



**ScHARe**

# ScHARe Repository Introduction

October 16, 2024

**Deborah Duran**, PhD • NIMHD

**Luca Calzoni**, MD MS PhD Cand. • NIMHD

**Elif Dede Yildirim**, 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

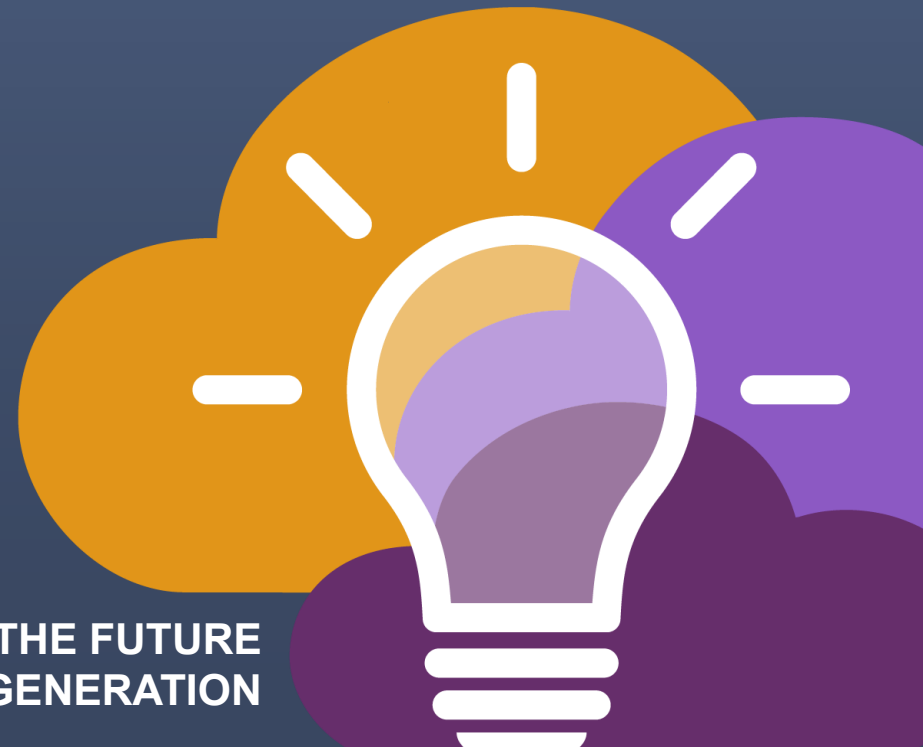
**I am interested in (check all that apply):**

- 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](https://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

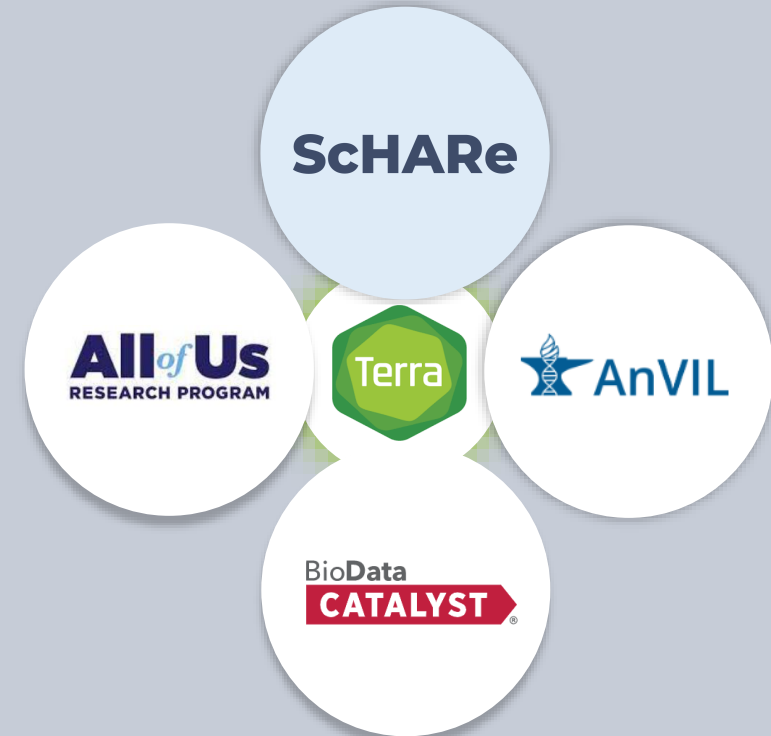
- Secure workspaces
- Data storage
- Computational resources
- Tutorials (how to)
- Copy-and-paste code in Python and R
- Learning Terra on ScHARe prepares you to use other NIH platforms



