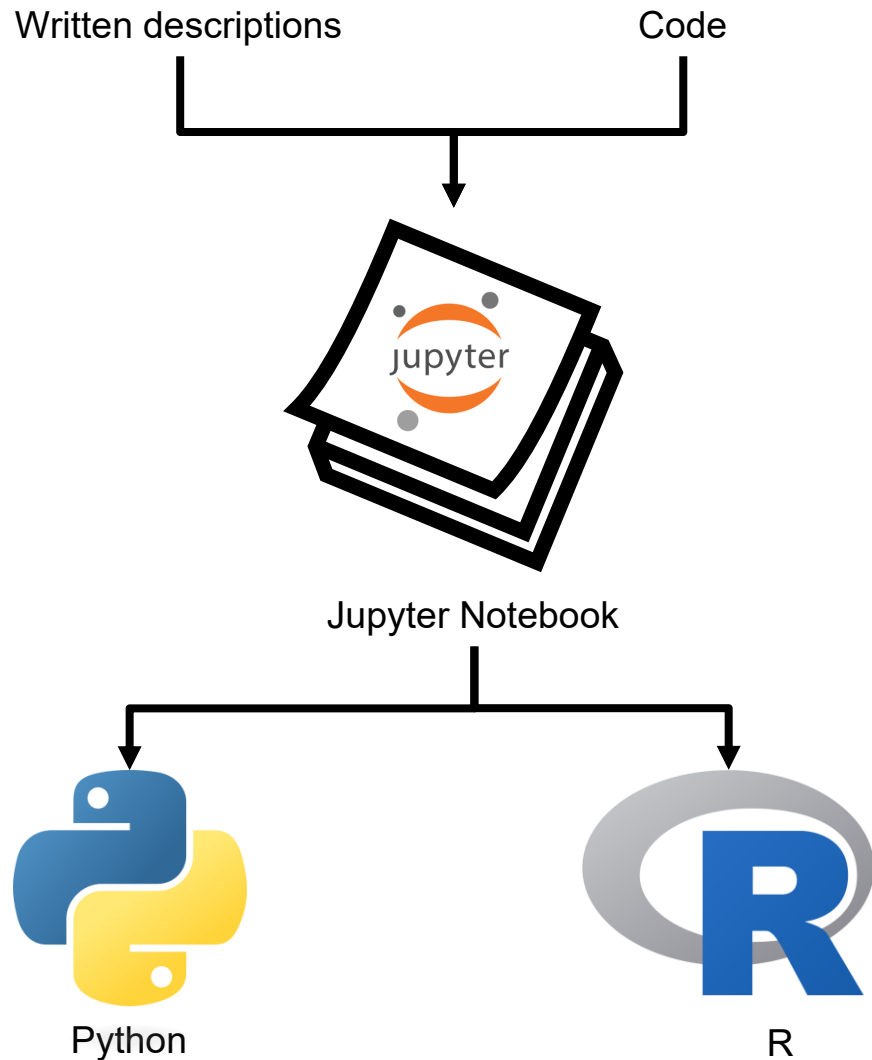


ScHARe Workspace: Data Tab

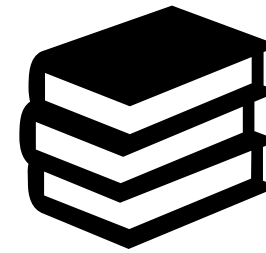


ScHARe Workspace: Analyses Tab

Data Analytics tools are found in the analyses tab



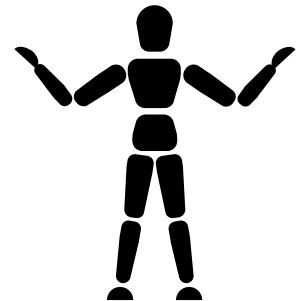
On the SCHARE Workspace



Reference
Notebooks



Tutorial
Notebooks

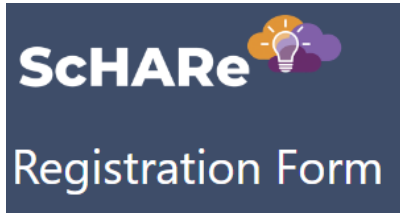


Model
Notebooks

Today's mission

Getting Set Up on Terra

Register for ScHARe



Create a Terra Account



Link a billing account

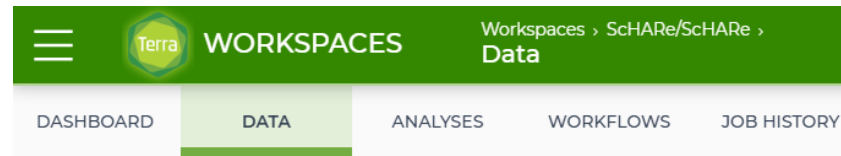


A tour of the ScHARe Workspace

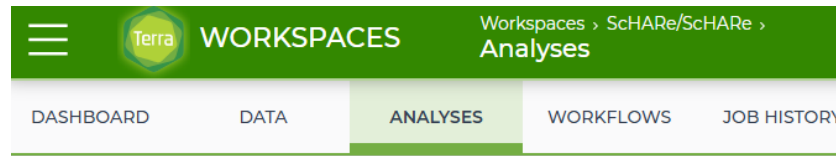
Terra Terminology



ScHARe Workspace: Data Tab



ScHARe Workspace: Analyses Tab



Slido Poll

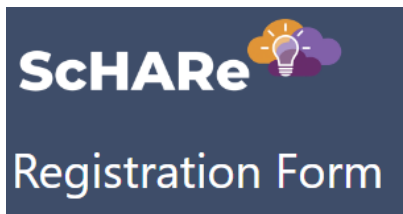
Which of the following is TRUE about the Data Tab in the SCHARE Terra Workspace?

- A. The datasets have Personal Identifiable Information in them
- B. The datasets are public data and are free to use
- C. The datasets require a Data Use Agreement to be signed to use
- D. The datasets are from untrustworthy sources

Today's mission

Getting Set Up on Terra

Register for SchARE



Create a Terra Account



Link a billing account

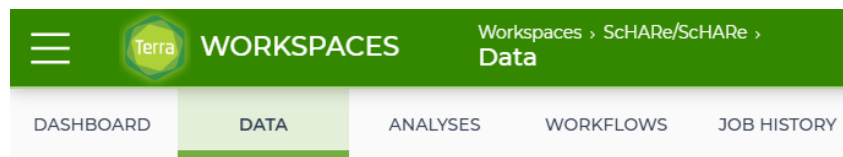


A tour of the SchARE Workspace

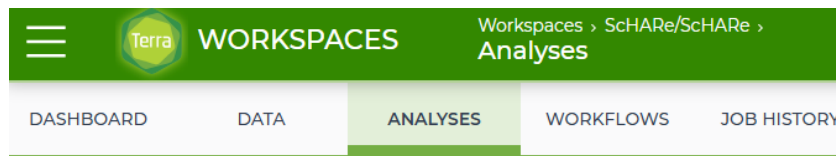
Terra Terminology



SchARE Workspace: Data Tab



SchARE Workspace: Analyses Tab



Your first steps

Creating a Workspace

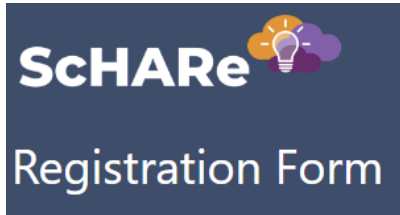
Pull data from the ACS

Simple analysis and visualization

Today's mission

Getting Set Up on Terra

Register for SchARe



Create a Terra Account



Link a billing account

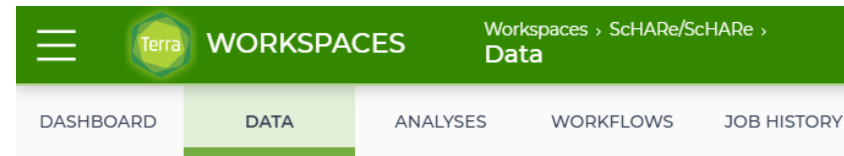


A tour of the SchARe Workspace

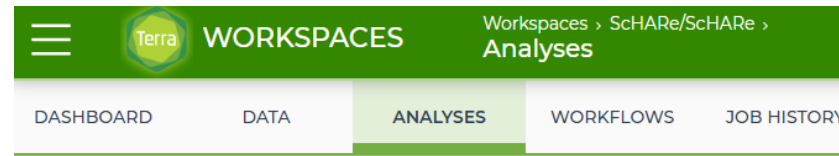
Terra Terminology



SchARe Workspace: Data Tab

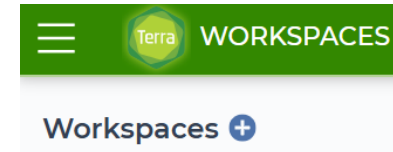


SchARe Workspace: Analyses Tab



Your first steps

Creating a Workspace

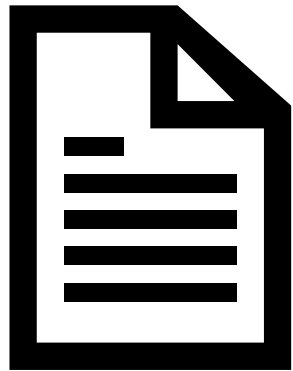


Pull data from the ACS

Simple analysis and visualization

Terra can take in data sources 3 ways

1. Upload your own data



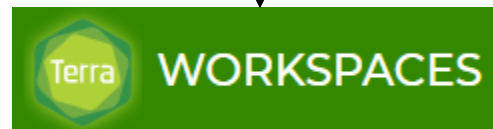
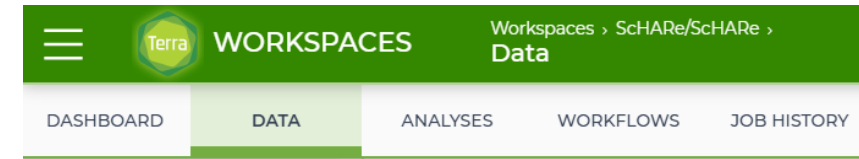
my_data.csv

2. Pull in Google-hosted datasets

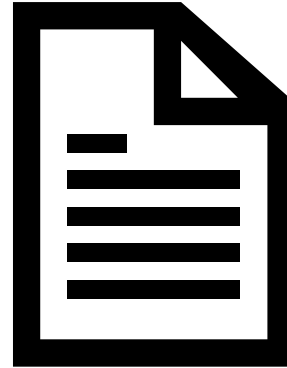


Google
Big Query

3. Use datasets on the SCHARE Workspace



Let's pull in some data from the Census



County Data (2023)

DASHBOARD

DATA

ANALYSES

WORKFLOWS

JOB HISTORY

Import Data

EditOpen With...ExportSettings

0 rows selected

TABLES

acs

28 results

14A_MainTable... (307)

0DiseaseAndCon... (34)

1EconomicStabi... (81)

0EducationAcces... (54)

0HealthBehavi... (20)

0HealthCareAcce... (44)

13MultipleCateg... (55)

0NeighborhoodAn... (11)

0SocialAndCommu... (8)

REFERENCE DATA

No references have been added.
Add reference data

A_MainTableDatasets_id

American_Community_Survey_2010

American_Community_Survey_2011

American_Community_Survey_2012

American_Community_Survey_2013

American_Community_Survey_2014

American_Community_Survey_2015

American_Community_Survey_2016

American_Community_Survey_2017

American_Community_Survey_2018

American_Community_Survey_2019

American_Community_Survey_2021

American_Community_Survey_2022

American_Community_Survey_2023

Atlas_of_Rural_and_Small_Town_America

Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Multiple Categories

Economic Stability

Data

ACS_County_2010.csv

ACS_County_2011.csv

ACS_County_2012.csv

ACS_County_2013.csv

ACS_County_2014.csv

ACS_County_2015.csv

ACS_County_2016.csv

ACS_County_2017.csv

ACS_County_2018.csv

ACS_County_2019.csv

ACS_County_2021.csv

ACS_County_2022.csv

ACS_County_2023.csv

RuralAtlasData24.xlsx

DataDictionary

ACS_2010_Variables.pdf

ACS_2011_Variables.pdf

ACS_2012_Variables.pdf

ACS_2013_Variables.pdf

ACS_2014_Variables.pdf

ACS_2015_Variables.pdf

ACS_2016_Variables.pdf

ACS_2017_Variables.pdf

ACS_2018_Variables.pdf

ACS_2019_Variables.pdf

ACS_2021_Variables.pdf

ACS_2022_Variables.pdf

ACS_2023_Variables.pdf

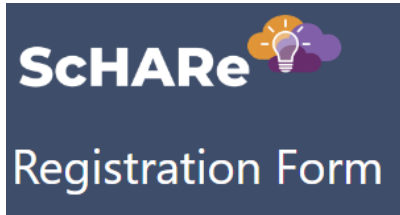
variableNameLookup.csv

Copy to clipboard

Today's mission

Getting Set Up on Terra

Register for SchARE



Create a Terra Account

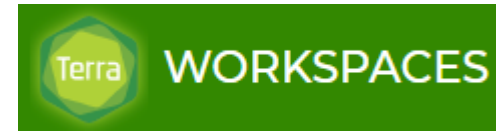


Link a billing account

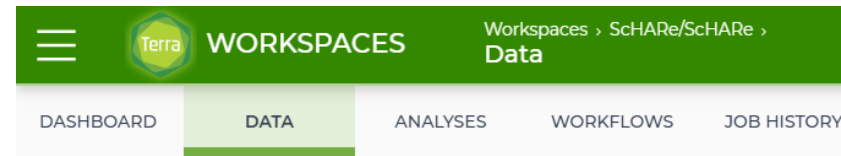


A tour of the SchARE Workspace

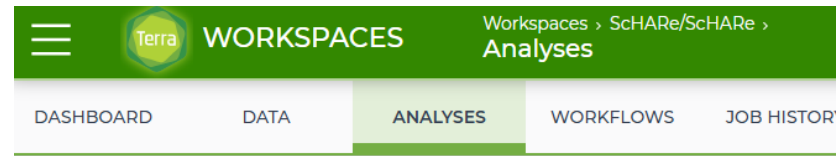
Terra Terminology



SchARE Workspace: Data Tab

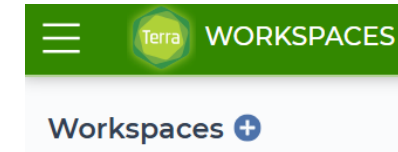


SchARE Workspace: Analyses Tab



Your first steps

Creating a Workspace

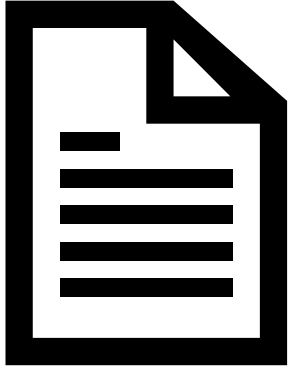


Get Census data

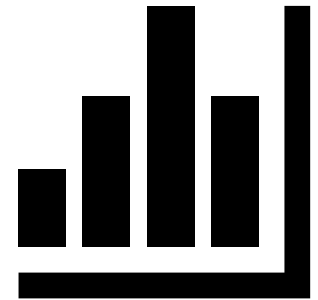


Simple analysis and visualization

Finally, let's visualize some of the data in Python



County Data (2023)



Median Income by County

Finally, let's visualize some of the data in Python

1. Copy the dataset from the SCHARE Workspace to your Workspace

```
!gsutil cp gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/General/ACS_County_2023.csv .
```

2. Bring in a library of code that helps us store the data

```
import pandas as pd
```

3. Load in the dataset from the Workspace

```
data = pd.read_csv('ACS_County_2023.csv')
```

4. Take a peek at the data to see if this worked

```
data.head(10)
```

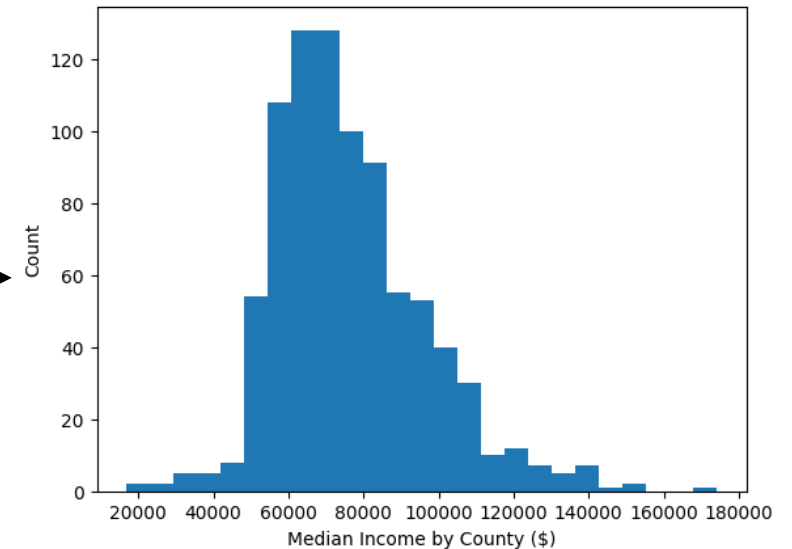
	GEO_ID	NAME	state	county	DP02_0001E	DP02_0002E	DP02_0003E	DP02_0004E	DP02_0005E	DP02_0006E	...	I
0	0500000US01003	Baldwin County, Alabama	1	3	105698	57455	16372	5658	2244	15984	...	
1	0500000US01015	Calhoun County, Alabama	1	15	46310	20559	7325	2114	1083	8399	...	
2	0500000US01043	Cullman County, Alabama	1	43	35961	19807	7399	2285	691	5868	...	
3	0500000US01049	DeKalb County, Alabama	1	49	27182	14118	5138	1991	963	4965	...	
4	0500000US01051	Elmore County, Alabama	1	51	33692	18114	5021	1563	404	5565	...	

5. Bring in a library of code for visualization

```
import matplotlib.pyplot as plt
```

6. Make a histogram of median household income by county

```
plt.hist(data['DP03_0062E'],25)
plt.xlabel('Median Income by County ($)')
plt.ylabel('Count')
```



from the codebook:

10/3/24, 8:47 PM		Census Data API: /data/2023/acs/1/profile/variables					
Name	Label	Concept	Required	Attributes	Limit	Predicate Type	Group
DP03_0061PE	Percent!!INCOME AND BENEFITS (IN 2023 INFLATION-ADJUSTED DOLLARS)!!Total households!!\$200,000 or more	Selected Economic Characteristics	not required	DP03_0061PEA , DP03_0061PM , DP03_0061PMA	0	float	DP03
DP03_0062E	Estimate!!INCOME AND BENEFITS (IN 2023 INFLATION-ADJUSTED DOLLARS)!!Total households!!Median household income (dollars)	Selected Economic Characteristics	not required	DP03_0062EA , DP03_0062M , DP03_0062MA	0	int	DP03

Putting it all together on Terra

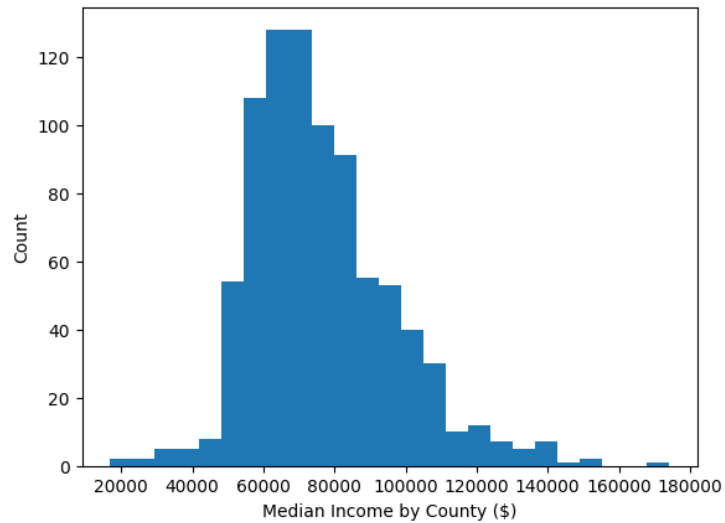
```
In [44]: # import packages for data storage
import pandas as pd

# copy dataset from SCHARE repository
!gsutil cp gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/General/ACS_County_2023.csv .

# read in dataset and preview
data = pd.read_csv('ACS_County_2023.csv')
data.head(10)
```

```
In [46]: # import package to visualize data
import matplotlib.pyplot as plt

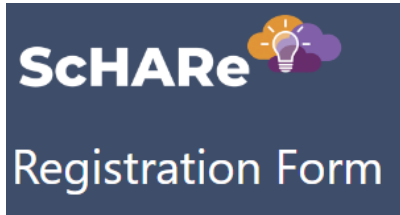
# make a histogram of median household income
plt.hist(data['DP03_0062E'],25)
plt.xlabel('Median Income by County ($)')
plt.ylabel('Count');
```



Today's mission

Getting Set Up on Terra

Register for SchARE



Create a Terra Account

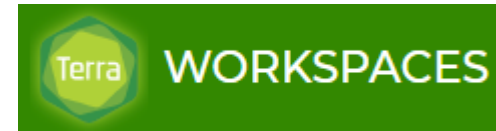


Link a billing account

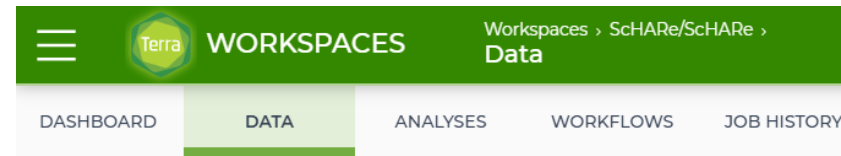


A tour of the SchARE Workspace

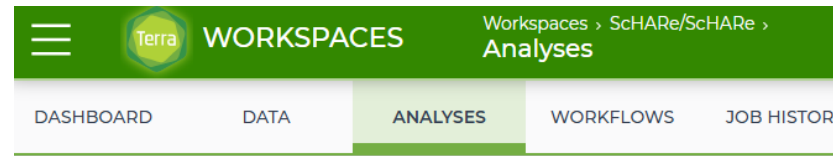
Terra Terminology



SchARE Workspace: Data Tab

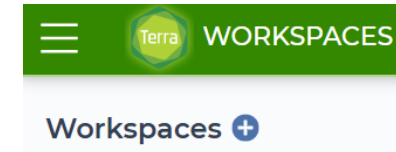


SchARE Workspace: Analyses Tab



Your first steps

Creating a Workspace



Pull data from the ACS



Simple analysis and visualization



Slido Poll

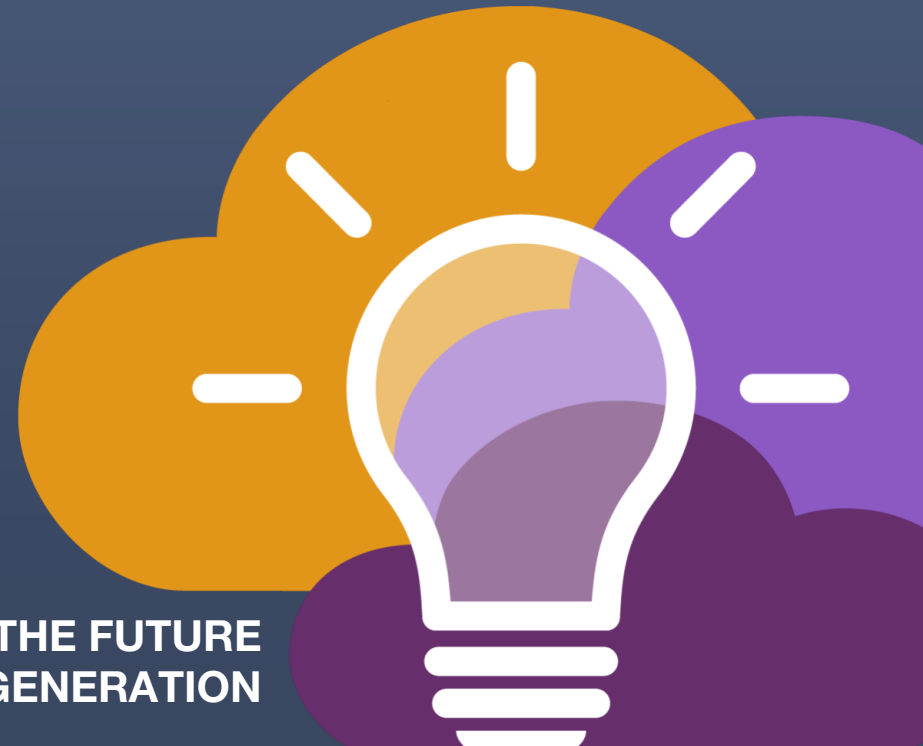
Which of the following is FALSE about the Data Tab in the SCHARE Terra Workspace?

- A. Terra can input your own data to use for analysis
- B. You can run your own analyses in the SCHARE Workspace
- C. Terra can input data from the SCHARE Workspace or Google
- D. Terra currently supports writing analysis code in R or Python

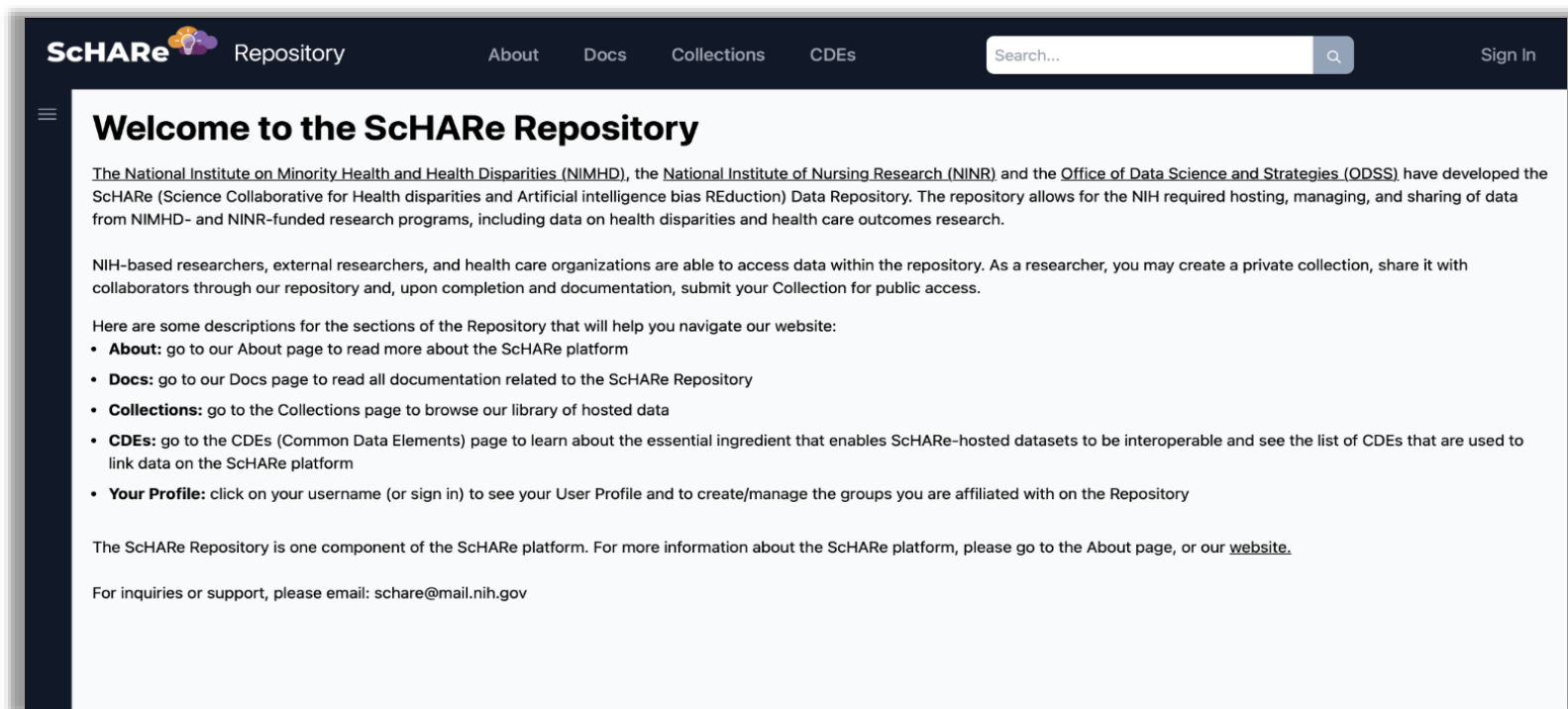
ScHARe

Repository
Background

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



ScHARe Repository



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



What can you do with the Repository?

Upload your own data

Store collected data and annotate with a data dictionary. Align data to the SchARE CDEs.

Harmonize data to CDEs

Map uploaded data to CDEs. Join your data with project or federated data via CDEs.

Browse for data

Find relevant federated national datasets or other project data.

Manipulate and aggregate data for analysis

Filter, sort, and select subsets for specific purposes. Link and aggregate datasets.

Control privacy levels and data sharing

Share confidential data among colleagues. Share public access data with the research community.

Data Analysis via SchARE Terra or local analysis platform



Key Features of the ScHARe Repository

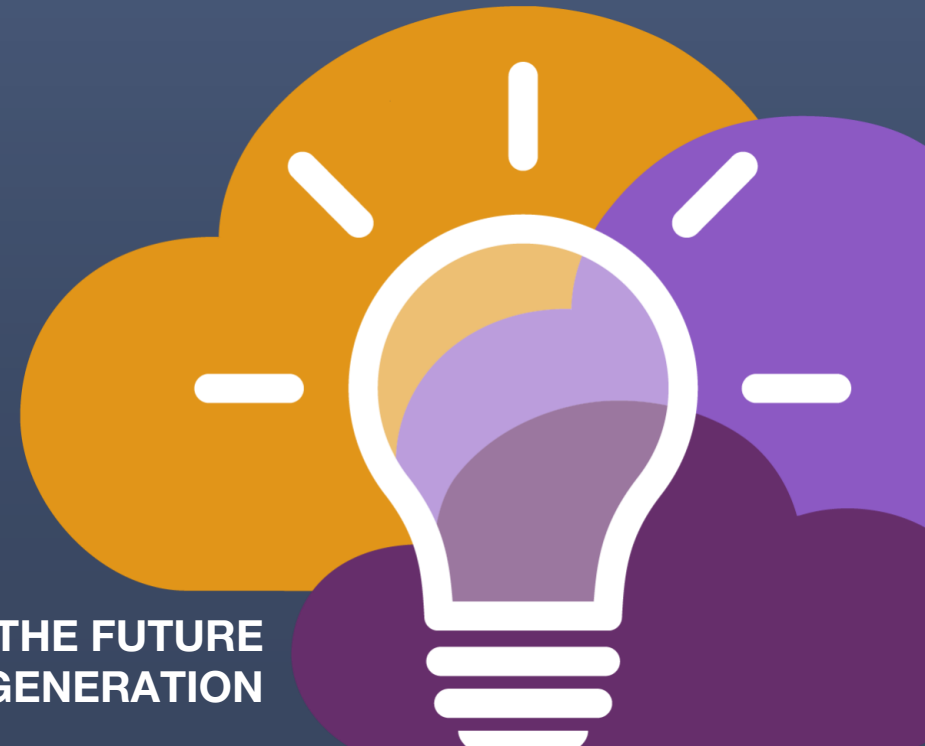
- Offers secure, long-term **storage** for project data
- Provides easy access to federated and other shared population science, Social Determinants of health (SDoH) **data sets**
- Leverages **CDEs** for organizing, aggregating, and harmonizing data
 - CDEs facilitate use of established semantic and coding systems
 - Aligns with ScHARe's focus on CDE centric data sharing
- Provides seamless tabular **data management**, aggregation, manipulation, and harmonization directly within the Repository
- Offers different **privacy levels** to meet individual data set requirements (public, restricted, confidential, private)
- Allows **sharing data** with other research collaborators



ScHARe

Getting Started

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Visit the ScHARe Repository!

<https://schare-repository.nimhd.nih.gov/>


For today's training, please use our test site:

<https://test-schare.nimhd.nih.gov/>


Data uploaded to this site will not be retained.




Log In to the Repository

 Repository

AboutDocsCollectionsCDEs

Search... 

Sign In



Welcome to the SchARE Repository


The [National Institute on Minority Health and Health Disparities \(NIMHD\)](#), the [National Institute of Nursing Research \(NINR\)](#) and the [Office of Data Science and Strategies \(ODSS\)](#) have developed the SchARE (Science Collaborative for Health disparities and Artificial intelligence bias REduction) Data Repository. The repository allows for the NIH required hosting, managing, and sharing of data from NIMHD- and NINR-funded research programs, including data on health disparities and health care outcomes research.

NIH-based researchers, external researchers, and health care organizations are able to access data within the repository. As a researcher, you may create a private collection, share it with collaborators through our repository and, upon completion and documentation, submit your Collection for public access.

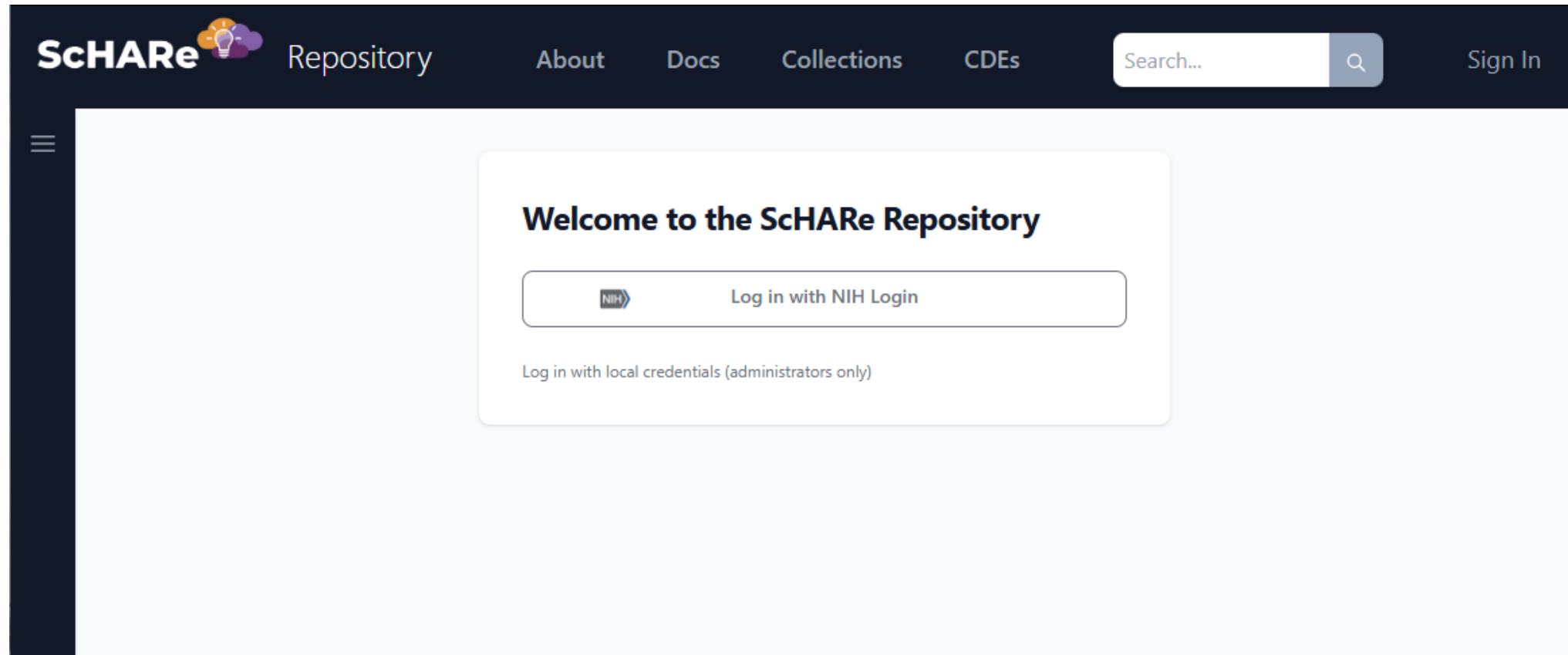
Here are some descriptions for the sections of the Repository that will help you navigate our website:

- **About:** go to our About page to read more about the SchARE platform
- **Docs:** go to our Docs page to read all documentation related to the SchARE Repository
- **Collections:** go to the Collections page to browse our library of hosted data
- **CDEs:** go to the CDEs (Common Data Elements) page to learn about the essential ingredient that enables SchARE-hosted datasets to be interoperable and see the list of CDEs that are used to link data on the SchARE platform
- **Your Profile:** click on your username (or sign in) to see your User Profile and to create/manage the groups you are affiliated with on the Repository

The SchARE Repository is one component of the SchARE platform. For more information about the SchARE platform, please go to the About page, or our [website](#).

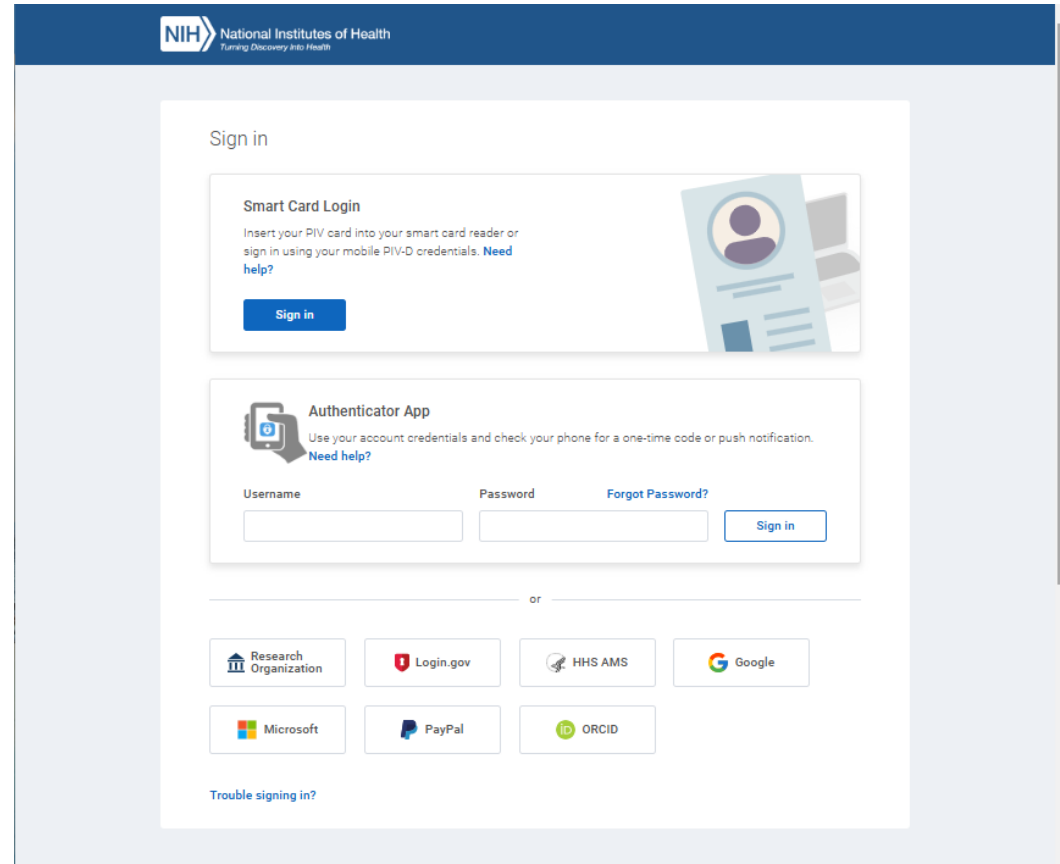


Log In to the Repository



Log In to the Repository

- Select the most appropriate login method
 - NIH Smart Card Login for NIH affiliates
 - Otherwise, Research Organization, Login.gov, HHS AMS preferred



The screenshot shows the NIH Sign in page. At the top is the NIH logo and the text "National Institutes of Health" and "Turning Discovery Into Health". Below this is a "Sign in" heading. The page features two main login sections. The first is "Smart Card Login", which includes instructions to insert a PIV card or use mobile PIV-D credentials, a "Need help?" link, and a "Sign in" button. The second is "Authenticator App", which includes instructions to use account credentials and check for a one-time code or push notification, a "Need help?" link, and a "Sign in" button. Below these sections are input fields for "Username" and "Password", and a "Forgot Password?" link. A horizontal line with "or" in the center separates these from a grid of social login buttons: "Research Organization", "Login.gov", "HHS AMS", "Google", "Microsoft", "PayPal", and "ORCID". At the bottom left of the grid is a "Trouble signing in?" link. A "WARNING NOTICE" is visible at the very bottom of the page.