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



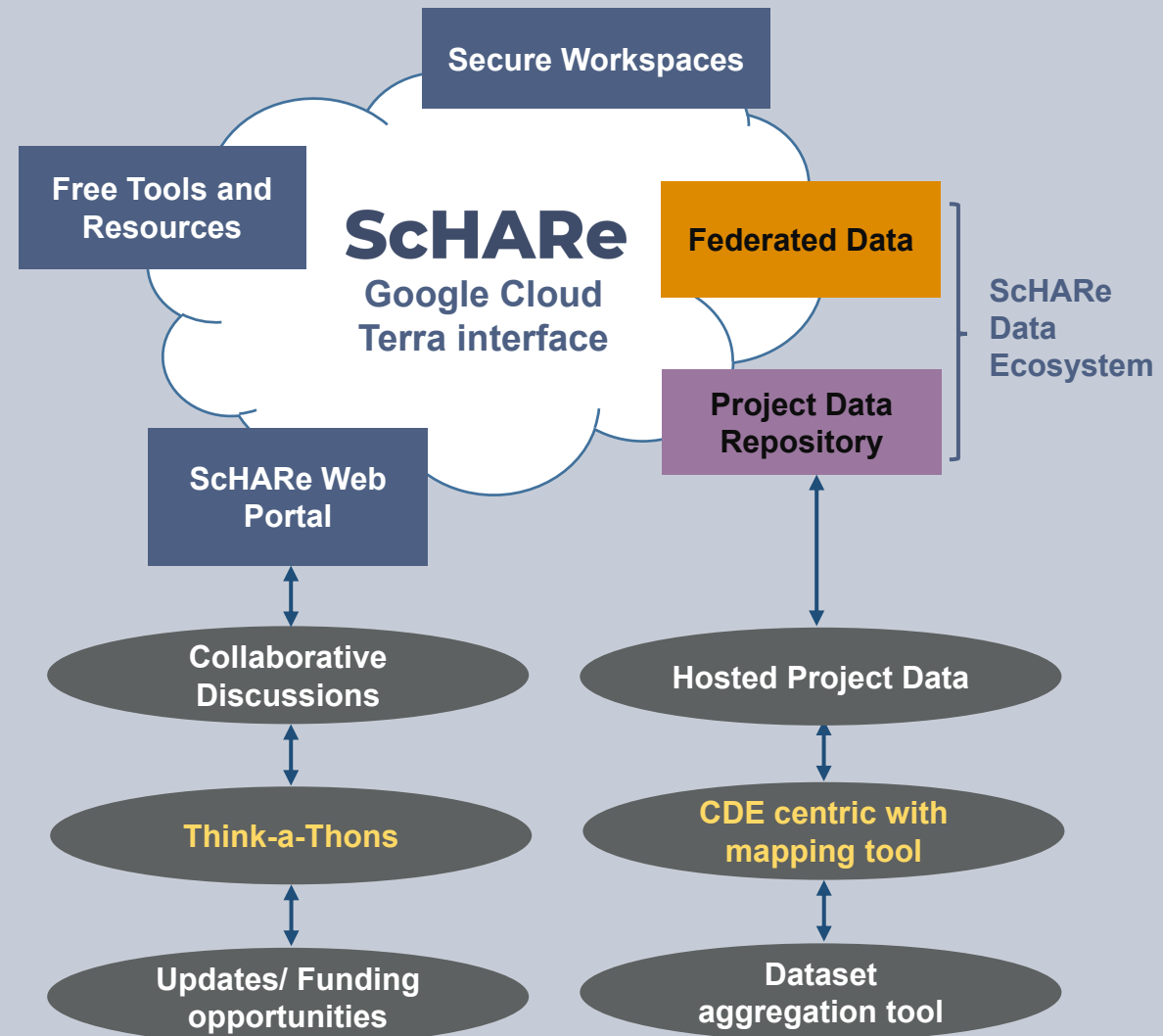


# ScHARe Components

## 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

The screenshot displays the ScHARe Terra interface with a 'Share Workspace' modal open. The modal includes a 'User email' input field with an 'ADD' button, a 'Current Collaborators' list, and a 'Share with Support' toggle. The background shows a workspace list with columns for Name, recently viewed dates, and search filters.

Name	Recently Viewed
ScHARe	Viewed Apr 14, 2023, 11:58 AM
ScHARe Think-a-Thons	Viewed Apr 10, 2023, 10:00 AM

**Share Workspace**

User email:  **ADD**

**Current Collaborators**

calzonil2@nih.gov	Owner	<input checked="" type="checkbox"/> Can share	<input checked="" type="checkbox"/> Can compute
ScHARe-Contractors@firecloud.org	Writer	<input type="checkbox"/> Can share	<input type="checkbox"/> Can compute
ScHARe-Read-Only-Access@firecloud.org	Reader	<input type="checkbox"/> Can share	<input type="checkbox"/> Can compute

Share with Support:  No **CANCEL** **SAVE**

- 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

WORKSPACES  
Workspaces > ScHARe/ScHARe > Analyses

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

Your Analyses + START

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

Modular codes

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

WORKSPACES  
Workspaces > ScHARe/ScHARe > ANALYSES

DASHBOARD DATA ANALYSES

WORKFLOWS

Find a Workflow

Suggested Workflows

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

Find Additional Workflows

Dockstore  
Browse WDL workflows in Dockstore, an open platform used by the GA4GH for sharing Docker-based workflows.

- Modular codes developed for reuse
- Adding SAS

# ScHARe Terra interface: access to datasets

What data?

The screenshot shows the ScHARe Terra interface in the 'Analyses' tab. The notebook '00\_List of Datasets Available on ScHARe.ipynb' is open, displaying the 'The ScHARe Data Ecosystem' section. This section provides a comprehensive list of datasets available in the ScHARe Data Ecosystem, categorized by content. The categories include:

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

Where?

The screenshot shows the ScHARe Terra interface in the 'Data' tab. The 'Data' table is displayed, showing a list of tables with search and filter options. The 'EconomicStability' table is highlighted, showing its contents.

TABLES	EDIT	OPEN WITH...	EXPORT	SETTINGS	0 rows selected
<input type="checkbox"/> 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					0.185

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:

**250+**  
FEDERATED  
PUBLIC  
DATASETS

## Public datasets

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

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

**CDE**  
FOCUSED  
REPOSITORY

## Funded datasets

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

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

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

# ScHARe Ecosystem

OVER 260 DATA SETS CENTRALIZED

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

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

with the addition of:

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

The screenshot shows the Terra WORKSPACES Data interface. The top navigation bar includes 'DASHBOARD', 'DATA', 'ANALYSES', 'WORKFLOWS', and 'JOB HISTORY'. The 'DATA' tab is active, displaying an 'IMPORT DATA' button and a search bar for tables. A list of tables is shown on the left, with 'EconomicStability (62)' highlighted. The main table on the right lists datasets with columns for 'EconomicStability\_id' and 'SizeGb'. The datasets listed include 'FoodAccessResearchAtlasData2010', 'CurrentPopulationSurvey\_FoodSecuritySupplement\_2011', 'CurrentPopulationSurvey\_FoodSecuritySupplement\_2012', 'CurrentPopulationSurvey\_FoodSecuritySupplement\_2013', 'CurrentPopulationSurvey\_FoodSecuritySupplement\_2014', 'AHS\_National\_Household\_2015', 'AHS\_National\_Mortgage\_2015', 'AHS\_National\_Person\_2015', 'AHS\_National\_Project\_2015', and 'CurrentPopulationSurvey\_FoodSecuritySupplement\_2015'.

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





# ScHARe Ecosystem: ScHARe hosted datasets

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

260+ datasets

- What are the Social Determinants of Health?