NIH National Institutes of Health
Turning Discovery Into Health

Sign in

Smart Card Login
Insert your PIV card into your smart card reader or sign in using your mobile PIV-D credentials. [Need help?](#)
Sign in

Authenticator App
Use your account credentials and check your phone for a one-time code or push notification. [Need help?](#)
Username Password [Forgot Password?](#) **Sign in**

or


Research Organization Login.gov HHS AMS Google
Microsoft PayPal ORCID

[Trouble signing in?](#)

WARNING NOTICE



Log In to the Repository

 Sign in with Google

Choose an account

to continue to nih.gov




ScHARe Demo
scharedemo@gmail.com




Use another account



Log In to the Repository

 Sign in with Google

Sign in to nih.gov

 scharedemo@gmail.com ▼

By continuing, Google will share your name, email address, language preference, and profile picture with nih.gov. See nih.gov's Privacy Policy and Terms of Service.


You can manage Sign in with Google in your [Google Account](#).

Cancel

Continue



Log In to the Repository

 **National Institutes of Health**
Turning Discovery Into Health

NIH - Information Sharing Consent

Welcome `scharedemo@gmail.com`, (not `scharedemo@gmail.com`?
[Click here](#))

SCHAREgDE is requesting access to the following information:


- **Basic profile information**
First Name, Last Name, User Id and Email

By clicking agree, you allow this application to use your information in accordance to their terms of service and privacy policies. The information is needed to facilitate your access to this application. The information will not be shared with other applications and it will be retained for a limited time. [NIH Privacy Policy](#).


GrantDeny





Log In to the Repository


 Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

Terms of Use

To continue, you need to accept the terms of this website.

Terms and Conditions

Code of Conduct

Privacy Policy

Community Guidelines

SchARE Data Ecosystem and Research Collaboration Platform

Terms and Conditions


1. Acceptance of Terms

By accessing or using the SchARE Repository ("the Repository"), you agree to abide by these Terms and Conditions. These terms govern your use of the Repository's website, services, and tools, in addition to the Data Use Agreement (DUA) that governs the use of the publicly available data.


☐ I accept the above terms and conditions. *

Accept All


Reject





Log In to the Repository


 Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

Terms of Use

To continue, you need to accept the terms of this website.

Terms and Conditions ✓

Code of Conduct

Privacy Policy

Community Guidelines

ScHARe Data Ecosystem and Research Collaboration Platform

Terms and Conditions


1. Acceptance of Terms

By accessing or using the ScHARe Repository ("the Repository"), you agree to abide by these Terms and Conditions. These terms govern your use of the Repository's website, services, and tools, in addition to the Data Use Agreement (DUA) that governs the use of the publicly available data.


☒ I accept the above terms and conditions. *

Accept All


Reject





Log In to the Repository


 Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

Terms of Use

To continue, you need to accept the terms of this website.

Terms and Conditions ✓

Code of Conduct ✓

Privacy Policy

Community Guidelines

SchARE Data Ecosystem and Research Collaboration Platform

General Data Use Agreement


The SchARE General Data Use Agreement sets forth principles for responsible management and use of large-scale health disparities and population health data accessed through the SchARE Repository.

This agreement applies to all the data contained within the SchARE Repository, including publicly accessible data along with researcher-


☒ I accept the above code of conduct. *

Accept All


Reject





Log In to the Repository


ScHARe  Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

Terms of Use

To continue, you need to accept the terms of this website.

Terms and Conditions✓

Code of Conduct✓

Privacy Policy✓

Community Guidelines

ScHARe Data Ecosystem and Research Collaboration Platform

Privacy Policy

General Information


The ScHARe Privacy Policy adheres to the guidelines of both the [NIH Web Privacy Policy](#) and the [HHS Web Privacy Policy](#).

For site management, information may be collected for statistical purposes


☒ I accept the above privacy policy. *

Accept All


Reject





Log In to the Repository


ScHARe  Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

Terms of Use

To continue, you need to accept the terms of this website.

Terms and Conditions✓

Code of Conduct✓

Privacy Policy✓

Community Guidelines✓

ScHARe Data Ecosystem and Research Collaboration Platform


Community Guiding Principles

This Community Guiding Principles document outlines expectations of every member of the Science Collaborative for Health disparities and Artificial intelligence bias REduction (ScHARe) community, which includes institutions and individuals who use ScHARe resources, expertise, and data. The ScHARe community is committed to providing a welcoming and inspiring environment for all community members. Failure to abide by this

☒ I accept the above community guidelines. *

Accept All

Reject



ScHARe Repository

Learn more about the site

Browse for data
(private and public)

View CDE definitions

Search across all data that
you can see
(private and public)

Quick links to recently
accessed and your
own data

Log in, log out, and user
information

The screenshot shows the ScHARe Repository website interface. The top navigation bar includes links for 'About', 'Docs', 'Collections', and 'CDEs', along with a search bar and a user profile dropdown labeled 'scharedemo'. A left sidebar contains 'My Collections' and 'Starred' links. The main content area features a heading 'About ScHARe and the ScHARe Repository' followed by a paragraph describing the platform's development by NIMHD, NINR, and ODSS. Below this is a paragraph about the program's aim to increase participation in data science and cloud computing. A 'Mission' section follows, listing three bullet points: providing streamlined access to datasets, offering a project data repository for enhanced interoperability, and providing a data analytics and research collaboration platform. A decorative lightbulb icon is located in the bottom right corner of the slide.

ScHARe Repository

About Docs Collections CDEs Search... scharedemo

My Collections >

Starred >

About ScHARe and the ScHARe Repository

ScHARe (Science Collaborative for Health Disparities and Artificial intelligence bias REduction) is a cloud-based research collaboration platform developed by the [The National Institute on Minority Health and Health Disparities \(NIMHD\)](#), the [National Institute of Nursing Research \(NINR\)](#) and the [Office of Data Science and Strategies \(ODSS\)](#)

The aim of the ScHARe program is to increase participation of people from underrepresented populations in data science and cloud computing, so that all populations can benefit from the research opportunities afforded by Big Data.

An integral part of the program is the [ScHARe Think-a-Thons](#) webinar series, conceived to help prepare underrepresented researchers, students, and their collaborators to use the ScHARe platform and learn how to leverage Big Data and cloud computing. Participants share knowledge and skills, and form cross-disciplinary, multi-level collaborations around innovative research projects that can lead to breakthrough publications.

Mission

Here are some descriptions for the sections of the Repository that will help you navigate our website:

- Providing streamlined and centralized access to relevant datasets—including social determinants of health and other social science data.
- Offering a project data repository centered on core common data elements for enhanced data interoperability and compliance with the NIH Data Management and Sharing Policy.
- Providing a data analytics and research collaboration platform for biomedical researchers and their collaborators to access the same data and run analyses together in secure online spaces.
- Encouraging researchers to leverage Big Data and advanced artificial intelligence analytic tools.

ScHARe Repository - Docs



ScHARe Repository

About Docs Collections CDEs Search... scharedemo

My Collections >

Starred >

Welcome to the Repository Documents Library!

ScHARe Terms and Services

Community Guiding Principles

This Community Guiding Principles document outlines expectations of every member of the Science Collaborative for Health disparities and Artificial intelligence bias REduction (ScHARe) community, which includes institutions and individuals who use ScHARe resources, expertise, and data. The ScHARe community is committed to providing a welcoming and inspiring environment for all community members. Failure to abide by this Community Guiding Principles document may result in revocation of approved access to datasets obtained through the ScHARe ecosystem and access to the ScHARe platform.


These principles were established by the ScHARe Community to encourage and enable a highly collaborative research environment to address and reduce health disparities, improve health outcomes, and mitigate artificial intelligence bias. ScHARe Community members are expected to honor these guiding principles, be open to questions, reserve judgement, treat each other with respect, and foster engagement regardless of the career level or knowledge base.

Table of Contents

- Welcome to the Repository Documents Library!
- • ScHARe Terms and Services
- • Community Guiding Principles
 - • • Guiding Principles
 - • • Ethics Statement
 - • • Diversity and Inclusion Statement
 - • • Communication Etiquette
 - • • Keeping on Track
 - • • Audio and Visual Presence
 - • • Public and Private Messaging
 - • • Respect and Communicate team and personal boundaries.
 - • • Resolving Issues and Conveying Concerns
 - • General Data Use



ScHARe Repository - Collections

ScHARe 

Repository


About


Docs

Collections


CDEs


Search...



 scharedemo ▾

☰


 My Collections >

 Starred >

Collections


Create New Collection

My Collections



You don't have any collections yet - why not create a new one using the button above?

Shared with Me



You don't have any collections shared with you right now!

Public Collections

grouped by: Domains of Influence

Biological

Example NMHSS Analysis


LIVE

This contains data from the 2018 National Mental Health Services Survey (N-MHSS)...

7 months ago



ScHARe Repository - CDEs

ScHARe 


Repository



About

Docs


Collections


CDEs

Search... 

 scharedemo 

☰

 My Collections >

 Starred >

Common Data Elements

A major obstacle in advancing health disparities research is linking data between different studies. To surmount this, the ScHARe Repository makes use of a set of shared language, definitions, and an internal reference system associated with common categories in health disparities data. This system is known as Common Data Elements (CDEs), and the use of CDEs across the Repository enables the linking of datasets and a new generation of health disparities research.

The ScHARe Repository leverages the following CDE sets:

ScHARe 19 CDEs

ScHARe core common data elements

Age

Age >

number | 5 definitions | 1 concept | PDJBIGXjO:0001

Age Units >

string | 5 definitions | 1 concept | valueSet | GNVEdFQR0:0001

Birthplace

Birthplace - US >


string | 5 definitions | 3 concepts | 0qP44x2oX:0001

Birthplace - Country Outside US >

string | 4 definitions | 2 concepts | _qZulkipzA:0001



ScHARe Repository - User Profile

ScHARe 

Repository


About

Docs


Collections


CDEs

Search...

 scharedemo

☰

 My Collections

 Starred

Signed in as **scharedemo**

Sign out

Common Data Elements

A major obstacle in advancing health disparities research is linking data between different studies. To surmount this, the ScHARe Repository makes use of a set of shared language, definitions, and an internal reference system associated with common categories in health disparities data. This system is known as Common Data Elements (CDEs), and the use of CDEs across the Repository enables the linking of datasets and a new generation of health disparities research.

The ScHARe Repository leverages the following CDE sets:

ScHARe

19 CDEs

ScHARe core common data elements

Age

Age >

number | 5 definitions | 1 concept | PDjBiGXJO:0001

Age Units >

string | 5 definitions | 1 concept | valueSet | GNVoeDFQR0:0001

Birthplace

Birthplace - US >

string | 5 definitions | 3 concepts | 0qP44x2oX:0001

Birthplace - Country Outside US >


string | 4 definitions | 2 concepts | _qZulkipzA:0001

Postal Zip Code >


string | 5 definitions | 1 concept | eT7V_ETn_6:0001





ScHARe Repository - User Profile


 Repository


AboutDocsCollectionsCDEs

Search... 

 scharedemo ▾



 My Collections >

 Starred >

User Profile: ScHARe Demo

User Details

Username	scharedemo
User email	scharedemo@gmail.com
Subject ID	436625d1-348c-4144-8d9f-2ec00283bc1c
Account ID	a:9f51e3d7-40f0-4b9a-95c2-a5c7d483a9a3
First Login	10 minutes ago

Profile Visibility


Your user profile's visibility controls what others see when they search for you on this site or when they view data that you have published such as public collections. Your options are:

- **Public** - Your full name and username will be visible to all users, including users who are not logged in.
- **Unlisted** - Your profile will only be visible to users who are logged in and who are members of a group that you are a member of.

Visibility:



ScHARe Repository - User Profile

ScHARe 

Repository


About


Docs

Collections


CDEs


Search...



 scharedemo ▾

☰

 My Collections >

 Starred >

User Profile: ScHARe Demo

User Details

Username	scharedemo
User email	scharedemo@gmail.com
Subject ID	436625d1-348c-4144-8d9f-2ec00283bc1c
Account ID	a:9f51e3d7-40f0-4b9a-95c2-a5c7d483a9a3
First Login	11 minutes ago

Profile Visibility

Your user profile's visibility controls what others see when they search for you on this site or when they view data that you have published such as public collections. Your options are:

- **Public** - Your full name and username will be visible to all users, including users who are not logged in.
- **Unlisted** - Your profile will only be visible to users who are logged in and who are members of a group that you are a member of.

Visibility:

Unlisted ▾

Update profile visibility





ScHARe

ScHARe Repository Introduction

November 20, 2024

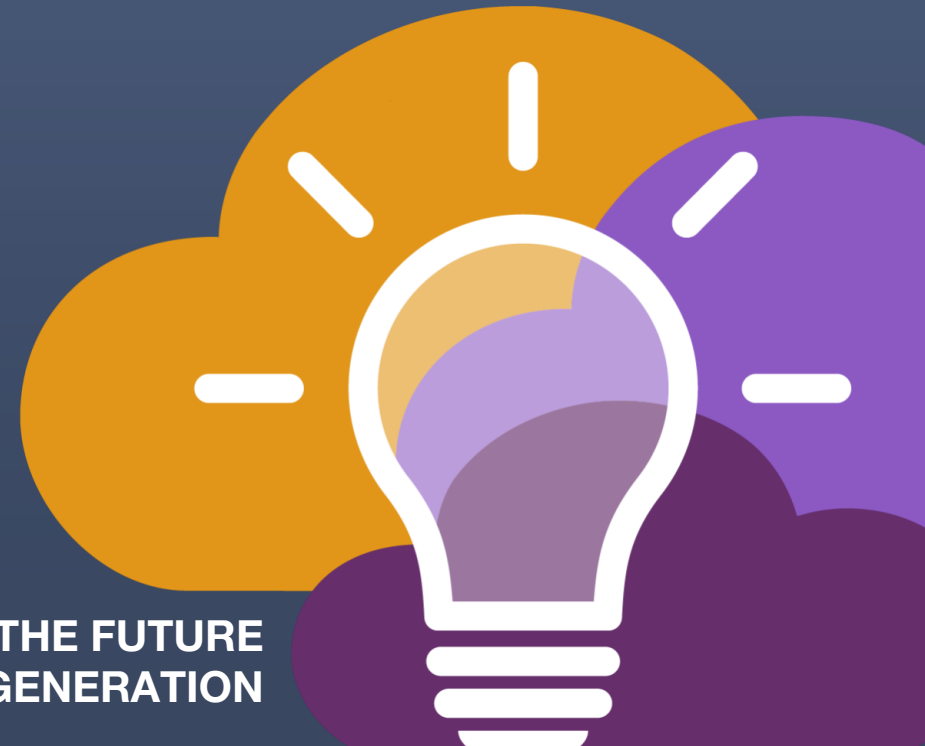
Deborah Duran, PhD • NIMHD
Elif Dede Yildirim, PhD • NIMHD
Mark Aronson, PhD • NIMHD



ScHARe

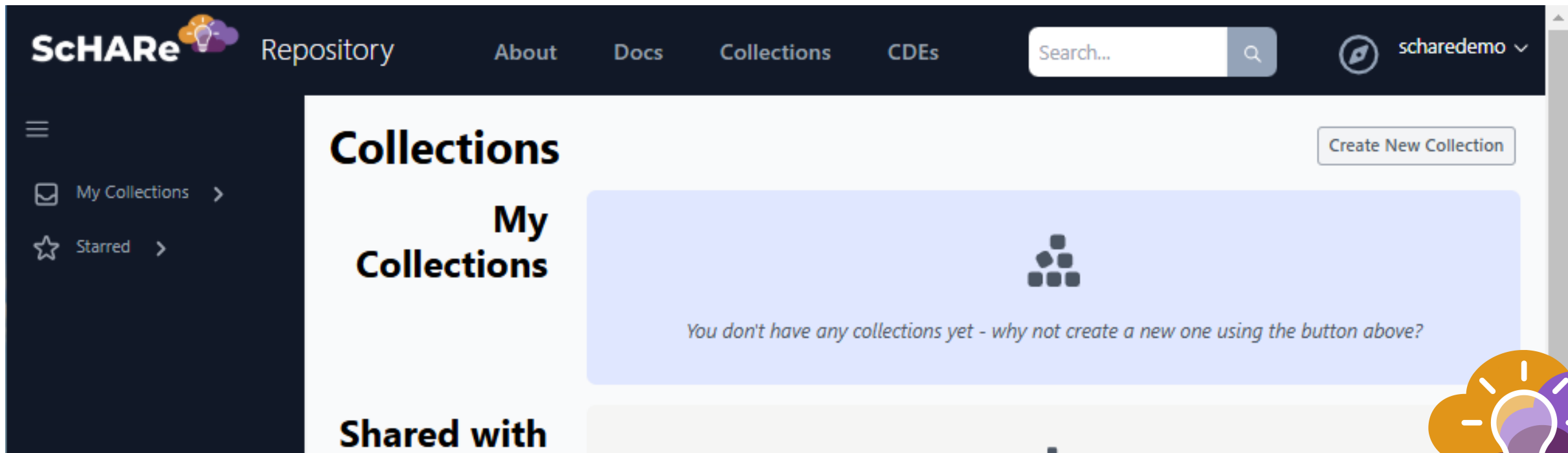
Creating a
Collection

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION




Creating a Collection

- **Collections** are a place where you can describe and store your data and any related metadata. They can be shared with colleagues and published when you're ready. You can upload any file type!
- The first step in uploading data is to create a collection



Creating a Collection

 Repository


About

Docs

Collections

CDEs

Search...

 scharedemo

Recent

My Collections

Starred

Create New Collection

Newly created collections are private by default

Collection Name

CDE Set

SchARE (19 CDEs)

Abstract

Domains of Influence

Select all that apply...


- Provide a readable name and a brief abstract (description) of the data to be uploaded
- These details can be updated later if needed



Creating a Collection



Recent >

 My Collections >

☆ Starred >

Create New Collection

Newly created collections are private by default >



Collection Name

demo-files

CDE Set ①

ScHARe (19 CDEs)




Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Enter a Tag

Project CDEs 

Levels of Influence

Select all that apply...



Domains of Influence

Select all that apply...



Research Areas

Select all that apply...



Creating a Collection

ScHARe

Repository

About

Docs

Collections

CDEs

Search...

scharedemo

Recent

My Collections

Starred

Common Data Elements

A major obstacle in advancing health disparities research is linking data between different studies. To surmount this, the ScHARe Repository makes use of a set of shared language, definitions, and an internal reference system associated with common categories in health disparities data. This system is known as Common Data Elements (CDEs), and the use of CDEs across the Repository enables the linking of datasets and a new generation of health disparities research.

The ScHARe Repository leverages the following CDE sets:

ScHARe

19 CDEs

ScHARe core common data elements

Age

Age >

number | 5 definitions | 1 concept | PDjBiGXjO:0001

Age Units >

string | 5 definitions | 1 concept | valueSet | GNVoeDFQR0:0001

Birthplace

Birthplace - US >

string | 5 definitions | 3 concepts | 0qP44x2oX:0001

Birthplace - Country Outside US >

string | 4 definitions | 2 concepts | _qZu

Tags

My Collections >

Starred >

Collection Name

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Enter a Tag

Advanced User-Defined Metadata >

Submit

Cancel

Adolescents

Age

Alcohol Use

American Indian or Alaska Native

Asian or Asian American

Black or African American

Children

Chronic Diseases

Clinical Trial Data

Community Health Programs

County-Level Data

Data Science: AI/ML/LLM

Data from Electronic Health Records (EHRs)

Diet and Nutrition

Disability

Discrimination and Racism

Select all that apply...

Select all that apply...

Select all that apply...

Select all that apply...

Apply relevant tags and project-level CDEs to your dataset



Tags

My Collections >

Starred >

Collection Name

demo-files

CDE Set ⓘ

SCHARE (19 CDEs)

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Cross-Sectional Study x

State-Level Data x

Enter a Tag

Project CDEs ⓘ

Levels of Influence

Select all that apply...

Domains of Influence

Select all that apply...

Research Areas

Select all that apply...

Research Focus

Select all that apply...

Advanced User-Defined Metadata >

Submit

Cancel



Project-Level CDEs: Levels of Influence

My Collections >

Starred >

Collection Name

demo-files

CDE Set ⓘ

SCHaRe (19 CDEs)

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Cross-Sectional Study x

State-Level Data x

Enter a Tag

Project CDEs ⓘ

Levels of Influence

Domains of Influence

Research Areas

Research Focus

Select all that apply...

Individual

Interpersonal

Community


Societal

Select all that apply...

Advanced User-Defined Metadata >

Submit

Cancel



Project-Level CDEs: Domains of Influence

☆ Starred >

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Cross-Sectional Study x

State-Level Data x

Enter a Tag

Project CDEs ⓘ

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

Research Areas

Research Focus

Biological

Behavioral

Physical/Built Environments

✓ Sociocultural Environment

Health Care Systems and Clinical Care

Advanced User-Defined Metadata >

Submit

Cancel



Project-Level CDEs- Research Areas

☆ Starred >

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Cross-Sectional Study x

State-Level Data x

Enter a Tag

Project CDEs ⓘ

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

Research Areas

Select all that apply...

Research Focus

Minority health study focused on a one race or ethnic population and not addressing a health disparity.

Health Disparity Outcomes

Other Health Outcomes / Healthcare Delivery

Advanced User-Defined Metadata >

Submit

Cancel



Project-Level CDEs: Research Focus

☆ Starred >

Abstract

November 20 Think-a-Thon Demonstration Files

Tags

Cross-Sectional Study x

State-Level Data x

Enter a Tag

Advanced User-Defined Metadata >

Submit

Cancel

Project CDEs ⓘ

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

Research Areas

Select all that apply...

Research Focus

Select all that apply...

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



ScHARe 

About

Docs

Collections

CDEs

Search

scharedemo ▾

i

Collection was successfully created.

✕



Operations ▼

Abstract

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

➤ **Links and Documents**

> Data Items

- Metadata and Statistics

Access Level ①



Private

Analysis Readiness



Not Ready >

CDE Compliance -

ScHARe ①



0 / 19 CDEs assigned



The Collection Main Page

Click the collection name to return to the main page

Actions on the collection itself (edit metadata, share, delete)

Browse and upload documentation

Sidebar with key information

Browse and upload data items

SchARE Repository About Docs Collections CDEs Search... scharedemo

scharedemo / **demo-files** / LIVE Operations

demo-files

Abstract
November 20 Think-a-Thon Demonstration Files

Levels of Influence: Individual Domains of Influence: Sociocultural Environment

Links and Documents

No links or documents have been uploaded yet.

Add New

Data Items

Metadata and Statistics

Access Level
Private

Analysis Readiness
Not Ready

CDE Compliance - SchARE
0 / 19 CDEs assigned

Tags
Topics tagged in this collection
Cross-Sectional Study
State-Level Data



Slido Poll

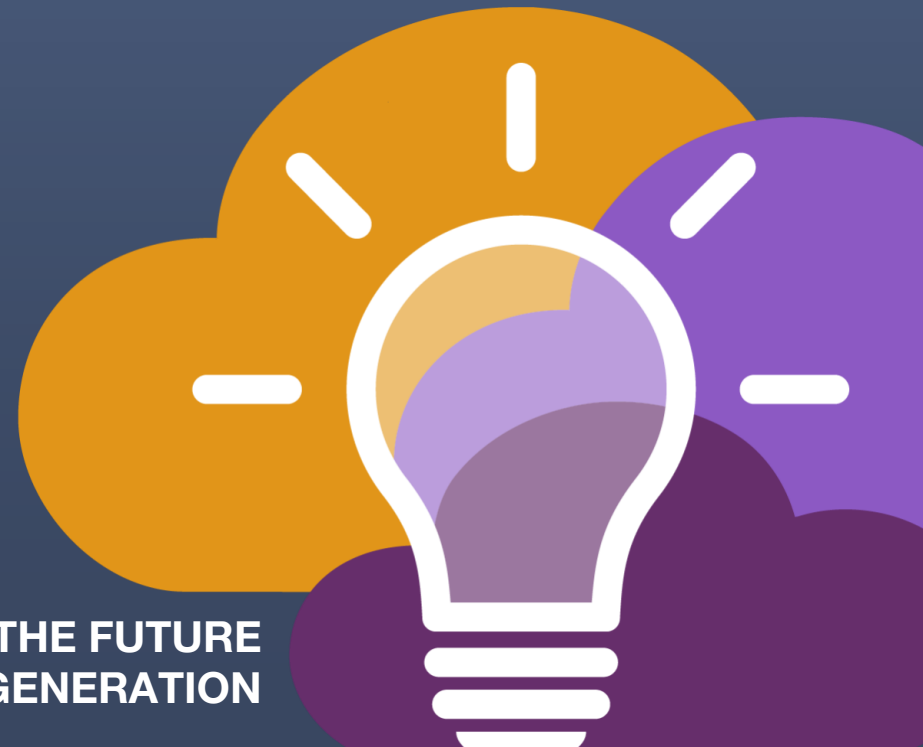
Which of the following data types are you most comfortable sharing publicly?

- a) Health data
- b) Environmental data
- c) Behavioral data
- d) Other

ScHARe

Uploading your
first Data Set

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Uploading a Data File

Upload any document, presentation, etc.

Link to papers through PubMed

The screenshot displays the Schare demo interface. At the top, the breadcrumb navigation shows 'scharedemo / demo-files / LIVE'. On the right, there are icons for a link, a star, and an 'Operations' dropdown menu. The main content area is titled 'demo-files' and includes an 'Abstract' section with the text 'November 20 Think-a-Thon Demon' and 'Levels of Influence Individual'. Below this, there is a section for 'Links and Documents' with an 'Add New' button. A modal window titled 'Add Document or Link' is open in the center. It contains three buttons: 'Upload', 'Pubmed Search', and 'Link'. Below these is a 'TYPE' dropdown menu with the text 'Select one...'. The 'UPLOAD' section has a 'Choose File' button and the text 'No file chosen'. There are input fields for 'TITLE' (with the placeholder 'Title...') and 'DESCRIPTION'. At the bottom of the modal are 'Add' and 'Cancel' buttons. On the right side of the interface, there are several status sections: 'Access Level' (Private), 'Analysis Readiness' (Not Ready), 'CDE Compliance - SchARe' (0 / 19 CDEs assigned), and 'Tags' (Topics tagged in this collection).

scharedemo / demo-files / LIVE

demo-files

Abstract
November 20 Think-a-Thon Demon
Levels of Influence
Individual

Links and Documents

Add New

Data Items

Metadata and Statistics

Add Document or Link

Upload Pubmed Search Link

TYPE
Select one...

UPLOAD
Choose File No file chosen

TITLE
Title...

DESCRIPTION

Add Cancel

Access Level ①
Private

Analysis Readiness
Not Ready

CDE Compliance - SchARe ①
0 / 19 CDEs assigned

Tags
Topics tagged in this collection
Cross-Sectional Study
State-Level Data

Uploading a Data File

You can upload to the SchARe Repository...

Any file

- Documents (PDF, docx, etc.)
- Archives (zip, etc.)

Some file formats can be previewed in the web interface

Anything uploaded can be downloaded

Tabular data

- CSV and TSV
- Excel (xlsx, xls)
- SAS (XPORT V5/6, not V8)
- SPSS
- STATA
- Parquet

Tabular data can be viewed in the web interface, plus much more...



Uploading a Data File

The screenshot displays the SchARE Repository interface. The top navigation bar includes the SchARE logo, a search bar, and links for About, Docs, Collections, and CDEs. The user is logged in as 'scharedemo'. The main content area shows the 'demo-files' collection, which is currently in 'LIVE' status. The collection's abstract describes 'November 20 Think-a-Thon Demonstration Files'. It lists 'Levels of Influence' as 'Individual' and 'Domains of Influence' as 'Sociocultural Environment'. A large dashed box with an upload icon and the text 'Drag and Drop or Browse Files to Upload' is prominently displayed. To the right, a sidebar provides details on 'Access Level' (Private), 'Analysis Readiness' (Not Ready), 'CDE Compliance - SchARE' (0 / 19 CDEs assigned), and 'Tags' (Cross-Sectional Study, State-Level Data).

SchARE Repository

About Docs Collections CDEs Search... scharedemo

scharedemo / **demo-files** / LIVE

demo-files

Abstract
November 20 Think-a-Thon Demonstration Files

Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

> Links and Documents

▼ Data Items

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ
Private

Analysis Readiness
✕ Not Ready >

CDE Compliance - SchARE ⓘ
✕ 0 / 19 CDEs assigned

Tags
Topics tagged in this collection

Cross-Sectional Study

State-Level Data



Uploading a Data File



☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

 Operations ▾

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

> Links and Documents

> Data Items

Drag and Drop or [Browse Files](#) to Upload

exa...csv


×

Upload Files


Cancel All

> Metadata and Statistics


Access Level ⓘ

 Private

Analysis Readiness

 Not Ready >

CDE Compliance - SchARe ⓘ


 0 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data



Uploading a Data File

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

🔗 ⌛ ☆ Operations ▾

demo-files

Abstract
November 20 Think-a-Thon Demonstrati

Levels of Influence
Individual

Don
Soci

> Links and Documents

▼ Data Items

STATUS	NAME	SIZE
📄	example_data.csv	25 KB

Page 1 of 1 >

📁

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ
🔒 Private

Analysis Readiness
⊗ Not Ready >

CDE Compliance - SchARe ⓘ
⊗ 0 / 19 CDEs assigned


Tags
Topics tagged in this collection
[Cross-Sectional Study](#) [State-Level Data](#)

File Upload Complete

The SchARe Repository uses the SchARe common data elements to enable data discovery, linking, and interoperability. To make your data more interoperable and ready for publication, we will guide you through the process of assigning common data elements to your data column headings. Click OK to continue.

TIP: You can cancel now and complete the process later using the "Configure CDEs" tool from the collection Operations menu.

OK Cancel



ScHARe

CDE Mapping: Option 1

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Getting Started with Mapping CDEs

- The Repository does not *enforce* CDEs, but it does *encourage* using CDEs
- Start with the data you have, and use the tools the repository offers to map as much of your data to CDEs as possible

Assign CDEs at Upload

The screenshot shows a modal window titled "Assign CDEs to Uploaded Data Columns" with the instruction "Assign your uploaded item's data columns to their appropriate common data elements." The "CDE set selected: SchARE" is indicated. The dialog contains a table for mapping:

SCHARE CDE	UPLOADED DATA COLUMNS	DATA VALID
Age		
Age	Age	✓
Age Units		
Birthplace		
Birthplace - US		
Birthplace - Country Outside US		
Postal Zip Code		
Race/Ethnicity Self-Identification		
Sex at Birth		
Gender		
Gender		

Buttons at the bottom: Assign, More Options, Cancel.

Map CDEs using Dataviews

The screenshot shows the "Map CDEs using Dataviews" interface. The "Source data from:" is set to "test_data.xlsx". The "Map Into" section shows a mapping for "Race/Ethnicity Self-Identification".

SOURCE VALUE	TARGET VALUE
Black	Black or African American
African American	Black or African American
American Indian	American Indian or Alaska Native
Alaska Native	American Indian or Alaska Native

Target Column: race_mapped. Options: New column (selected), Replace source column.

When source value is not found in map, target value is: ☒ NULL, ☐ Source value, ☐ Constant: value.



Assigning CDEs at Upload

☰

Recent >

My Collections >

★ Starred >

scharedemo / demo-files / LIVE

🔗 ⌚ ☆ Operations ▾

demo-files

Abstract

November 20 Think-a-Thon Demonstrati

Levels of Influence

Individual

Don

Soci

> Links and Documents

▼ Data Items

STATUS	NAME	SIZE
📄	example_data.csv	25 KB

<< < Page 1 of 1 > >>

📁

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Match CDEs with column headings

The first step is to automatically detect alignments between your data and the SchARE CDE set. Unless otherwise indicated, please select SchARE below.

Select the CDEs you want to assign:

SchARE (19 CDEs)

CDE matching method:

Using column headers

Next Cancel

Access Level ⓘ

🔒 Private

Analysis Readiness

⊗ Not Ready >

CDE Compliance - SchARE ⓘ

⊗ 0 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data

💡

Assigning CDEs at Upload

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

Operations

demo-files

Abstract

November 20 Think-a-Thon Demonstrati

Levels of Influence

Individual

Don

Soci

> Links and Documents

> Data Items

STATUS	NAME	SIZE
	example_data.csv	25 KB

<< Page 1 of 1 >>

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Match CDEs with column headings

The first step is to automatically detect alignments between your data and the SchARE CDE set. Unless otherwise indicated, please select SchARE below.

Select the CDEs you want to assign:

SchARE (19 CDEs)

CDE matching method:

Using column headers

✓ Using column headers

Manual assignment (no search)

Access Level ⓘ

Private

Analysis Readiness

Not Ready >

CDE Compliance - SchARE ⓘ

0 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data

Assigning CDEs at Upload

Some of the assignments will be correct -- double-check the validation and the column name matches what you expect

Complete or cancel the process, or use "More Options" to go to a detailed tool for CDE assignment

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**

SCHARE CDE	UPLOADED DATA COLUMNS	DATA VALID
Age		
123 Age	age ↵	✓
Age Units	age_units ↵	✓
Birthplace		
Birthplace - US		✓
Birthplace - Country Outside US		✓
Postal Zip Code	zip_code ↵	✓
Race/Ethnicity Self-Identification	race_ethnicity_1 ↵	✗
Sex at Birth	sex_at_birth ↵	✓
Gender		
Gender	gender ↵	✗

Assign

More Options

Cancel

Some of the assignments will be wrong or missing -- add anything that should be added, and double-check the validation



Assigning CDEs at Upload

Recent >

My Collections >

Starred >

scharedemo / de

demo

Abstract

November

Levels of Individual

Links a

Data It

STAT

Metadata and Statistics

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**

Postal Zip Code	zip_code	✓
Race/Ethnicity Self-Identification	race_ethnicity_1	✗
Sex at Birth		
Gender		
Gender - Select Other		
Gender - Specify		
Sexual Orientation		
Sexual Orientation - Select Other		
Sexual Orientation - Specify		
Marital Status	marital_status	✓
Education		

Assign More Options Cancel

Row 1 "no" is not one of ["American Indian or Alaska N...

Row 2 "yes" is not one of ["American Indian or Alas...

Row 3 "no" is not one of ["American Indian or Alask...

Row 4 "no" is not one of ["American Indian or Alask...

Row 5 "yes" is not one of ["American Indian or Alas...

Row 6 "no" is not one of ["American Indian or Alask...

Row 7 "no" is not one of ["American Indian or Alask...

Row 8 "no" is not one of ["American Indian or Alask...

Row 9 "no" is not one of ["American Indian or Alask...

Row 10 "yes" is not one of ["American Indian or Ala...

Row 11 "no" is not one of ["American Indian or Alas...

When an assigned CDE fails validation, hover over the red X to get details on why

NOTE: CDE validation errors can be corrected using dataviews (coming up later)

Assigning CDEs at Upload

The screenshot shows a web application interface with a dark sidebar on the left containing navigation links: 'Recent', 'My Collections', and 'Starred'. The main content area is titled 'scharedemo / de' and displays a list of items including 'Abstract', 'November', 'Levels of Individual', 'Links a', and 'Data It'. A modal dialog titled 'Assign CDEs to Uploaded Data Columns' is open in the center. The dialog has a subtitle 'Assign your uploaded item's data columns to their appropriate common data elements.' and a 'More Info' link. It shows the 'CDE set selected: SchARe'. The dialog contains a table with columns for data column names, dropdown menus for selecting CDEs, and status indicators (green checkmarks for assigned, red X for unassigned). A dropdown menu is open for 'race_ethnicity_1', showing a list of available CDEs. At the bottom of the dialog are three buttons: 'Assign', 'More Options', and 'Cancel'. In the bottom right corner of the image, there is a decorative graphic of a lightbulb inside a cloud.

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**

Uploaded Data Column	Assigned CDE	Status
Postal Zip Code	zip_code	✓
Race/Ethnicity Self-Identification	race_ethnicity_1	✗
Sex at Birth	record_id	✓
Gender		
Gender	age	✗
Gender - Select Other	age_units	✗
Gender - Specify	race_ethnicity_1	✓
Sexual Orientation		
Sexual Orientation	race_ethnicity_2	✓
Sexual Orientation - Select Other	race_ethnicity_3	
Sexual Orientation - Specify	race_ethnicity_4	
Marital Status	race_ethnicity_5	
Education	race_ethnicity_6	
	race_ethnicity_7	
	zip_code	✓
	sex_at_birth	
	gender	
	gender_select_other	
	gender_specify	

Assign **More Options** **Cancel**

Assigning CDEs at Upload

Recent >

My Collections >

Starred >

scharedemo / de

demo

Abstract

November

Levels of Individual

Links and

Data Items

STAT

Metadata and Statistics

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**

Postal Zip Code	zip_code	✓
Race/Ethnicity Self-Identification		
Sex at Birth	sex_at_birth	✓
Gender		
Gender	gender	✗
Gender - Select Other		
Gender - Specify		
Sexual Orientation		
Sexual Orientation		
Sexual Orientation - Select Other		
Sexual Orientation - Specify		
Marital Status		
Education		

Row 1 "None of these describe me" is not one of ["Man", "Woman", "Non-...]

Row 2 "None of these describe me" is not one of ["Man", "Woman", "Non-...]

Row 3 "Trans" is not one of ["Man", "Woman", "Non-...]

Row 4 "Trans" is not one of ["Man", "Woman", "Non-...]

Row 6 "Male" is not one of ["Man", "Woman", "Non-...]

Row 7 "Trans" is not one of ["Man", "Woman", "Non-...]

Row 8 "Female" is not one of ["Man", "Woman", "Non-...]