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

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



[https://www.cdc.gov/about/priorities/social-determinants-of-health-at-cdc.html?CDC\\_AAref\\_Val=https://www.cdc.gov/about/sdoh/index.html](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 06_How to access plot and save data from public BigQuery datasets using Python 3.ipynb
```

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

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

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

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

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

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



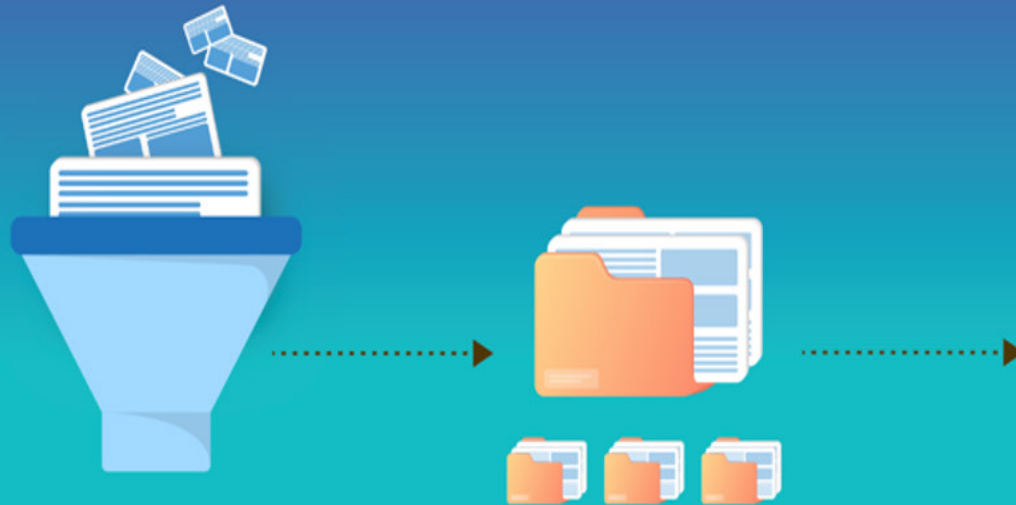
## CDE benefits:

- 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](https://cde.nlm.nih.gov/resources)



# ScHARe Core CDEs

PhenX Toolkit

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

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

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

\* Project Level CDEs

## NIH Endorsed



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

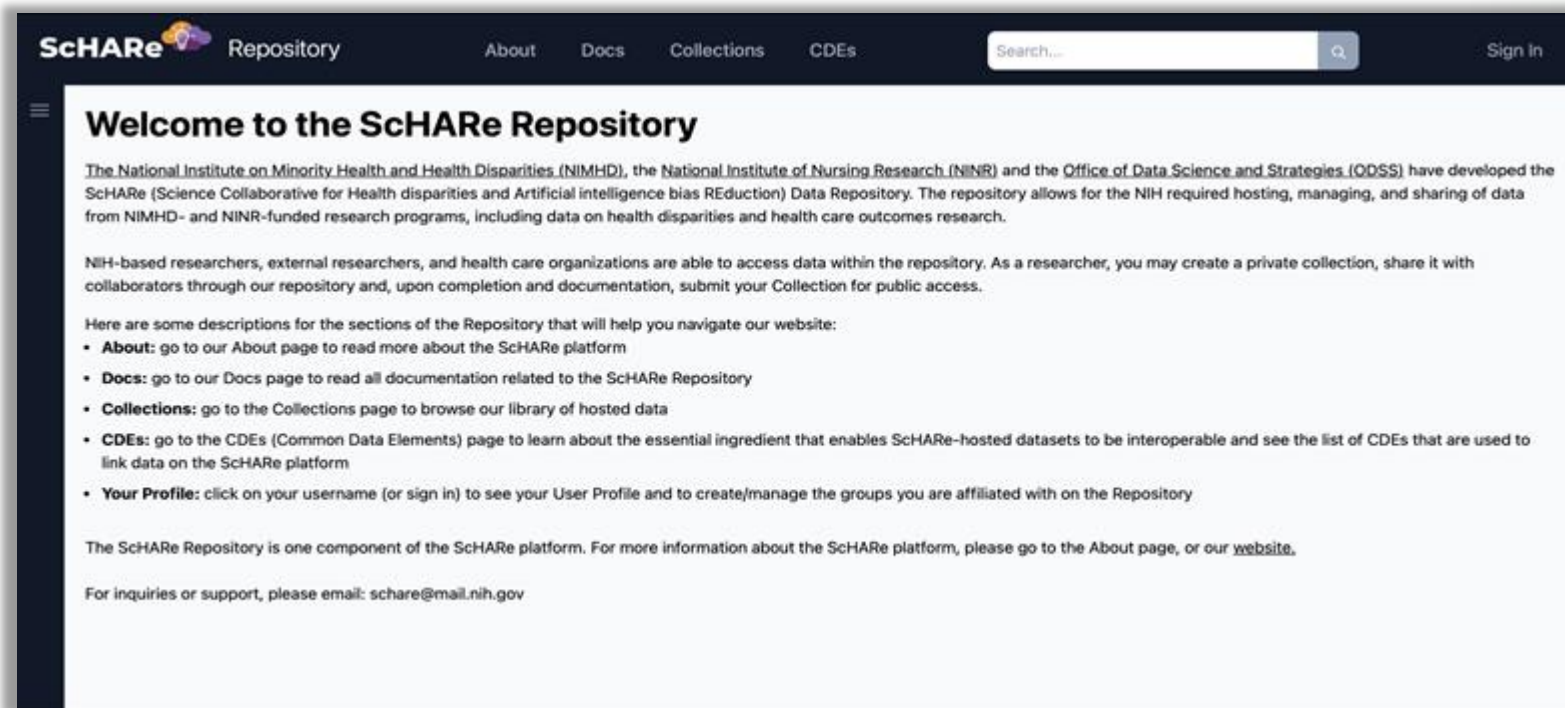
**NIH CDE Repository:**

[cde.nlm.nih.gov/home](https://cde.nlm.nih.gov/home)

**PhenX Toolkit:**

[www.nimhd.nih.gov/resources/phenx/](https://www.nimhd.nih.gov/resources/phenx/)

# ScHARe Repository



The screenshot shows the ScHARe Repository website. The header includes the ScHARe logo, navigation links for About, Docs, Collections, and CDEs, a search bar, and a Sign In button. The main content area features a welcome message, a paragraph describing the repository's origin and purpose, a paragraph about user access, a list of navigation tips, and contact information.

**ScHARe Repository** About Docs Collections CDEs 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](#).

For inquiries or support, please email: [schare@mail.nih.gov](mailto:schare@mail.nih.gov)

- 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



# ScHARe

## Research Think-a-Thons

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



# Generational career & discipline exchange





# Think-a-Thons

## Goals:

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



## 1. TUTORIAL AND TARGETED THINK-A-THONS

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



## 2. RESEARCH THINK-A-THONS

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

**Register:**  
[bit.ly/think-a-thons](https://bit.ly/think-a-thons)



# Think-a-Thon tutorials

[bit.ly/think-a-thons](https://bit.ly/think-a-thons)

February

**Artificial Intelligence and Cloud Computing 101**

March

**ScHARe 1 – Accounts and Workspaces**

April

**ScHARe 2 – Terra Datasets**

May

**ScHARe 3 – Terra Google-hosted Datasets**

June

**ScHARe 4 – Terra ScHARe-hosted Datasets**

July

**An Introduction to Python for Data Science – Part 1**

August

**An Introduction to Python for Data Science – Part 2**

September

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

October

**Preparing for AI 1: Common Data Elements and Data Aggregation**

November

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

January