Row 10 "Trans" is not one of ["Man", "Woman", "Non-...]

Row 11 "Male" is not one of ["Man", "Woman", "Non-...]

Row 13 "Male" is not one of ["Man", "Woman", "Non-...]

Row 14 "Trans" is not one of ["Man", "Woman", "Non-...]

Assign

More Options

Cancel

Access Level ⓘ

Private

Analysis Readiness

Not Ready >

CDE Compliance - SCHARE ⓘ

0 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data

Assigning CDEs at Upload

Recent >

My Collections >

Starred >

scharedemo / de

demo

Abstract

November

Levels of Individual

Links a

Data It

STAT

Metadata and Statistics

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**

Postal Zip Code	zip_code	▼	✓
Race/Ethnicity Self-Identification		▼	
Sex at Birth	sex_at_birth	▼	✓
Gender			
Gender		▼	
Gender - Select Other	gender_select_other	▼	✓
Gender - Specify	gender_specify	▼	✓
Sexual Orientation			
Sexual Orientation		▼	
Sexual Orientation - Select Other		▼	
Sexual Orientation - Specify		▼	
Marital Status	marital_status	▼	✓
Education		▼	

Assign More Options Cancel

Access Level ⓘ

Private

Analysis Readiness

Not Ready >

DE Compliance - SCHARe ⓘ

0 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study State-Level Data

Viewing Uploaded Data

Click the item name to view the item's data and more information

scharedemo / demo-files / LIVE

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files



Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

Viewing Uploaded Data

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE
	 example_data.csv	5 minutes ago	25 KB

Page 1 of 1

Drag and Drop or [Browse Files](#) to Upload


> Metadata and Statistics

All data is initially private (only you can access it).


Access Level

 Private

Analysis Readiness

 Ready >

CDE Compliance - SchARe

 8 / 19 CDEs assigned

Tags

Topics to

Cross-Section

View the CDEs that you have assigned
(coming up next)



Viewing Uploaded Data

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

[🔗](#) [🕒](#) [★](#)

Operations ▾

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

> Links and Documents

> Data Items

STATUS	NAME	CREATED	SIZE
<div><div>📄</div><div>Storage: ✓ ok</div><div>Parser: ✓ ok</div><div>Validator: ✓ ok</div></div>	data.csv	5 minutes ago	25 KB

⏪ < Page 1 of 1 >

📁

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ

🔒 Private

Analysis Readiness

✓ Ready >

CDE Compliance - SchARE ⓘ

🕒 8 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data

The item status shows all good or any problems in storage, parsing, or validation

https://test-schare.nimhd.nih.gov/collections/e6d24c56-a0aa-4b89-b8f6-437245084a69/view/LIVE/example_data.csv?s=meta

Viewing Uploaded Data

Click the collection name to return to the collection main page

View the original file, the parsed data as a table, the item's data dictionary, and relevant metadata

scharedemo / [demo-files](#) / LIVE / example_data.csv

File Table Dictionary Meta 25 KB | 10 minutes ago | text/csv | status:

Item Operations ▾

record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth	gender
1943	92	Years	no	no	no	yes	no	no	no	28752	Male	None of these describe me
6688	76	Years	yes	no	no	no	no	no	no	10032		None of these describe me
9784	18	Years	no	no	yes	no	no	yes	no	07501		Trans
5193	91	Years				no	yes	no	no			Trans
8502	86	Years				no	no	no	no			Prefer not to answer
1210	72	Years				no	yes	no	no			Male
3563	79	Years				no	no	no	no			Trans
1172	46	Years	no	no	no	no	no	yes	no	37208	Prefer not to answer	Female
2618	80	Years	no	no	no	yes	no	no	no	35405	Male	Prefer not to answer
6809	58	Years	yes	no	no	no	no	no	no	11769	None of these describe me	
8405	55	Years	no	no	no	yes	no	no	no	77433	Female	

For columns with CDEs assigned, click the info popup to view the CDE details

Perform operations on the data (download, rename, move, delete, etc.)





ScHARe

ScHARe Repository Introduction

November 20, 2024

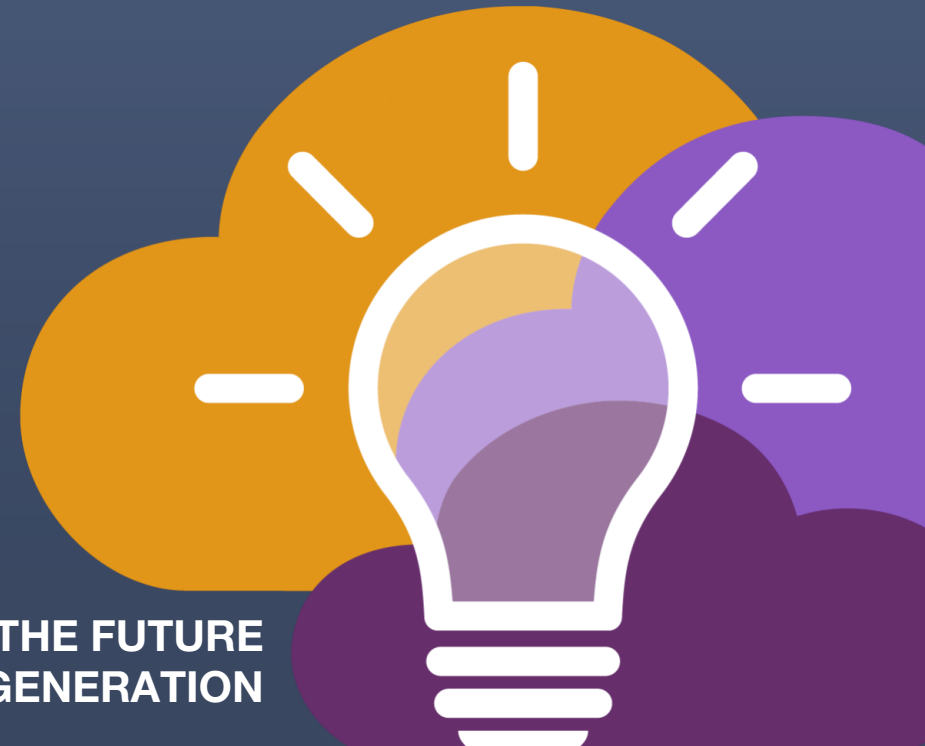
Deborah Duran, PhD • NIMHD
Elif Dede Yildirim, PhD • NIMHD
Mark Aronson, PhD • NIMHD



ScHARe

CDE Mapping

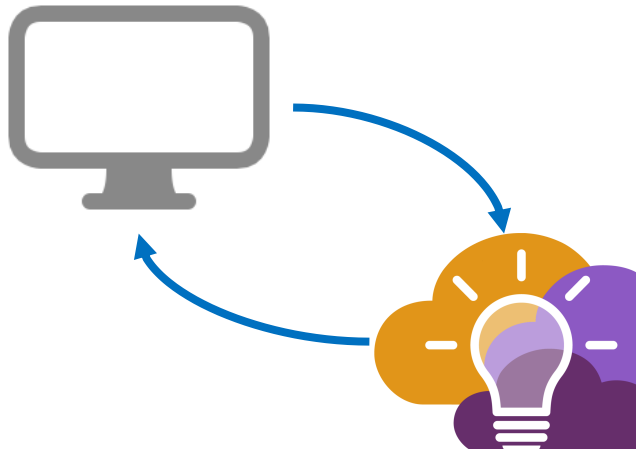
BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Using Dataviews to Map CDEs

If your data doesn't conform to the CDEs as it was uploaded, you have two options:

**Use your own tools to adjust the data
and re-upload**



**Use a Dataview to Map to CDEs
within the Repository**

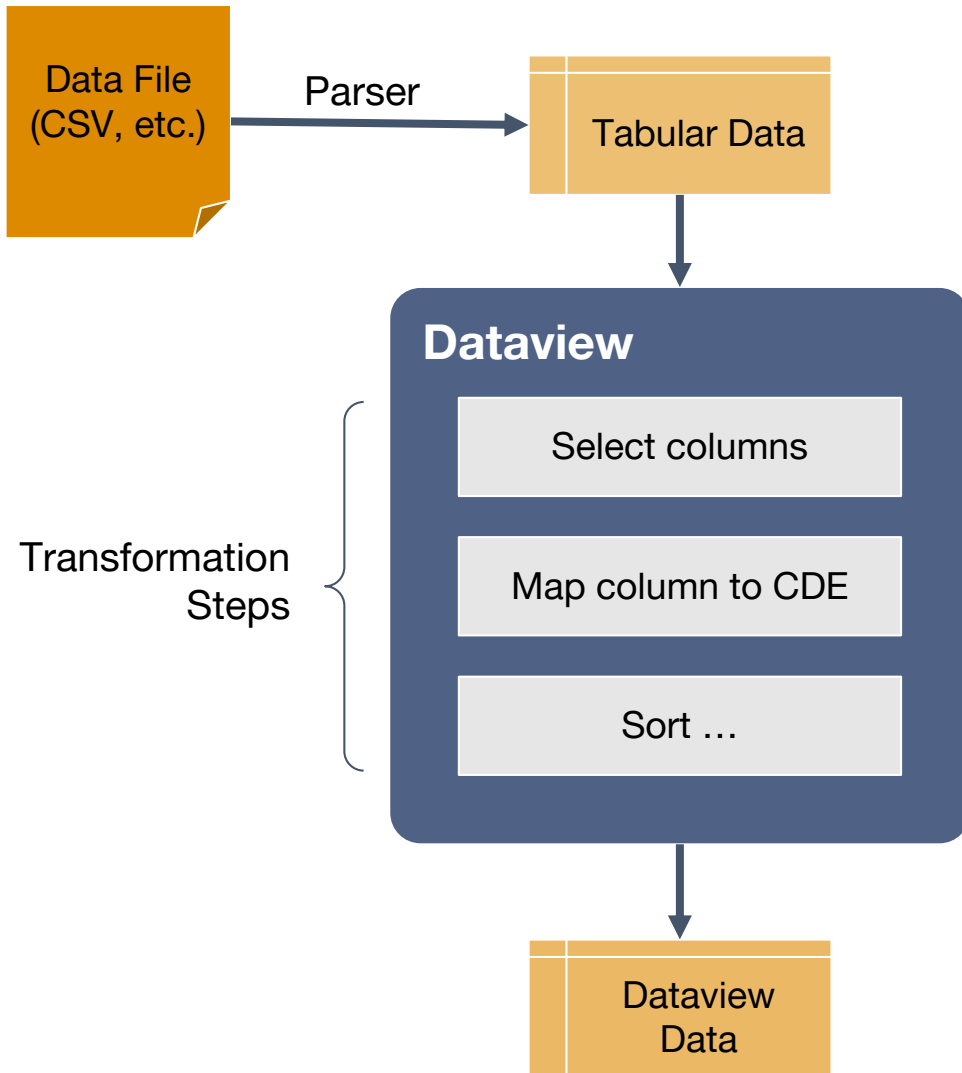
The screenshot shows the SchARE Repository interface. The main content area displays the 'Map Into' configuration for a dataview named 'mapped test data'. The source data is 'test_data.xlsx'. The configuration includes a 'Source Column' dropdown set to 'Race/Ethnicity Self-Identification', a 'Value Map from Source to Target' table, and a 'Target Column' dropdown set to 'race_mapped'. The 'Value Map' table lists source values and their corresponding target values.

SOURCE VALUE	TARGET VALUE
Black	Black or African American
African American	Black or African American
American Indian	American Indian or Alaska Native
Alaska Native	American Indian or Alaska Native

Below the table, there are input fields for 'source value' and 'target value', and a 'When source value is not found in map, target value is:' section with options for NULL, Source value, and Constant: value.



What is a Dataview?




Dataviews take data from one or more sources, apply a series of transformation steps to that data (*filtering, sorting, mapping, etc.*) resulting in a new table of data as output.

Uses of Dataviews:


- Creating subsets of data
- Hiding PHI/PII for publishing
- Summarizing individual-level data into subsets and estimates
- Joining multiple datasets together
- **Mapping to CDEs**
- ... many others!




Mapping CDEs via Dataview

ScHARe  Repository

About Docs Collections CDEs

Search... 

 scharedemo ▾


Recent >

My Collections >



Starred >

scharedemo / demo-files / LIVE / example_data.csv

File Table Dictionary Meta

25 KB | an hour ago | text/csv | status: 

Item Operations ▾

record_id	age 	age_units 	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_cod
1943	92	Years	no	no	no	yes	no	no	no	28752
6688	76	Years	yes	no	no	no	no	no	no	10032
9784	18	Years	no	no	yes	no	no	yes	no	07501
5193	91	Years	no	no	no	no	yes	no	no	26726
8502	86	Years	yes	no	no	no	no	no	no	99737
1210	72	Years	no	no	no	no	yes	no	no	52223
3563	79	Years	no	yes	no	no	no	no	no	53186
1172	46	Years	no	no	no	no	no	yes	no	37208
2618	80	Years	no	no	no	yes	no	no	no	35405

Rename...

Copy To...

Link To...

Move To Folder...

Configure Table

Assign Data Dictionary...

Create Dataview...

Import as REDCap Data Dictionary

Export Table Data...

Download

Delete

None of these



Step 1: Create a new Dataview

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / example_data.csv

File Table Dictionary Meta 25 KB | 10 minutes ago | text/csv | status:

Item Operations ▾

Item Info

Item ID

fb06dec8-0ac2-4c6d-ad62-b4adf5f3277c

Item Type

file

Name

example_data.csv

Folder ID

ROOT

Created on

2024-11-20T01:35:00

Item URL

pigeon://test-schare.nimhd.nih.gov/e6d24c56-a0aa-4b89-b8f6-437245084a69/LIVE/fb06dec8-0ac2-4c6d-ad62-b4adf5f3277c

Metadata

KEY	VALUE
content-type	text/csv
Content-type	text/csv

Storage

URL

pigeon+datastore://pigeon_s3/360a562b-e312-474e-abcd-d944da509425

Size

25994 bytes

Checksum

{ "sha-256": "dSjkNHqi7UweXRJRBOb0ALMkXrqFIfo3m8C4nU7hKIU=" }

Create Dataview

Dataview name

Create Dataview

Cancel

Step 1: Create a new Dataview

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / example_data.csv

File

Table

Dictionary

Meta

25 KB | 10 minutes ago | text/csv | status:

Item Operations ▾

Item Info

Item ID

fb06dec8-0ac2-4c6d-ad62-b4adf5f3277c

Item Type

file

Name

example_data.csv

Folder ID

ROOT

Created on

2024-11-20T01:35:00

Item URL

pigeon://test-schare.nimhd.nih.gov/e6d24c56-a0aa-4b89-b8f6-437245084a69/LIVE/fb06dec8-0ac2-4c6d-ad62-b4adf5f3277c

Metadata

KEY	VALUE
content-type	text/csv
Content-type	text/csv

Storage

URL

pigeon+datastore://pigeon_s3/360a562b-e312-474e-abcd-d944da509425

Size

25994 bytes

Checksum

{ "sha-256": "dSjkNHqi7UweXRJRBOb0ALMkXrqFIfo3m8C4nU7hKIU=" }

Create Dataview

example-dataview

Create Dataview

Cancel

Step 2: Add Mapping Steps to your Dataview

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / example-dataview

Advanced Explorer Table Dictionary Meta

559 bytes | a few seconds ago | text/prql | status: ⌛

Item Operations ▾

Source data from: example_data.csv ↴

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

+ Add Step ▾

Libraries: ⓘ

+ Add Library

+ Add Data Elements

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record_id	age ⓘ	age_units ⓘ	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code ⓘ	sex_at_birth ⓘ	gender
1943	92	Years	no	no	no	yes	no	no	no	28752	Male	None of these
6688	76	Years	yes	no	no	no	no	no	no	10032	Male	
0784	18	Years	no	no	yes	no	no	yes	no	07501	Male	

Step 2: Add Mapping Steps to your Dataview

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / example-dataview

Advanced Explorer Table Dictionary Meta

559 bytes | a few seconds ago | text/prql | status: ⌛

Item Operations ▾

Source data from: example_data.csv ↗

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

on Preview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

age_units ⓘ	race_ethnicity_1 ⓘ	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code ⓘ	sex_at_birth ⓘ	gender
Years	no	no	no	yes	no	no	no	28752	Male	None of these
Years	yes	no	no	no	no	no	no	10032	Male	
0784	18	Years	no	no	no	no	no	07501	Male	

Select Columns

Filter Rows

Sort

Shuffle

Join

Take Rows

Rename Column

Map Column

Aggregate Rows

Step 2: Add Mapping Steps to your Dataview

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / example-dataview

Advanced Explorer Table Dictionary Meta

559 bytes | a few seconds ago | text/prql | status: ⌵

Item Operations ▾

Source data from: example_data.csv

↴

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

10

↓

Map Into

Source Column

Select column... ▾

* Column selection is required

Value Map from Source to Target

source value

target value

+

* At least one map entry is required

When source value is not found in map, target value is:
☒ NULL ☐ Source value ☐ Constant: value

Target Column

new column name

New column ☐ Replace source column

* Target Column is required

↓

+ Add Step ▾


Libraries: ⓘ

+ Add Library

+ Add Data Elements

Clear Dataview

Save



Step 2: Add Mapping Steps to your Dataview

Map Into

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

source value

target value

+

Target Column

gender_mapped

New column

Replace source column

When source value is not found in map, target value is:

☒ NULL ☐ Source value ☐ Constant: value

+ Add Step

Libraries:

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth	gender
1210	72	Years	no	no	no	no	yes	no	no	52223	Female	Male
1943	92	Years	no	no	no	yes	no	no	no	28752	Male	None of
6688	76	Years	yes	no	no	no	no	no	no	10032	Male	

Step 2: Add Mapping Steps to your Dataview

Map Into

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man
Female	Woman
Trans	Transgender
Nonbinary	Non-Binary
None of these describe me	None of these describe me

source value

target value

+

Target Column

gender_mapped

New column

Replace source column

When source value is not found in map, target value is:

☒ NULL ☐ Source value ☐ Constant: value

+ Add Step

Libraries:

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record id age age units race ethnicity 1 race ethnicity 2 race ethnicity 3 race ethnicity 4 race ethnicity 5 race ethnicity 6 race ethnicity 7 zip code sex at birth



Step 3: Assign CDEs to Data Columns

Map Into

Source Column

gender

age

age_units

zip_code

sex_at_birth

gender_select_other

age

Age Units

Postal Zip Code

Sex at Birth

Gender - Select Other

N/A

N/A

N/A

N/A

N/A

✕

✕

✕

✕

✕

Add new data element annotation:

Select a column...

Add

From:

CDEs

Data Dictionary

Select a data element...

Clear

CDE set selected: SCHARE

Done

+ Add Step

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record id

age

age_units

race ethnicity 1

race ethnicity 2

race ethnicity 3

race ethnicity 4

race ethnicity 5

race ethnicity 6

race ethnicity 7

zip code

sex at birth



Step 3: Assign CDEs to Data Columns

Map Into

Source Column

gender

+ Add Step ▾

Results >

Transformation Preview

↓

Add Data Elements

You can enhance your dataview by adding new data elements or modifying existing ones that were derived from its source data.

COLUMN	DATA ELEMENT	REPRESENTS PV	
age	Age	N/A	✕
age_units	Age Units	N/A	✕
zip_code	Postal Zip Code	N/A	✕
sex_at_birth	Sex at Birth	N/A	✕
gender_select_other	Gender - Select Other	N/A	✕

Add new data element annotation:

From:

CDEs

Data Dictionary

gender_mapped

Gender

CDE set selected: SchARE

OPTIONAL: Select a PV...

Add

Clear

Done

Clear Dataview

Save Dataview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

record id

age

age_units

race_ethnicity_1

race_ethnicity_2

race_ethnicity_3

race_ethnicity_4

race_ethnicity_5

race_ethnicity_6

race_ethnicity_7

zip_code

sex at birth



Step 3: Assign CDEs to Data Columns

Results ▾

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

No errors to report! 📄

Transformation Preview

y ⓘ	measure_0	measure_1	measure_2	measure_3	measure_4	measure_5	measure_6	measure_7	measure_8	measure_9	gender_mapped ⓘ
62	0.5455549357622266	1	0.5884900154230546	0.12911211858353722	57	33	0.7170385313006806	0.8788787389282339	0.5467863499436898	None of these describe me. I would like to consider additional options	
90	0.9939486148968586	18	0.8889859573142674	0.9745095524972497	2	68	0.7794942573077963	0.34939465386130986	0.2571283059006645	None of these describe me. I would like to consider additional options	
92	0.1187476169408952	1	0.6512384712588953	0.8971788616590552	57	15	0.5303711405755923	0.45273835850696076	0.4082502293654543	Transgender	
60	0.4452622773910989	10	0.5971200227987504	0.7493012469195882	98	92	0.8712062558952113	0.4832045355262965	0.9295209737850779	Transgender	
53	0.971610341237526	55	0.8817829051636139	0.05976090957174052	77	88	0.5610231545798023	0.6188743418431822	0.8298724376862494	Man	
15	0.6678690757541459	12	0.35150592643125234	0.016162277012149118	98	24	0.4741302934986361	0.3494528319110445	0.5733822336463569	Transgender	
93	0.20797803936217862	88	0.5164279582660238	0.17957863745725022	18	6	0.3757660830284647	0.6016044994045969	0.35718866008810335	Woman	
49	0.33569642444900216	25	0.8284451772569406	0.1251105503540758	92	56	0.7920621573885018	0.46157047423521724	0.5849628410398715	Transgender	
55	0.4669418786022268	59	0.9681471694553496	0.1829245483682186	38	69	0.16838001266365477	0.654926782920557	0.4987917893025189		
34	0.31022787630950976	53	0.2198646710684642	0.024882302497037267	87	99	0.38764603267169684	0.7459103083590588	0.9586660722941287		

rows per page: 25 ▾ 1-10 of 1



Step 3: Assign CDEs to Data Columns

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

Target Column

gender_mapped

New column

Replace source column

Add Data Elements

You can enhance your dataview by adding new data elements or modifying existing ones that were derived from its source data.

COLUMN	DATA ELEMENT	REPRESENTS PV
gender_mapped	Gender	N/A
age	Age	N/A
age_units	Age Units	N/A
zip_code	Postal Zip Code	N/A
sex_at_birth	Sex at Birth	N/A

Add new data element annotation:

From:

CDEs

Data Dictionary

Select a column...

Select a data element...

CDE set selected: SCHARE

Add

Clear

Done

+ Add Step

Results

No errors to report

Transformation Preview

agency	measure_0	measure_1	measure_2	measure_3	measure_4	measure_5	measure_6	measure_7	measure_8	measure_9	gender
62	0.5455549357622266	1	0.5884900154230546	0.12911211858353722	57	33	0.7170385313006806	0.8788787389282339	0.5467863499436898		

Clear Dataview

Save Dataview

0 validation errors



Step 3: Assign CDEs to Data Columns

Source Column

gender

value map from Source to Target

SOURCE VALUE	TARGET VALUE
--------------	--------------

Target Column

gender_mapped

Add Data Elements

You can enhance your dataview by adding new data elements or modifying existing ones that were derived from its source data.

COLUMN	DATA ELEMENT	REPRESENTS PV	
gender_mapped	Gender	N/A	
age	Age	N/A	✕
age_units	Age Units	N/A	✕
zip_code	Postal Zip Code	N/A	✕
sex_at_birth	Sex at Birth	N/A	✕

Add new data element annotation:

From: CDEs Data Dictionary

race_ethnicity_1 Race/Ethnicity Self-Identification CDE set selected: SchARE American Indian or Alaska Native

Add

Clear

Done

+ Add Step

Results

No errors to report

Transformation Preview

agency	measure_0	measure_1	measure_2	measure_3	measure_4	measure_5	measure_6	measure_7	measure_8	measure_9	gender
62	0.5455549357622266	1	0.5884900154230546	0.12911211858353722	57	33	0.7170385313006806	0.8788787389282339	0.5467863499436898		

Clear Dataview Save Dataview

0 validation errors

Step 3: Assign CDEs to Data Columns

Source Column

gender

value map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

Target Column

gender_mapped

New column

Replace source column

Add Data Elements

You can enhance your dataview by adding new data elements or modifying existing ones that were derived from its source data.

COLUMN	DATA ELEMENT	REPRESENTS PV
gender_mapped	Gender	N/A
race_ethnicity_1	Race/Ethnicity Self-Identification	American Indian or Alaska Native
age	Age	N/A
age_units	Age Units	N/A
zip_code	Postal Zip Code	N/A

Add new data element annotation:

From:

CDEs

Data Dictionary

Select a column...

Select a data element...

CDE set selected: SCHARE

Add

Clear

Done

+ Add Step

Results

No errors to report

Transformation Preview

efficiency	measure_0	measure_1	measure_2	measure_3	measure_4	measure_5	measure_6	measure_7	measure_8	measure_9
62	0.5455549357622266	1	0.5884900154230546	0.12911211858353722	57	33	0.7170385313006806	0.8788787389282339	0.546786349943689	

Clear Dataview

Save Dataview

0 validation errors



Step 3: Assign CDEs to Data Columns

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results ▾

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

No errors to report! 🙌

Transformation Preview

record_id	age ⓘ	age_units ⓘ	race_ethnicity_1 ⓘ	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code ⓘ	sex_at_birth ⓘ	gender
1943	92	Years	no	<div><div>Race/Ethnicity Self-Identification : American Indian or Alaska Native (string)</div><div>Concepts ▾</div><div>C41259: American Indian or Alaska Native C17049: Race C16564: Ethnic Group C74528: Self-Report</div><div>Definitions ></div></div> <div>valueDomain dataElement dataElement dataElement</div>				no	no	28752	Male	None of these describe me
6688	76	Years	yes					no	no	10032	Male	None of these describe me
9784	18	Years	no					yes	no	07501	Male	Trans
5193	91	Years	no					no	no	26726	None of these describe me	Trans
1210	72	Years	no					no	no	52223	Female	Male
3563	79	Years	no					no	no	53186	None of these describe me	Trans
1172	46	Years	no	no	no	no	no	yes	no	37208	Prefer not to answer	Female
6809	58	Years	yes	no	no	no	no	no	no	11769	None of these describe me	Female
8502	86	Years	yes	no	no	no	no	no	no	99737	Prefer not to answer	Female



Step 3: Assign CDEs to Data Columns

+ Add Step ▾

Results ▾

No errors to report

Transformation

record_id	age										
1943	92										
6688	76	Years	yes	no	no	no	no	no	no	10032	
9784	18	Years	no	no	yes	no	no	yes	no	07501	
5193	91	Years	no	no	no	no	yes	no	no	26726	

Trans

Transgender

Clear Dataview

Save Dataview

✓ 0 validation errors

6	race_ethnicity_7	zip_code	sex_at_birth
no		28752	Male

Add Data Elements

You can enhance your dataview by adding new data elements or modifying existing ones that were derived from its source data.

race_ethnicity_2	Race/Ethnicity Self-Identification	Asian or Asian American
race_ethnicity_3	Race/Ethnicity Self-Identification	Black or African American
race_ethnicity_4	Race/Ethnicity Self-Identification	Hispanic, Latino, or Spanish
race_ethnicity_5	Race/Ethnicity Self-Identification	Native Hawaiian or Other Pacific Islander
race_ethnicity_6	Race/Ethnicity Self-Identification	Middle Eastern or North African
race_ethnicity_7	Race/Ethnicity Self-Identification	White

Add new data element annotation:

From: CDEs Data Dictionary

Select a column...

Select a data element...

CDE set selected: SCHARE

Add

Clear

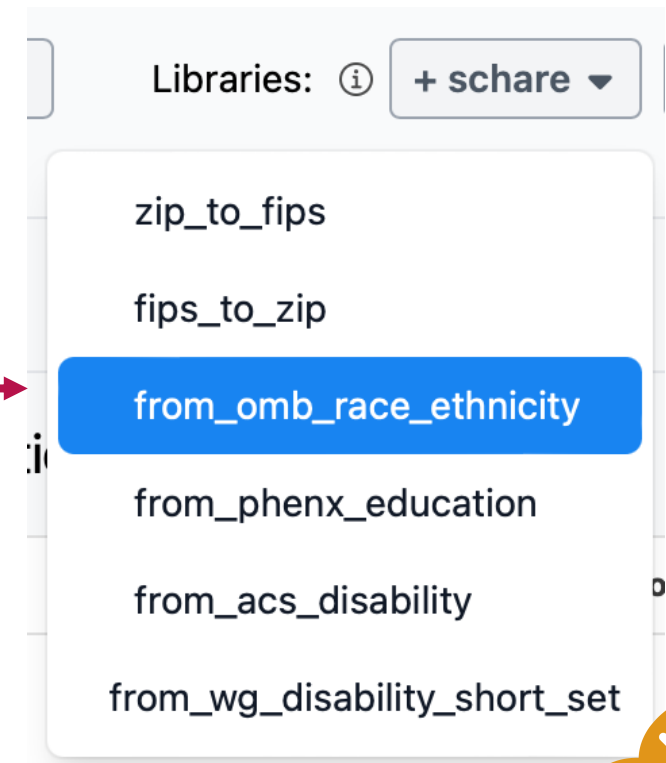
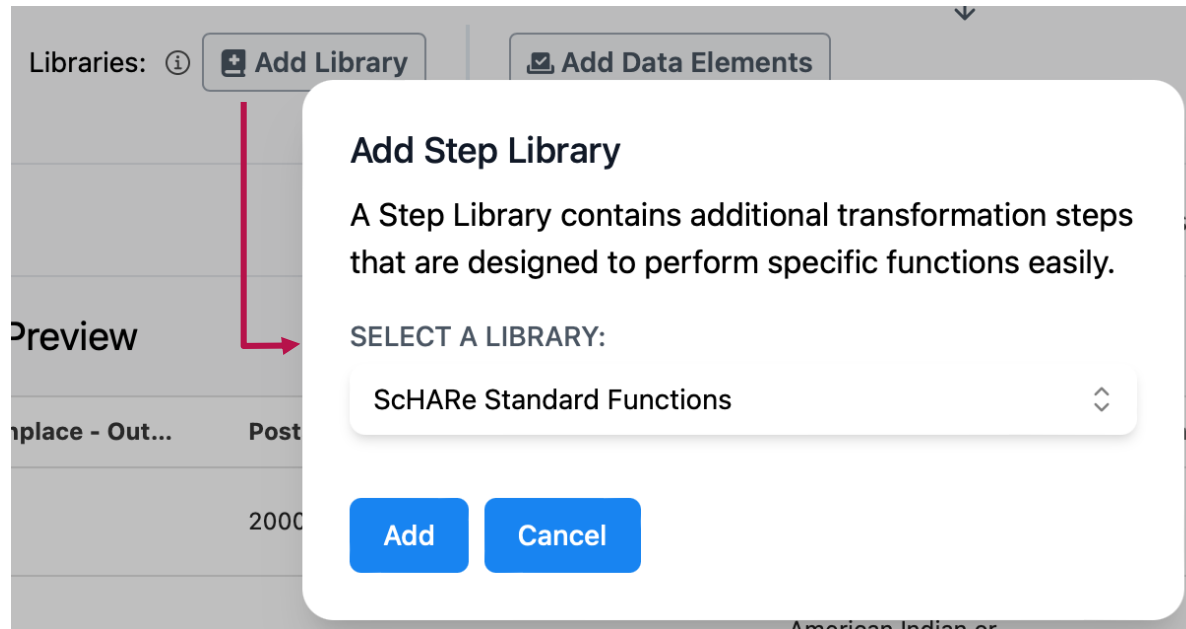
Done



Mapping CDEs via Dataview

The system also has a number of specific mapping functions available, for cases when the mapping is standardized or when the mapping is more complex than can be handled by the generic mapping functions.

Please contact our support team if you need any help with custom mapping functions!



Slido Poll

What conditions would make you more likely to share your data publicly?

- a) Proper anonymization and privacy safeguards
- b) Clear contribution to scientific advancement
- c) Personal or community benefit
- d) Legal and ethical assurances
- e) I wouldn't share it under any circumstances

ScHARe

Sharing Data

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Viewing your Collection

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

🔗 ⌂ ☆ Operations ▾

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

See whether your data is well-annotated or what you should improve

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
🟢	example_data.csv	34 minutes ago	25 KB	⋮
🟢	example-dataview	20 minutes ago	2 KB	⋮

<< Page 1 of 1 >>

📁

Drag and Drop or [Browse Files](#) to Upload

Access Level ⓘ

🔒 Private

Analysis Readiness

🟢 Ready >

CDE Compliance - SchARE ⓘ

🕒 9 / 19 CDEs assigned

Tags

Topics tagged in

Cross-Sectional Study

View the CDEs that you have assigned

CDE Compliance and Analysis Readiness

Analysis Readiness - a simple metric on whether your data is ready for downstream use

1. Did you assign **metadata** to your collection (tags, project-level CDEs)?
2. Have you partially or fully **assigned CDEs** to your data?
3. Is the data **accessible** to the system (no broken links)?
4. Does the data pass **validation** according to the assigned CDEs?

Analysis Readiness

✓ Ready ▾

Metadata: Complete

DEs Assigned: Partially Complete

Data Access: Passing

Validations: Passing

CDE Compliance - ScHARe ⓘ

🕒 9 / 19 CDEs assigned

Age Annual Household Inco...

English Proficiency Gender

Household Member Size Marital Status

Postal Zip Code Race/Ethnicity Self-Ide...

Sex at Birth

CDE Compliance - Showing how many CDEs have been assigned across the data in the collection

- Click on the metric to pop up the list of CDEs assigned
- Click on an individual CDE to find more information about that CDE



Access Levels and Sharing Data

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

Operations ▾

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

> Links and Documents

> Data Items

STATUS	NAME	CREATED	SIZE	
	example_data.csv	34 minutes ago	25 KB	⋮
	example-dataview	20 minutes ago	2 KB	⋮

<< Page 1 of 1 >>

Drag and Drop or [Browse Files](#) to Upload

Access Level ⓘ

Private

Analysis Readiness

✓ Ready ▾

Metadata: Complete

DEs Assigned: Partial

Data Access: Passing

Validations: Passing

CDE Compliance - SchARe ⓘ

9 / 19 CDEs assigned

Age

Annual Household Income

English Proficiency

Gender

Household Member Size

Marital Status

Postal Zip Code

Race/Ethnicity Self-Identified

Sex at Birth

Tags

Topics tagged in this collection

Configure CDEs

Add Readme

Add Dictionary...

Add Folder...

Add Link...

Make Private

Share...

Make Public

Edit

Delete Collection

Access Levels and Sharing Data

scharedemo / demo-files / LIVE

Recent > My Collections > Starred >

demo-files

Abstract
November
Levels of Individual

Links and Data Items

STAT

DE Compliance - SchARE

9 / 19 CDEs assigned

Age Annual Household Income English Proficiency Gender Household Member Size Marital Status

Tags
Topics tagged in this collection

Access Level ①
Private

Analysis Readiness
Ready
Metadata: Complete
DEs Assigned: Partially Complete
Data Access: Passing
Validations: Passing

Share Collection

Users, groups, and collections with access:

ID	ROLE
ScHARe Demo (scharedemo)	ADMIN

Share with: Users Groups Collections

This collection's access level is currently set to **Private**.
To share this collection with others, you must first set the access level to **Confidential**.

Make Confidential


Done

<< Page 1 of 1 >

You have control over how your data is shared on the SchARE Repository. By default, all collections start out as **Private**.



Access Levels and Sharing Data

ScHARe 

Repository

About

Docs

Collections

CDEs

Search...

scharedemo

scharedemo

Recent

My Collections

Starred

scharedemo / demo-files / LIVE

Collection authorization successfully updated.

Operations

demo-files

Abstract

November

Levels of Individual

Links and

Data Items

STAT

Example data view

Access Level

Confidential

Analysis Readiness

Ready

Metadata: Complete

DEs Assigned: Partially Complete

Data Access: Passing

Validations: Passing

DE Compliance - ScHARe

9 / 19 CDEs assigned

Age

Annual Household Income

English Proficiency

Gender

Household Member Size

Marital Status

Postal Zip Code

Race/Ethnicity Self-Id...



Sex at Birth

Page 1 of 1

Drag and Drop or [Browse Files](#) to Upload

Share Collection

Users, groups, and collections with access:

ID	ROLE	
 SchARe Demo (scharedemo)	ADMIN	

Share with:

Users

Groups

Collections

Search for a user...

Role: Reader

Share

Done



Access Levels and Sharing Data

The screenshot shows the SchARe Demo interface with the 'demo-files' collection selected. The 'Access Level' is set to 'Confidential'. A 'Share Collection' dialog box is open, displaying a table of users with access:

ID	ROLE
SCHARe Demo (scharedemo)	ADMIN

Below the table, the 'Share with' section has three tabs: 'Users', 'Groups', and 'Collections'. The 'Users' tab is selected, and a dropdown menu is open, showing a list of users:

- Elif
- Elif Dede Yildirim (dedeyildirim2)
- Elif Dede Yildirim (elif.dedeyildirim)
- Elif Dede Yildirim (elifdede)

The 'Role' is set to 'Reader', and a 'Share' button is visible.

After changing your collection's access level to Confidential, you can share it with other users, groups, and collections.



Access Levels and Sharing Data

The screenshot shows a 'Share Collection' dialog box overlaid on a web interface. The dialog has a title 'Share Collection' and a subtitle 'Users, groups, and collections with access:'. Below this is a table with two columns: 'ID' and 'ROLE'. The table contains one entry: 'SCHARe Demo (scharedemo)' with the role 'ADMIN'. To the right of the table is a trash icon. Below the table are three tabs: 'Users', 'Groups', and 'Collections'. The 'Users' tab is selected. Below the tabs, there is a search bar containing 'Elif Dede Yildirim' and a dropdown arrow. To the right of the search bar is a 'Role:' label followed by a dropdown menu. The dropdown menu is open, showing four options: 'Viewer', 'Reader' (which is highlighted), 'Writer', and 'Admin (full control)'. To the right of the role dropdown is a blue 'Share' button. At the bottom left of the dialog is a blue 'Done' button. The background of the dialog shows a blurred view of the web interface, including a 'demo-files' section, an 'Access Level' dropdown set to 'Confidential', and a 'Analysis Readiness' section with a 'Ready' status and various completion metrics.

Access Levels

The access level of a collection defines the maximum permissions that can be used to share it with others. The following access levels are supported:

- **Private:** Only the collection's owner can access
- **Confidential:** The collection can be shared with named users
- **Controlled:** The collection can be shared with members of a controlled access group, as well as named users
- **Public:** The collection can be read by any user, including those not logged in; it can also be shared with named users



Access Levels and Sharing Data

ScHARe Repository

About Docs Collections CDEs Search...

scharedemo / demo-files / LIVE

Collection authorization successfully updated.