**Preparing for AI 3: Computational Data Science Strategies 101**

February/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**

## SPECIAL EVENTS

- 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

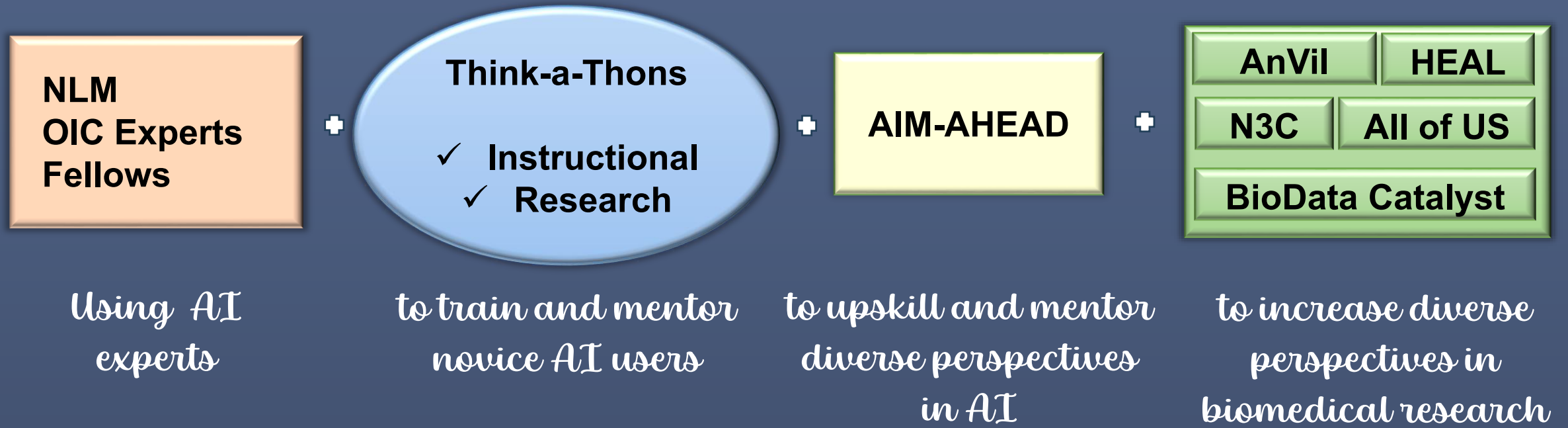
***Accessible:*** wide audience

***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



**ScHARe**

# ScHARe Repository Introduction

October 16, 2024

**Deborah Duran**, PhD • NIMHD

**Luca Calzoni**, MD MS PhD Cand. • NIMHD

**Elif Dede Yildirim**, PhD • NIMHD



# ScHARe

Repository  
Background

BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION



# What is the ScHARe Repository?

The ScHARe Repository is a data repository for hosting and sharing population science data.

The repository also offers tools to explore, aggregate and manipulate that data.

## Datasets Collection

Relevant to health disparities and health care outcomes research, including SDOH data.

## Data Repository

Hosting, managing, and sharing of data from federated funded research programs across all communities



## Bias Mitigation Tools + Resources

Manage bias associated with datasets and algorithms, informing ethical and inclusive decisions.

## Collaborative Workspaces

Access and analyze public and controlled-access data in collaborative spaces.

# 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

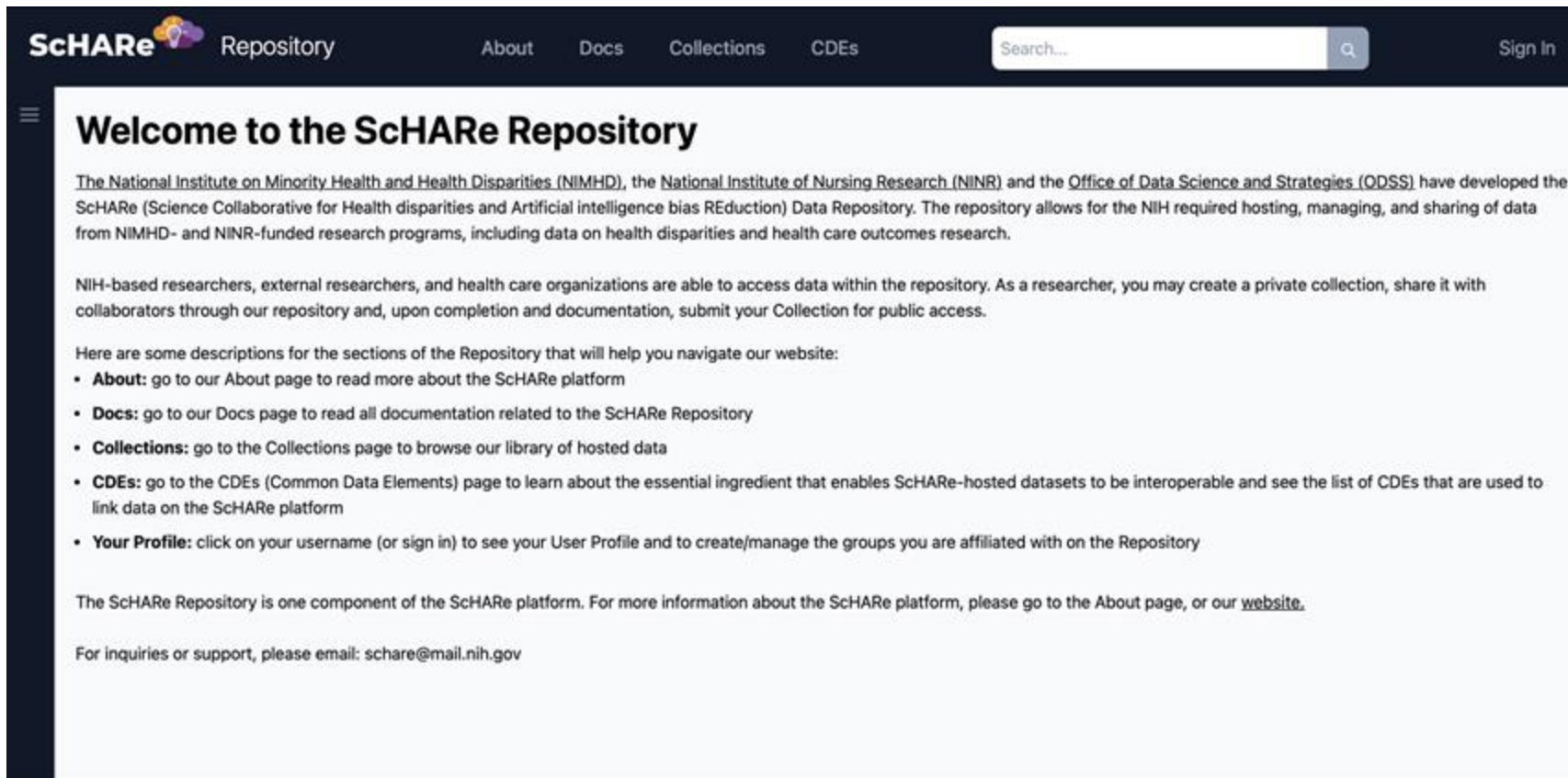
Getting Started

BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION



# Visit the ScHARe Repository!

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



The screenshot shows the ScHARe Repository website. The header includes the ScHARe logo, the word "Repository", and navigation links for "About", "Docs", "Collections", and "CDEs". There is a search bar and a "Sign In" button. The main content area features a "Welcome to the ScHARe Repository" heading, followed by a paragraph describing the repository's development by NIMHD, NINR, and ODSS. Below this is a paragraph about data access for researchers and health care organizations. A section titled "Here are some descriptions for the sections of the Repository that will help you navigate our website:" contains a bulleted list of links to various parts of the site. At the bottom, there is a note about the repository being part of the ScHARe platform and contact information for inquiries.

**ScHARe** Repository

About Docs Collections CDEs 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](#).

For inquiries or support, please email: [schare@mail.nih.gov](mailto:schare@mail.nih.gov)



# Visit the ScHARe Repository!

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

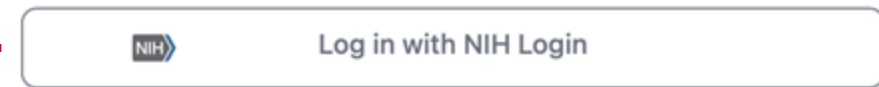
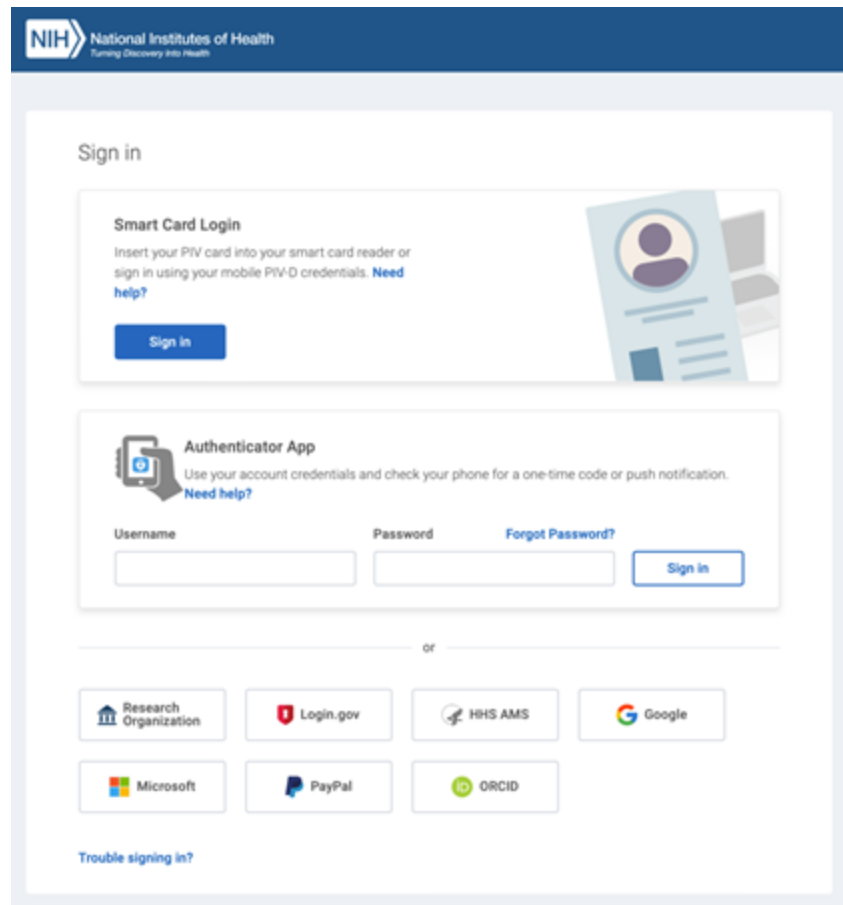
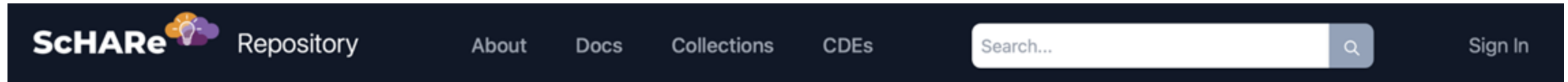
For today's training, please use our test site:

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

Data uploaded to this site will not be retained.



# Log In to the Repository



- **Select the most appropriate login method**
  - NIH Smart Card Login for NIH affiliates
  - Otherwise, Research Organization, Login.gov, HHS AMS preferred
- **Be sure to accept the Terms of Use**

# A Quick Tour of the Site

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)