Recent My Collections Starred

demo

Abstract
November
Levels of Individual

Links and Data Items

STAT

Drag and Drop or [Browse Files](#) to Upload

Share Collection

Users, groups, and collections with access:

ID	ROLE	
ScHARe Demo (scharedemo)	ADMIN	
Elif Dede Yildirim (elifdede)	READER	

Share with: **Users** Groups Collections

Search for a user... Role: Reader Share


Done

Access Level ①
Confidential

Analysis Readiness
Ready
Metadata: Complete
DEs Assigned: Partially Complete
Data Access: Passing
Validations: Passing

DE Compliance - ScHARe ①
9 / 19 CDEs assigned
Age Annual Household Income
English Proficiency Gender
Household Member Size Marital Status
Postal Zip Code Race/Ethnicity Self-Id...
Sex at Birth

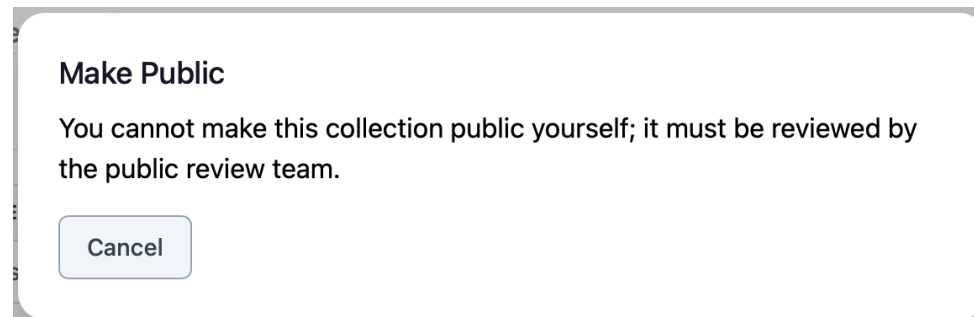
<< < Page 1 of 1 >



Publishing Data

We are still working on the final version of the process; however, it will essentially be:

1. Prepare your data according to the published guidelines
2. Add the Public Review group as Admin on your collection
 - a. This is necessary for them to be able to alter your collection's Access Level on your behalf
3. Notify the Public Review group by email
 - a. They will review your data for compliance with the data sharing guidelines
 - b. If any changes are necessary, you can work with the reviewer
 - c. Once complete, they will create a static version of your collection and set its Access Level to Public





ScHARe

ScHARe Repository Introduction

November 20, 2024

Deborah Duran, PhD • NIMHD
Elif Dede Yildirim, PhD • NIMHD
Mark Aronson, PhD • NIMHD



ScHARe

Data Aggregation

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Downloading data to your own computer

Use the item's menu to download the original file

- Best for working directly with raw data

You can also export and download the tabular data in CSV, TSV, or Parquet format

- Best for working with mapped data, summary data, and other processed data

The screenshot displays the Schare demo interface. The top navigation bar shows the path 'scharedemo / demo-files / LIVE'. The main content area is titled 'demo-files' and contains an abstract, levels of influence, and domains of influence. Below this, there is a section for 'Data Items' with a table listing files. A context menu is open over the 'example-data.csv' file, showing options like 'Export Table Data...', 'Download', and 'Delete'. The right-hand panel shows 'Access Level' as 'Confidential', 'Analysis Readiness' as 'Ready', and 'CDE Compliance - SCHARE' with 9 / 19 CDEs assigned. There are also tags for 'Cross-Sectional Study' and 'State-Level Data'.

STATUS	NAME	CREATED	SIZE
	example-data.csv	an hour ago	25 KB
	example-dataview		

Drag and Drop or [Browse Files](#) to Upload

Metadata and Statistics

Access Level ①
Confidential

Analysis Readiness
Ready >

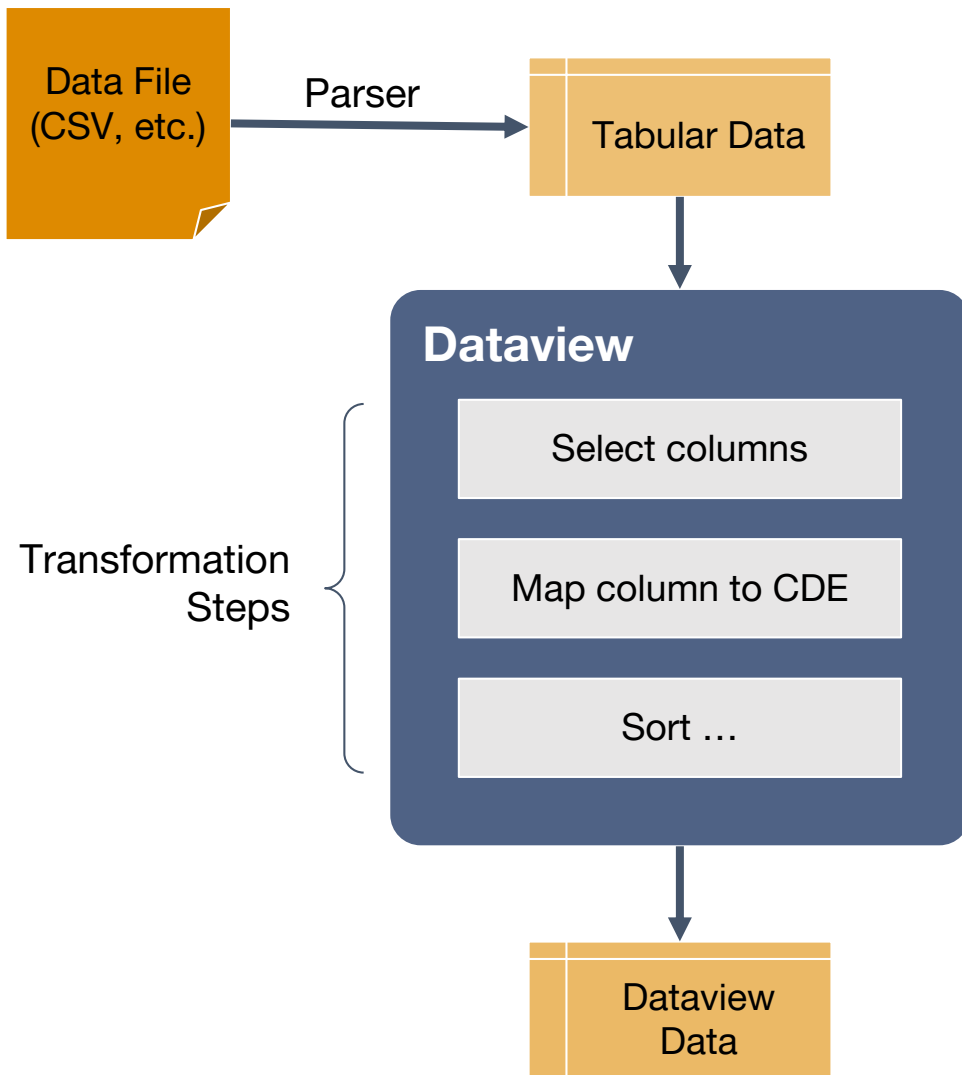
CDE Compliance - SCHARE ①
9 / 19 CDEs assigned

Tags
Topics tagged in this collection
Cross-Sectional Study State-Level Data



Using Dataviews to Aggregate Data

Dataviews take data from one or more sources, apply a series of transformation steps to that data (*filtering, sorting, mapping, etc.*) resulting in a new table of data as output.



Uses of Dataviews:

- Creating subsets of data
- Hiding PHI/PII for publishing
- Summarizing individual-level data into subsets and estimates
- **Joining multiple datasets together**
- Mapping to CDEs
- ... many others!



Browsing and Searching for Data

Collections

Create New Collection

My Collections

- demo-files** LIVE
November 20 Think-a-Thon Demonstration Files
18 hours ago

Shared with Me

You don't have any collections shared with you right now!

Public Collections

grouped by: Domains of Influence

Biological

- Example NMHSS Analysis** LIVE
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.
7 months ago

Health Care Systems and Clinical Care

- Example NMHSS Analysis** LIVE
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.
7 months ago
- Minority Health SVI** LIVE
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health...
8 months ago

Sociocultural Environment

- Minority Health SVI** LIVE
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health...
8 months ago

Repository

About Docs Collections CDEs

minority

Search

Filters

Apply Filters

Minority Health SVI
karl • 8 months ago
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Example NMHSS Analysis
karl • 7 months ago
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.

Scholar Repository - f1.2.1.d20241119 - b1.2.2.dev0-r94k313bc.d20241119

Look for improvements to the data organization and search experience in 2025!



Linking and Aggregating Data

If you find public data that you would like to combine with your project data, you can link it into your own collection and use it just like a file that you had uploaded yourself.

▼ Data Items

STATUS	NAME	CREATED	SIZE
	MH SVI Overview_11.19.2021.pdf	7 months ago	193 KB
	MH SVI Fact Sheet_7.15.2021.pdf	7 months ago	276 KB
	mh_svi_county_2018.csv	7 months ago	7.5 MB
	MinorityHealthSVI_DataDictionary_2018.csv	7 months ago	25 KB
	mh_svi_county-ScHARe	7 months ago	4 KB
	mh-svi-by-state-2018		
	MinorityHealthSVI_DataDictionary_2018.pdd.js...		

Copy To...
Link To...
Export Table Data...
Download

> Metadata and Statistics

Link Item To...

My Collections

- ✓ ScHARe Example Data 2

Public Collections

- test unlisted data
- Example NMHSS Analysis
- Minority Health SVI

Link Cancel

ScHARe Example Data 2

Abstract
For purposes of demonstration, this project collected data according to ...

Levels of Influence
Individual
Community

Domains of Influence
Biological
Behavioral

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE
	test_data.xlsx	3 hours ago	14 KB
	mapped test data	an hour ago	1 KB
	mh_svi_county-ScHARe	a few seconds ago	4 KB

<< Page 1 of 1 >>



Linking and Aggregating Data

If you find public data that you would like to combine with your project data, you can link it into your own collection and use it just like a file that you had uploaded yourself.

Minority Health SVI

Abstract

The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence
Community
Societal

Domains of Influence
Health Care Systems and Clinical Care
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	MH SVI Overview_11.19.2021.pdf	8 months ago	193 KB	⋮
	MH SVI Fact Sheet_7.15.2021.pdf	8 months ago	276 KB	⋮
	mh_svi_county_2018.csv	8 months ago	7.5 MB	⋮
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB	⋮
	mh_svi_county-ScHARe	8 months ago	4 KB	⋮
	mh-svi-by-state-2018	7 months ago	2 KB	⋮
	MinorityHealthSVI_DataDictionary_2018.pdd.json	5 months ago	87 KB	⋮

<< Page 1 of 1 >>

> Metadata and Statistics

Access Level ⓘ

Public

Analysis Readiness

✓ Ready >

CDE Compliance - ScHARe

1 / 19 CDEs assigned

Tags

Topics tagged in this collection

Minority Health Household Income
Household size Race
Health Care Delivery



Linking Data

Recent

My Collections >

Starred >

Minority Health SVI

Abstract

The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence

Community
Societal

Domains of Influence

Health Care Systems and Clinical Care
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	MH SVI Overview_11.19.2021.pdf	8 months ago	193 KB	⋮
	MH SVI Fact Sheet_7.15.2021.pdf	8 months ago	276 KB	⋮
	mh_svi_county_2018.csv	8 months ago	7.5 MB	⋮
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB	⋮
	mh_svi_county-ScHARe	8 months ago	4 KB	⋮
	mh-svi-by-state-2018			
	MinorityHealthSVI_DataDictionary_2018.pdd.json			

Copy To...

Link To...

Export Table Data...

Download

> Metadata and Statistics

Access Level ⓘ

Public

Analysis Readiness

Ready >

CDE Compliance - ScHARe

1 / 19 CDEs assigned

Tags

Topics tagged in this collection

Minority Health

Household Income

Household size

Race

Health Care Delivery



Linking Data

Recent >

My Collections >

Starred >

Minority Health SVI

Abstract

The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence

Community

Societal

Domains of Influence

Health Care

Sociocultural

Links and Documents

Data Items

STATUS	NAME	CREATED	SIZE
	MH SVI Overview_11.19.2021.pdf	months ago	193 KB
	MH SVI Fact Sheet_7.15.2021.pdf	months ago	276 KB
	mh_svi_county_2018.csv	8 months ago	7.5 MB
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB
	mh_svi_county-SchARE	8 months ago	4 KB
	mh-svi-by-state-2018	7 months ago	2 KB
	MinorityHealthSVI_DataDictionary_2018.pdd.json	5 months ago	87 KB

<< Page 1 of 1 >>

Metadata and Statistics

Link Item To...

My Collections

demo-files

Public Collections

ScHARe Think-a-Thon - Example Data

test unlisted data

Example NMHSS Analysis

Minority Health SVI

Link

Cancel

Access Level ⓘ

Public

Analysis Readiness

Ready >

CDE Compliance - SchARE

1 / 19 CDEs assigned

Tags

Topics tagged in this collection

Minority Health

Household Income

Household size

Race

Health Care Delivery

Linking Data

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE

Operations ▾

demo-files

Abstract
November 20 Think-a-Thon Demonstration Files

Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

> Links and Documents

< Data Items

STATUS	NAME	CREATED	SIZE	
	example-data.csv	8 hours ago	25 KB	
	example-dataview	7 hours ago	2 KB	
	mh_svi_county-SchARE	a few seconds ago	4 KB	

<< < Page 1 of 1 >

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ
 Confidential

Analysis Readiness
 Ready >

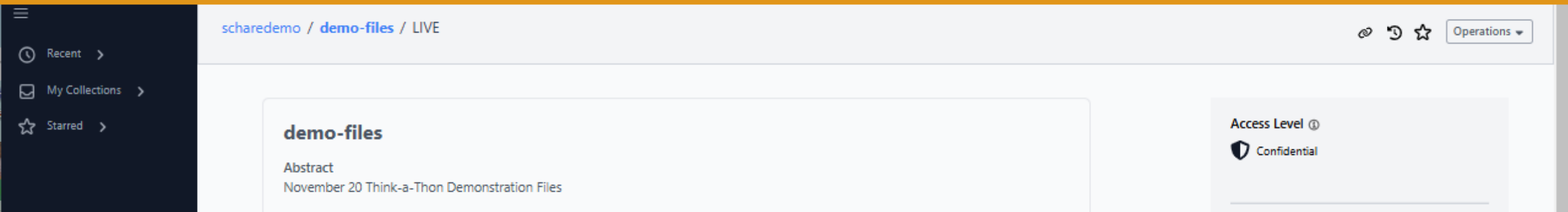
CDE Compliance - SchARE ⓘ
 9 / 19 CDEs assigned

Tags
Topics tagged in this collection

Cross-Sectional Study

State-Level Data

Aggregating Data



Once linked into your collection, you can create a dataview that combines the public data with your own project data.

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	example-data.csv	8 hours ago	25 KB	⋮
	example-dataview	7 hours ago	2 KB	⋮
	mh_svi_county-SchARe			
Drag and Drop or Browse Files to Upload				

- Rename...
- Copy To...
- Link To...
- Move To Folder...
- Configure Table
- Assign Data Dictionary...
- Create Dataview...
- Import as REDCap Data Dictionary

> Metadata and Statistics

CDE Compliance - SchARe ⓘ

🕒 9 / 19 CDEs assigned

Tags

Topics tagged in this collection

[Cross-Sectional Study](#) [State-Level Data](#)



Aggregating Data

My Collections >

Starred >

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence

Individual

Domains of Influence

Sociocultural Environment

> Links and Documents

> Data Items

STATUS	NAME	CREATED	SIZE
	example-data.csv	8 hours ago	25 KB
	example-dataview	7 hours ago	2 KB
	mh_svi_county-SchARE	minutes ago	4 KB

<< Page 1 of 1 >>

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ

Confidential

Analysis Readiness

Ready >

CDE Compliance - SchARE ⓘ

9 / 19 CDEs assigned

Tags

Topics tagged in this collection

Cross-Sectional Study

State-Level Data

Aggregating Data

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status: ⌵

Item Operations ▾

Source data from: example-dataview

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

on Preview

	age units	race ethnicity 1	race ethnicity 2	race ethnicity 3	race ethnicity 4	race ethnicity 5	race ethnicity 6	race ethnicity 7	zip code	sex at birth	gender	gender select
	Years	no	no	no	yes	no	no	no	28752	Male	None of these describe me	Trans man / Trans woman / FTM
	Years	yes	no	no	no	no	no	no	10032	Male	None of these describe me	Trans man / Trans woman / FTM
9784	18	Years	no	no	yes	no	no	yes	07501	Male	Trans	
5193	91	Years	no	no	no	no	yes	no	26726	None of these describe me	Trans	
8502	86	Years	yes	no	no	no	no	no	99737	Prefer not to answer	Prefer not to answer	

Select Columns

Filter Rows

Sort

Shuffle

Join

Take Rows

Rename Column

Map Column

Aggregate Rows

💡

Aggregating Data

☆ Starred >

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status:

Item Operations ▾

Source data from: [example-dataview](#) ↗

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

↓

Join Select

Join Table

Select table... ▾

* Join table selection is required

Dataview Column

Select column... ▾

* Column selection is required

Matching Column from Join Table

Select column... ▾

* Column selection is required

↓

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Transformation Incomplete

There are one or more errors in the transformation above. Please correct the errors to preview the transformation result.

Developer Data >

Aggregating Data

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta

575 bytes | a few seconds ago | text/prql | status:

Item Operations ▾

Source data from: [example-dataview](#) ↗

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

↓

Join Select

Join Table

Dataview Column

Matching Column from Join Table

↓

+ Add Step ▾

Libraries: ⓘ

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record_id	age ⓘ	age_units ⓘ	race_ethnicity_1 ⓘ	race_ethnicity_2 ⓘ	race_ethnicity_3 ⓘ	race_ethnicity_4 ⓘ	race_ethnicity_5 ⓘ	race_ethnicity_6 ⓘ	race_ethnicity_7 ⓘ	zip_code ⓘ	sex_at_birth ⓘ	gender ⓘ
1943	92	Years	no	no	no	yes	no	no	no	28752	Male	No th des me



Aggregating Data

☰

Recent >

My Collections >

Starred >

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status:

Item Operations ▾

Source data from: example-dataview

↷

Take

Number of rows ☐ Range of rows ☐ Limit output to Number of Rows

↓

Join Select

Join Table

mh_svi_county-SchARE

Dataview Column

zip_code

Matching Column from Join Table

zip_code

↓

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

on Preview

age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth	gender
Years	no	no	no	yes	no	no	no	28752	Male	No th des me

Select Columns

Filter Rows

Sort

Shuffle

+ Join

Take Rows

Rename Column

Selecting Variables

mh_svi_county-SchARE

zip_code

zip_code

Select

Available Columns

= ST_ABBR ⓘ

= FIPS ⓘ

= M_TOTPOP

= E_HU

= M_HU

= E_HH

Add All >>

Selected Columns

= zip_code ⓘ

= gender_mapped ⓘ

= sex_at_birth ⓘ

= STATE ⓘ

= COUNTY ⓘ

= LOCATION

= E_TOTPOP

<< Remove All

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record id	age ⓘ	zip_code ⓘ	gender_mapped ⓘ	sex_at_birth ⓘ	STATE ⓘ	COUNTY ⓘ	LOCATION	E_TOTPOP
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Avery	Avery County, North Carolina	
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Rutherford	Rutherford County, North Carolina	
1210	72	52223	Man	Female	IOWA	Delaware	Delaware County, Iowa	



Selecting Variables

Add All >>

<< Remove All

+ Add Step ▾

Libraries: ⓘ

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results >

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record_id	age ⓘ	zip_code ⓘ	gender_mapped ⓘ	sex_at_birth ⓘ	STATE ⓘ	COUNTY ⓘ	LOCATION	E_TOTPOP
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Avery	Avery County, North Carolina	17501
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Rutherford	Rutherford County, North Carolina	66532
1210	72	52223	Man	Female	IOWA	Delaware	Delaware County, Iowa	17258
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	McDowell	McDowell County, North Carolina	45109
8502	86	99737	Prefer not to answer	Prefer not to answer	ALASKA	Southeast Fairbanks	Southeast Fairbanks Census Area, Alaska	6876
9784	18	07501	Transgender	Male	NEW JERSEY	Passaic	Passaic County, New Jersey	504041
2618	80	35405	Prefer not to answer	Male	ALABAMA	Tuscaloosa	Tuscaloosa County, Alabama	206213
6809	58	11769	Transgender	None of these describe me	NEW YORK	Suffolk	Suffolk County, New York	1487901
6688	76	10032	None of these describe me. I would like to consider additional options	Male	NEW YORK	New York	New York County, New York	1632480
5193	91	26726	Transgender	None of these describe me	WEST VIRGINIA	Grant	Grant County, West Virginia	11641
3563	79	53186	Transgender	None of these describe me	WISCONSIN	Waukesha	Waukesha County, Wisconsin	398879
5193	91	26726	Transgender	None of these describe me	WEST VIRGINIA	Mineral	Mineral County, West Virginia	27278
1172	46	37208	Woman	Prefer not to answer	TENNESSEE	Davidson	Davidson County, Tennessee	6

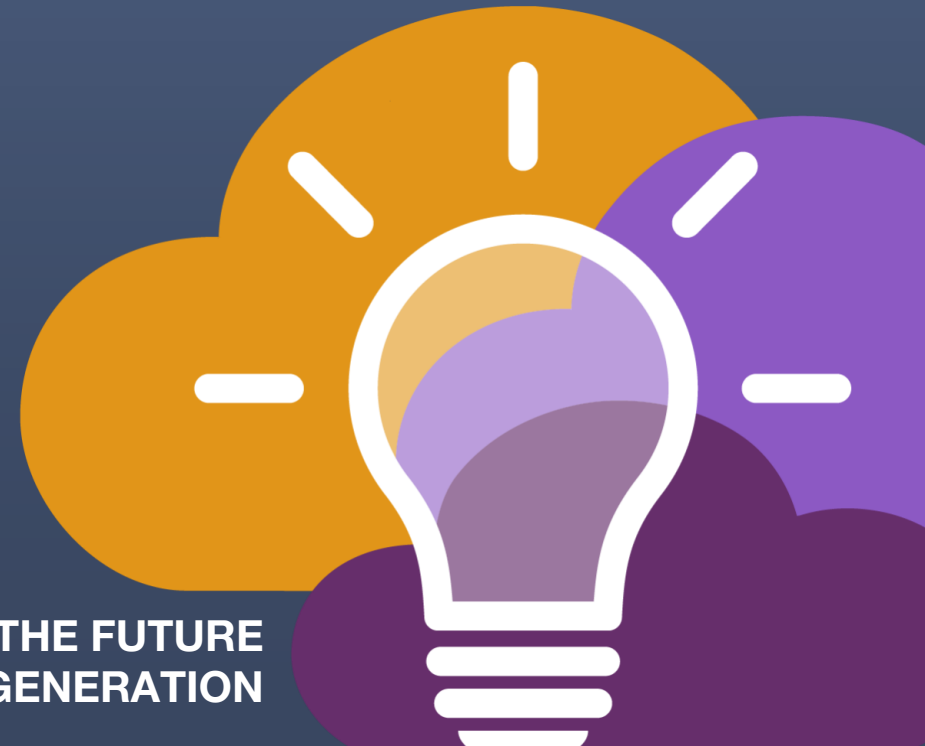
rows per page: 25 ▾



ScHARe

Data Analysis

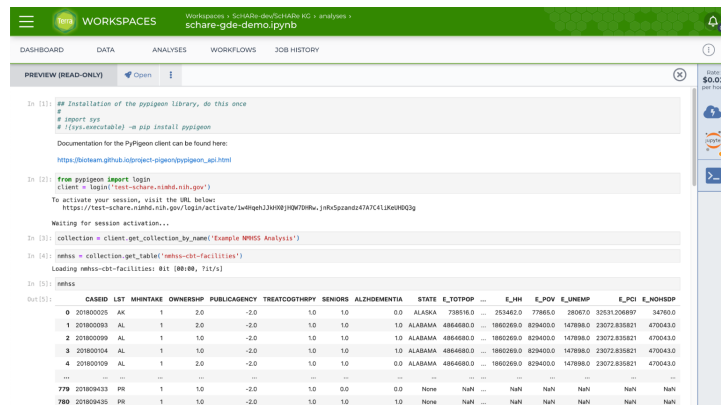
BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



How do I analyze or aggregate data from the ScHARe Repository?

Three quick options:

Connect to a Terra Jupyter notebook



```
In [1]: # Installation of the pyypipe library, do this once
# import pyypipe
# !pip install pyypipe
Documentation for the PyPipen client can be found here:
https://biocompare.com/pyypipe/pyypipe.html

In [2]: from pyypipe import login
client = login('test-schare-nchd.nih.gov')
To activate your session, visit the URL below:
https://test-schare-nchd.nih.gov/login/activate/14f1qth13H9BjHw70Rw_jmhSpaand47ATC1L6a8BQ3g
Waiting for session activation...

In [3]: collection = client.get_collection_by_name('Example NMHS Analysis')

In [4]: rehs = collection.get_table('rehs-cst-facilities')
Loading rehs-cst-facilities: 811 (88.18%, 711/s)

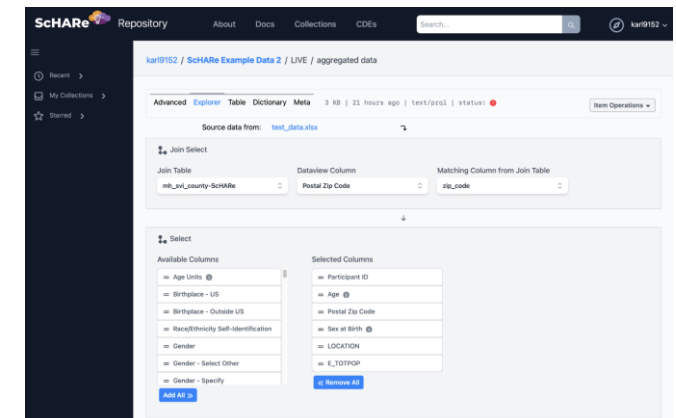
In [5]: rehs
```

CASEID	LIST	MINITAKE	OWNERSHIP	PUBLICAGENCY	THREATCOSTHPY	SENORS	ALZHEIMERIA	STATE	E_TOTPOP	E_HH	E_PSV	E_UNEMP	E_POI	E_HOHSOP
0	201800003	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	758516.0	253482.0	77860.0	280470.0
1	201800009	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	758516.0	253482.0	77860.0	280470.0
2	201800009	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	758516.0	253482.0	77860.0	280470.0
3	201800104	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	758516.0	253482.0	77860.0	280470.0
4	201800109	AL	1	2.0	-2.0	1.0	1.0	0.0	ALABAMA	4864680.0	758516.0	253482.0	77860.0	280470.0
...
779	201800433	PR	1	1.0	-2.0	1.0	0.0	0.0	None	NAN	NAN	NAN	NAN	NAN
780	201800435	PR	1	1.0	-2.0	1.0	1.0	1.0	None	NAN	NAN	NAN	NAN	NAN

Download to your own computer



Use Dataviews on the Repository



Look for a future Think-a-Thon where we will share more about how to link your data to Terra and how to use the ScHARe Repository to create aggregated data sets!



Analyzing Data on Terra (Jupyter Notebooks)

Terra

WORKSPACES

Workspaces > SchARE-dev/SchARE KG > analyses > schare-gde-demo.ipynb

0

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

PREVIEW (READ-ONLY) Open ⋮

In [1]:

```
## Installation of the pypigeon library, do this once
#
# import sys
# ![sys.executable] -m pip install pypigeon
```

Documentation for the PyPigeon client can be found here:
https://bioteam.github.io/project-pigeon/pypigeon_api.html

In [2]:

```
from pypigeon import login
client = login('test-schare.nimhd.nih.gov')
```

To activate your session, visit the URL below:
<https://test-schare.nimhd.nih.gov/login/activate/1w4HqehJJkHX0jHQW7DHRw.jnRx5pzandz47A7C4liKeUHDQ3g>

Waiting for session activation...

In [3]:

```
collection = client.get_collection_by_name('Example NMHSS Analysis')
```

In [4]:

```
nmhss = collection.get_table('nmhss-cbt-facilities')
```

Loading nmhss-cbt-facilities: 0it [00:00, ?it/s]

In [5]:

```
nmhss
```

Out[5]:

	CASEID	LST	MHINTAKE	OWNERSHP	PUBLICAGENCY	TREATCOGTHRPY	SENIORS	ALZHDementia	STATE	E_TOTPOP	...	E_HH	E_POV	E_UNEMP	E_PCI	E_NOHSDP
0	201800025	AK	1	2.0	-2.0	1.0	1.0	0.0	ALASKA	738516.0	...	253462.0	77865.0	28067.0	32531.206897	34760.0
1	201800093	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
2	201800099	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
3	201800104	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
4	201800109	AL	1	2.0	-2.0	1.0	1.0	0.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
...
779	201809433	PR	1	1.0	-2.0	1.0	0.0	0.0	None	NaN	...	NaN	NaN	NaN	NaN	NaN
780	201809435	PR	1	1.0	-2.0	1.0	1.0	1.0	None	NaN	...	NaN	NaN	NaN	NaN	NaN

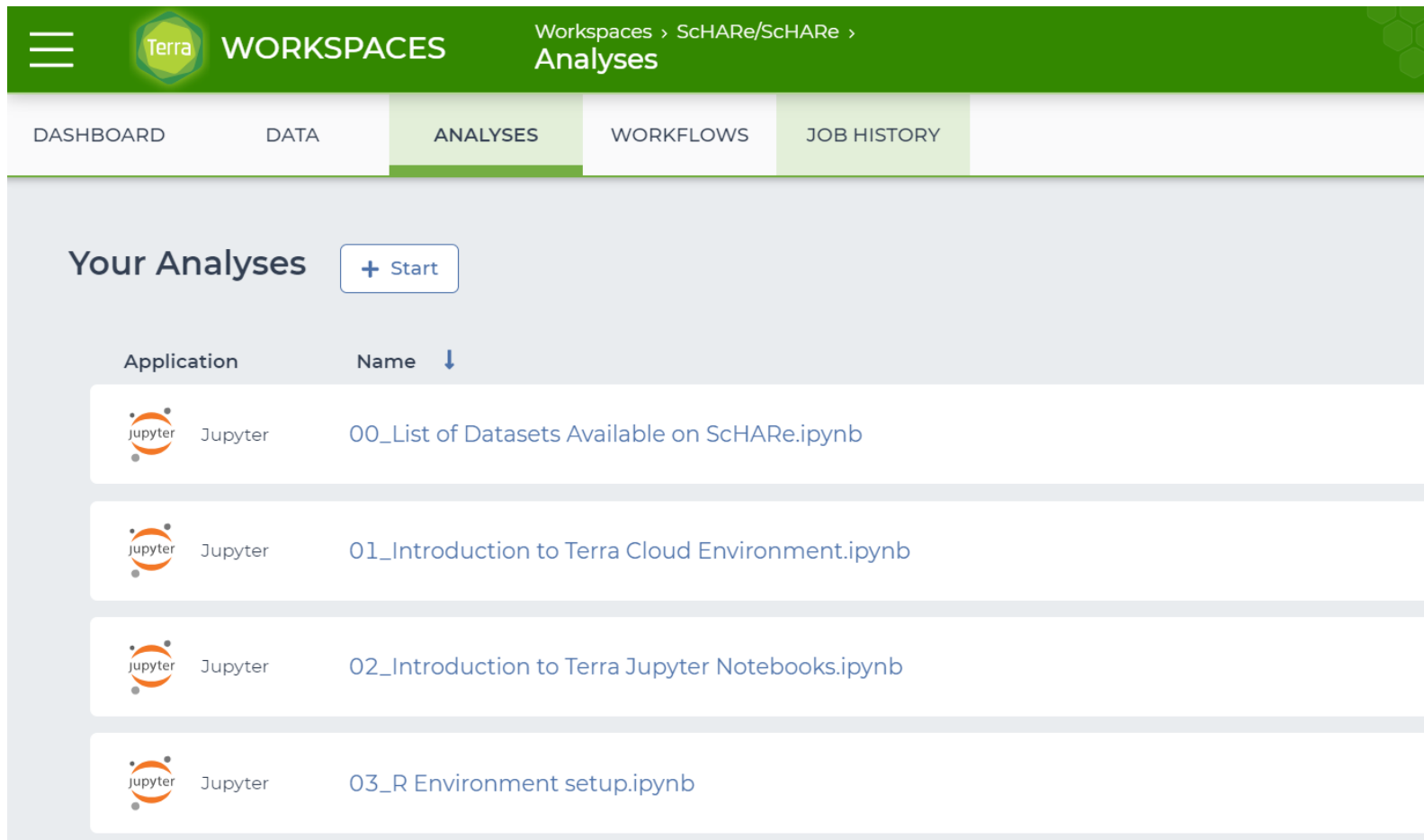
Rate: \$0.02 per hour

Jupyter

>



Analyzing Data on Terra (Jupyter Notebooks)




The screenshot shows the Terra WORKSPACES interface. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES", and a breadcrumb trail: "Workspaces > ScHARe/ScHARe > Analyses". Below this is a horizontal menu with tabs: "DASHBOARD", "DATA", "ANALYSES" (which is highlighted), "WORKFLOWS", and "JOB HISTORY". The main content area is titled "Your Analyses" and includes a "+ Start" button. Below this is a table with two columns: "Application" and "Name". The table lists four Jupyter notebooks, each with a Jupyter logo in the "Application" column and a filename in the "Name" column.

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



Analyzing Data on Terra (Jupyter Notebooks)

 **WORKSPACES**

Workspaces > ScHARe/ScHARe > analyses > 07_How to access plot and save data from ScHARe hosted datasets using Python 3.ipynb

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

PREVIEW (READ-ONLY) Open ⋮

Load dataset into a dataframe in our notebook Virtual Machine

Let's now get all of the data and load it into a pandas dataframe in our notebook Virtual Machine.

Pandas (<http://pandas.pydata.org/pandas-docs/stable/>) is a popular Python package for data manipulation to the Cloud Environment Virtual Machine running this notebook.

A dataframe is a 2-dimensional labeled data structure with columns of potentially different types. You can think of it like a spreadsheet or table. It is generally the most commonly used pandas object.

```
In [5]: # Let's create a dataframe. We will use the file location path copied earlier.
# Filepath = 'gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/HealthBehaviors/BRFSS2021.csv'


BRFSS_table = pandas.read_csv("gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/HealthBehaviors/BRFSS2021.csv", sep=',')
```

We have created our dataframe. Let's now take a look at the top of our table:

```
In [6]: BRFSS_table.head()
```

```
Out[6]:
```

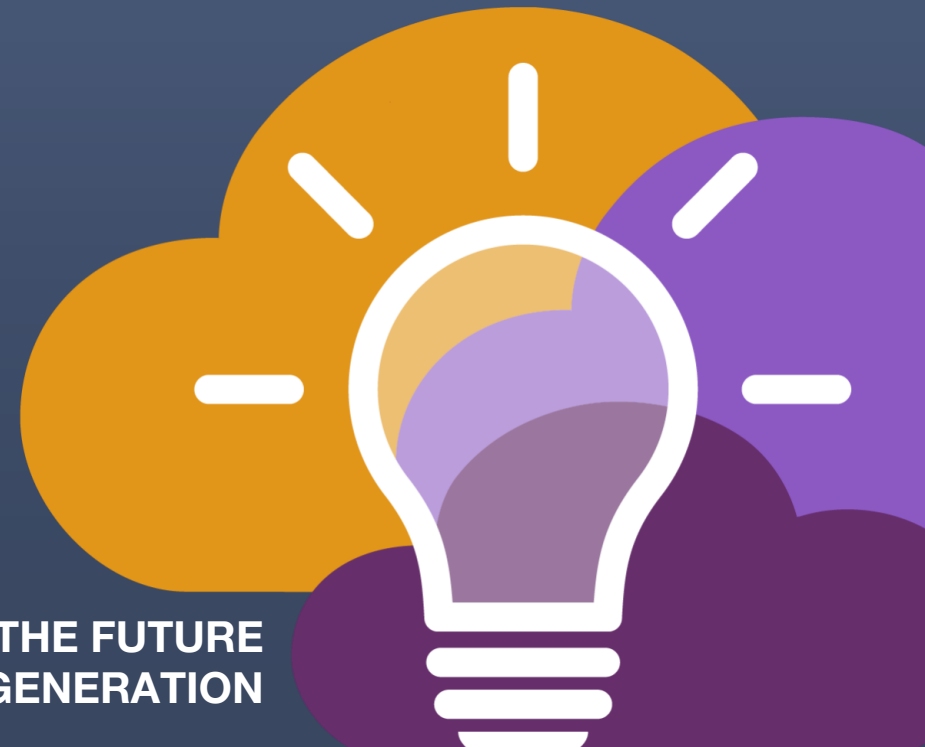
	_state	fmonth	idate	imonth	iday	iyear	dispcode	seqno	_psu	ctelenm1	...	_ftres1	_vegres1	_frutsu1	_vegesu1	_frlt1a	_vegl1a	_frt16a	_veg23a	_fruit
0	1	1	1192021	1	19	2021	1100	2021000001	2021000001	1.0	...	1	1	100.0	214.0	1	1	1	1	
1	1	1	1212021	1	21	2021	1100	2021000002	2021000002	1.0	...	1	1	100.0	128.0	1	1	1	1	
2	1	1	1212021	1	21	2021	1100	2021000003	2021000003	1.0	...	1	1	100.0	71.0	1	2	1	1	



ScHARe

Conclusion

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



ScHARe Terra

In Summary:

- Register for the ScHARe/Terra Platform
- Access federated data – especially SDoH and population science
- Create your own workspace – can share with others
- Assess computational tools, such as python and R coding

ScHARe Repository

In Summary: (Comply with the NIH Data Management and Sharing Policy Requirement)

- When getting started, first create a collection, provide metadata, and upload documents.
- Upload your data and use the system to help you map to the ScHARe CDEs.
- View your data, see CDE compliance and analysis readiness.
- Data can be shared with your colleagues, and can be made publicly available after review.
- You can use dataviews to create subsets of your data, join datasets together and more.
- Data can also be analyzed on the ScHARe Terra workspace.



We are here to support you!

We want to hear your questions, issues and comments about the SchARe Repository!

For any questions regarding how to use the Repository, please reach out to:

✉ schare@mail.nih.gov

Response time: within 24 hours

For Technical Support inquiries, contact:

✉ schare-repository-support@bioteam.net

🕒 Office Hours: Wednesdays, 4-5 PM US Eastern

Questions?



Slido Poll

What concerns or barriers do you have regarding sharing your data publicly?

Think-a-Thon poll

1. Rate how useful this session was:

- ☐ Very useful
- ☐ Useful
- ☐ Somewhat useful
- ☐ Not at all useful

Think-a-Thon poll

2. Rate the pace of the instruction for yourself:

- ☐ Too fast
- ☐ Adequate for me
- ☐ Too slow

Think-a-Thon poll

3. How likely will you participate in the next Think-a-Thon?

- ☐ Very interested, will definitely attend
- ☐ Interested, likely will attend
- ☐ Interested, but not available
- ☐ Not interested in attending any others

ScHARe

Thank you



ScHARe

Next Think-a-Thons:



bit.ly/think-a-thons

Register for ScHARe:



bit.ly/join-schare



schare@mail.nih.gov