Log in, log out, and user  
information

Quick links to recently  
accessed and your  
own data

The screenshot shows the SchARE Repository website interface. At the top, there is a navigation bar with the SchARE logo and the word 'Repository'. To the right of the logo are navigation links for 'About', 'Docs', 'Collections', and 'CDEs'. Further right is a search bar with the placeholder text 'Search...' and a magnifying glass icon. On the far right of the navigation bar is a user profile icon and a dropdown arrow.

Below the navigation bar, the main content area features a large heading: 'Welcome to the SchARE Repository'. Underneath this heading is a paragraph of introductory text. Below the text is a bulleted list of descriptions for the website sections. On the left side of the main content area, there is a dark sidebar with a hamburger menu icon at the top. Below the menu icon are three items: 'Recent' with a clock icon, 'My Collections' with a folder icon, and 'Starred' with a star icon. Each item has a right-pointing arrow.

**SchARE** Repository

About Docs Collections CDEs

Search...

## 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

Recent >

My Collections >

Starred >

# Viewing the ScHARe CDEs

- Browse through the list of ScHARe CDEs
- Expand an individual CDE's details to see definitions, permissible values, and concept codes

**ScHARe** Repository About Docs Collections CDEs Search... karl9152

## 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 | GNVvEdFQR0:0001

#### Birthplace

Birthplace - US v string | 5 definitions | 3 concepts | 0qP44x2oX:0001

Name	Data Type	ID	Version	Sensitivity
Birthplace - US	string	0qP44x2oX	0001	Not specified

Definitions

preferredQuestionText:  
Where were you born? (In the United States, including U.S. Territories--Puerto Rico, Guam, U.S. Virgin Islands, American Samoa and Northern Mariana Islands). If you were born outside the US, skip to the next question.

alsoKnownAs:  
<https://cde.nlm.nih.gov/api/de/0qP44x2oX>

longDescription:  
The location where someone was born - a U.S. state or territory

# ScHARe

Uploading your  
first Data Set

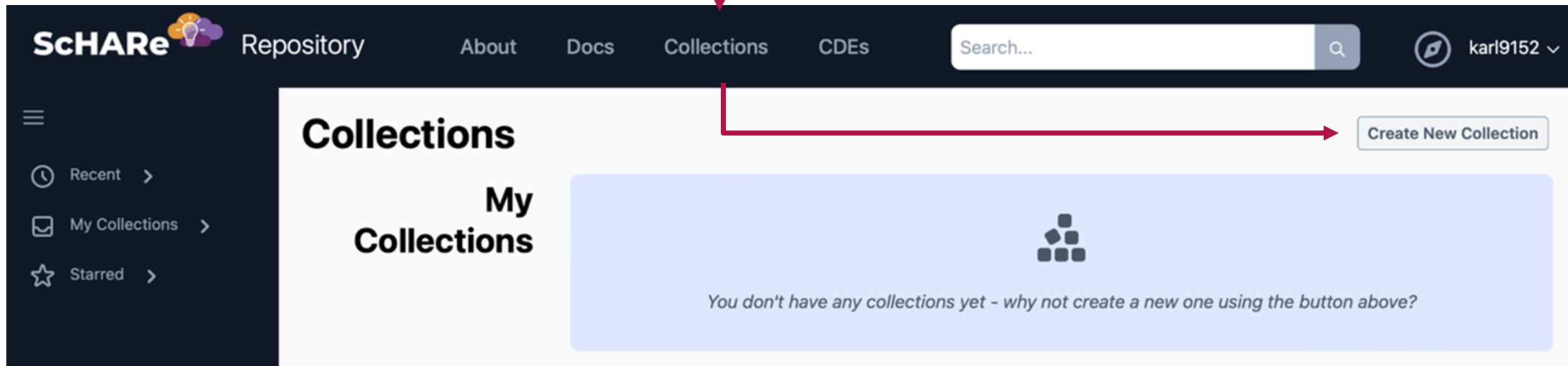
BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION





# Creating a Collection

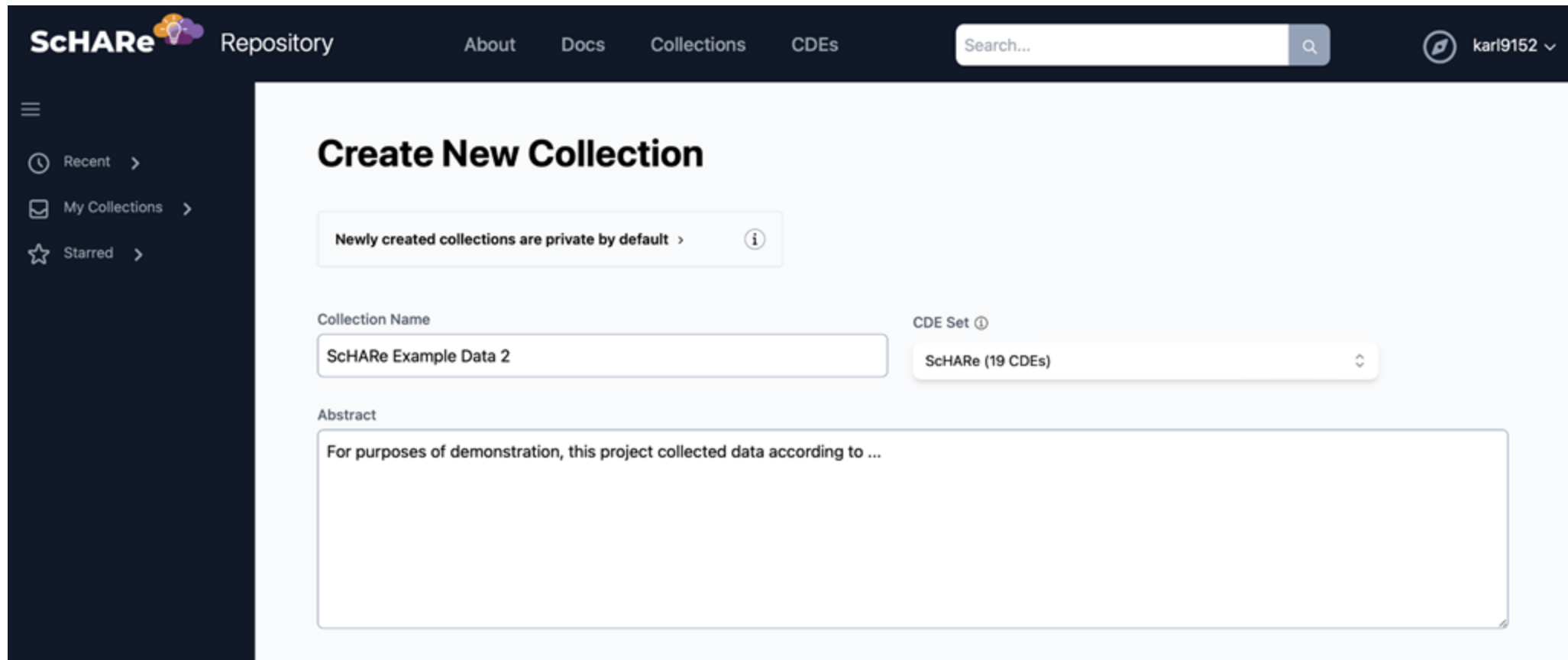
- **Collections** are a place where you can describe and store your data and any related metadata. They can be shared with colleagues and published when you're ready. You can upload any file type!
- The first step in uploading data is to create a collection



The screenshot displays the SchARE Repository interface. The top navigation bar includes the SchARE logo, 'Repository', 'About', 'Docs', 'Collections', and 'CDEs'. A search bar is located to the right of the navigation items, and the user's profile 'karl9152' is visible in the top right corner. On the left side, there is a sidebar with a menu icon and three items: 'Recent', 'My Collections', and 'Starred'. The main content area is titled 'Collections' and 'My Collections'. A large light blue box contains a collection icon and the text: 'You don't have any collections yet - why not create a new one using the button above?'. A red arrow points from the 'Collections' menu item in the top navigation bar to the 'Create New Collection' button in the top right corner of the main content area.

# Creating a Collection

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



The screenshot shows the 'Create New Collection' page in the ScHARe Repository. The page has a dark blue header with the ScHARe logo and navigation links for 'About', 'Docs', 'Collections', and 'CDEs'. A search bar and a user profile 'karl9152' are also visible in the header. On the left, a sidebar contains navigation options: 'Recent', 'My Collections', and 'Starred'. The main content area is titled 'Create New Collection' and features a notification: 'Newly created collections are private by default'. Below this, there are two input fields: 'Collection Name' with the value 'ScHARe Example Data 2' and 'CDE Set' with a dropdown menu showing 'ScHARe (19 CDEs)'. At the bottom, there is a large text area for the 'Abstract' containing the text: 'For purposes of demonstration, this project collected data according to ...'.

# Tags and Project-Level CDEs

- Apply relevant tags and project-level CDEs to your dataset
- Refer to the NIMHD Research Framework for more information on selecting project-level CDEs:
  - <https://www.nimhd.nih.gov/about/overview/research-framework/nimhd-framework.html>

The screenshot shows a web form with the following sections:

- Tags:** Two selected tags: "Diet and Nutrition" and "Food Insecurity". Below them is a text input field labeled "Enter a Tag".
- Project CDEs:**
  - Levels of Influence:** A dropdown menu with "Individual" and "Community" selected.
  - Domains of Influence:** A dropdown menu with "Select all that apply..." selected.
  - Research Areas:** A list of options: "Biological", "Behavioral", "Physical/Built Environments", "Sociocultural Environment", and "Health Care Systems and Clinical Care".
  - Research Focus:** A section with no visible options.
- Advanced User-Defined Metadata:** A section with a right-pointing arrow.
- Submit:** A blue button at the bottom left.

Callout boxes provide instructions:

- "Click Submit to continue" points to the Submit button.
- "Select tags from a dropdown list, or type in your own if none apply" points to the "Enter a Tag" input field.

# 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

		Levels of Influence*			
		Individual	Interpersonal	Community	Societal
Domains of Influence <i>(Over the Lifecourse)</i>	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 – 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



# 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
- Native Hawaiian and 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.

# The Collection Main Page

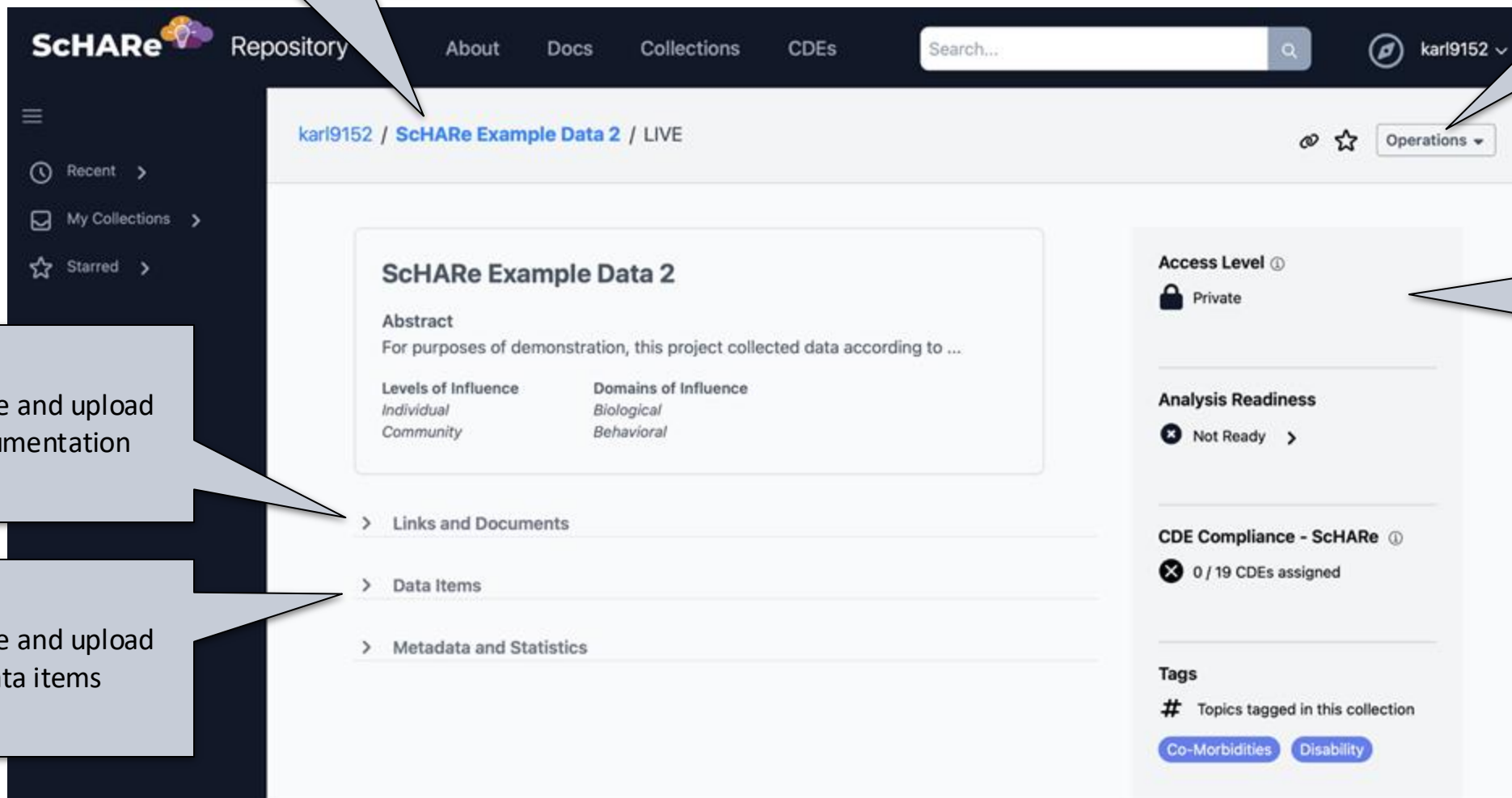
Click the collection name to return to the main page

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

Sidebar with key information

Browse and upload documentation

Browse and upload data items



## SCHARE Example Data 2

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

Levels of Influence	Domains of Influence
Individual	Biological
Community	Behavioral

> Links and Documents

> 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

Co-Morbidities Disability

# Adding Links and Documents

- Upload any document, presentation, etc.
- Link to papers through Pubmed

Links and Documents

No links or documents have been uploaded yet.

Add New

## Add Document or Link

Upload

Pubmed Search

Link

TYPE

Presentation

UPLOAD

Choose File



BioTeam\_ Dat...inciples.pdf ✓

TITLE

BioTeam - Data Ecosystem Principles

DESCRIPTION

A brief summary of trends in principles of data ecosystems

Add

Cancel

## Add Document or Link

Upload

Pubmed Search

Link

PUBMED SEARCH

ecosystem

Search

RESULTS

**A critical review on the active anti-viral metaboli...**

Naz A, Chowdhury A, Pareek S, Kumar P, Poddar NK  
J Complement Integr Med, (), 2024 Oct 10  
doi: 10.1515/jcim-2024-0186

**Advancing stem cell technologies for conservati...**

Hutchinson AM, Appeltant R, Burdon T, Bao Q, Bargaj...  
Development, 151(20), 2024 Oct 15  
pii: dev203116. doi: 10.1242/dev.203116

Add

Cancel

# Uploading a Data File

You can upload to the ScHARe Repository...

## Any file

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

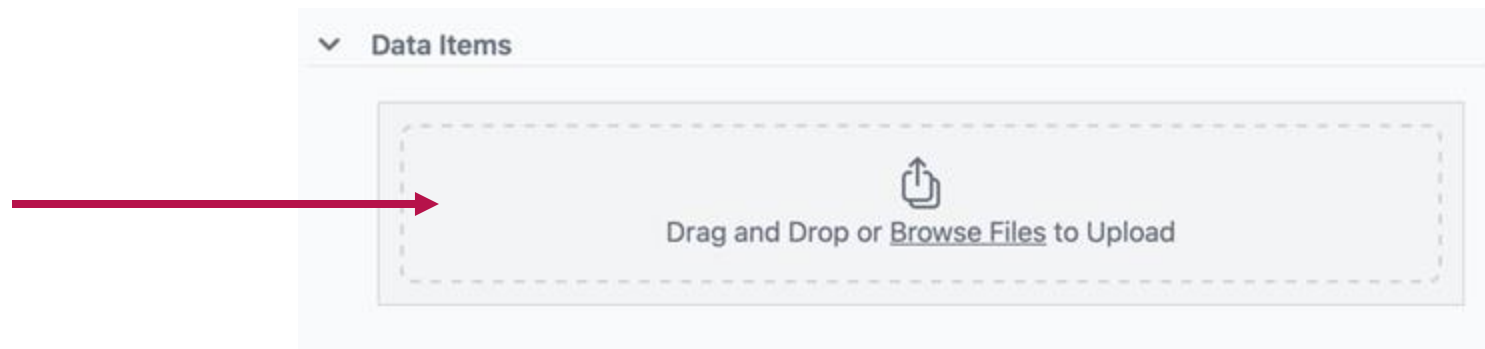
Some file formats can be previewed in the web interface

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

For today's training, we'll all use the same example data file.

## Visit:

<https://test-schare.nimhd.nih.gov/c/fwf4bfnf7b>

Or click "Collections" and find "ScHARe Think-a-Thon – Example Data"

**ScHARe Think-a-Thon - Example Data** LIVE

This collection holds example data for use during the ScHARe Think-a-Thon, October 2024.

30 minutes ago

... Expand the Data Items section

## Download:

The screenshot shows the 'Data Items' section of the ScHARe interface. A table lists data items with columns for STATUS, NAME, CREATED, and SIZE. The first row is highlighted in blue and contains a file icon, 'example-data.csv', '6 minutes ago', and '21 KB'. A context menu is open over this row, listing various actions: Rename..., Copy To..., Link To..., Move To Folder..., Configure Table, Assign Data Dictionary..., Create Dataview..., Import as REDCap Data Dictionary, Export Table Data..., Download (highlighted in blue), and Delete. A red arrow points to the three-dot menu icon in the table, and another red arrow points to the 'Download' button in the context menu.

STATUS	NAME	CREATED	SIZE
	example-data.csv	6 minutes ago	21 KB

- 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

# Viewing Uploaded Data

The screenshot displays the ScHARe Repository interface. At the top, the navigation bar includes 'ScHARe Repository', 'About', 'Docs', 'Collections', 'CDEs', and a search bar. The breadcrumb path is 'karl9152 / ScHARe Example Data 2 / LIVE'. The main content area features a title 'ScHARe Example Data 2', an abstract, and metadata including 'Levels of Influence' (Individual, Community) and 'Domains of Influence' (Biological, Behavioral). Below this is a table of data items with columns for STATUS, NAME, CREATED, and SIZE. A single item 'test\_data.xlsx' is listed with a green status icon and a creation time of 'an hour ago'. A 'Drag and Drop or Browse Files to Upload' area is visible at the bottom. On the right, a sidebar contains sections for 'Access Level' (Private), 'Analysis Readiness' (Ready), 'CDE Compliance - ScHARe' (2 / 19 CDEs assigned), and 'Tags' (Co-Morbidities, Disability).

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

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

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

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

# Viewing Uploaded Data

The screenshot shows the ScHARe Repository interface. At the top, there are navigation links: About, Docs, Collections, and a search bar. The main content area displays the file path: `karl9152 / ScHARe Example Data 2 / LIVE / test_data.xlsx`. Below the path, there are tabs for File, Table, Dictionary, and Meta. The 'Table' tab is active, showing a data table with columns: Participant ID, Age, Age Units, Birthplace - US, Birthplace - Out, Postal Zip Code, Race/Ethnicity, Sex at Birth, and Gender. The table contains five rows of data. A popup window is open over the 'Age' column, showing details for the CDE 'Age (number)'. The popup includes sections for Concepts, Definitions, and other metadata.

Click the collection name to return to the collection main page

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

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

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

Participant ID	Age	Age Units	Birthplace - US	Birthplace - Out	Postal Zip Code	Race/Ethnicity	Sex at Birth	Gender
0001	64	Years	AL					
0002	47	Years	NC					
0003								
0004								
0005								
0006	23	Years		Algeria				

Age (number)  
Concepts  
C25150: Age *dataElement*  
Definitions  
What is the person's age (in years if more than 24 months old or months if 24 months or younger)? *preferredQuestionText*  
<https://cde.nlm.nih.gov/api/de/PDjBiGXjO> *alsoKnownAs*  
The number of years (if more than 24 months old) or months (if 24 months or younger) that the person has been alive. *longDescription*  
National Longitudinal Survey *source*  
<https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/household/age>  
age, person, demographics *other*



**ScHARe**

# ScHARe Repository Introduction

October 16, 2024

**Deborah Duran**, PhD • NIMHD

**Luca Calzoni**, MD MS PhD Cand. • NIMHD

**Elif Dede Yildirim**, PhD • NIMHD



# ScHARe

## CDE Mapping

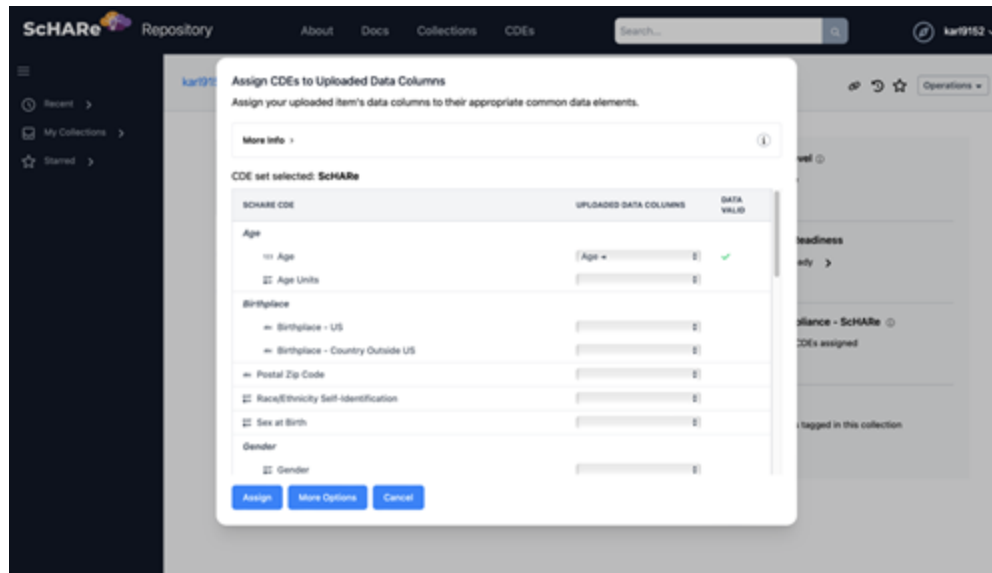


BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION

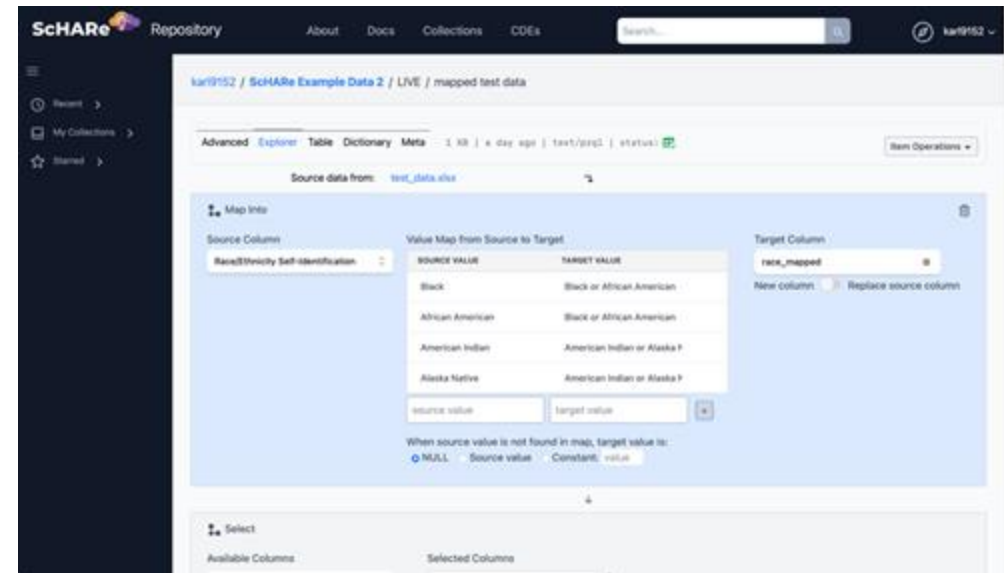
# Getting Started with Mapping CDEs

- 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



## Map CDEs using Dataviews



# Assigning CDEs at Upload

## 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

- After you upload a tabular data file, the system will recommend that you begin the process of assigning CDEs to your data column headings.

## 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



# Assigning CDEs at Upload

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

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

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
<b>Age</b>		
123 Age	Age	✓
Age Units	Age Units	✓
<b>Birthplace</b>		
Birthplace - US		
Birthplace - Country Outside US		
Postal Zip Code		
Race/Ethnicity Self-Identification	Race/Ethnicity Self-Identi	✗
<b>Gender</b>		
Gender		
Gender - Select Other		
Gender - Specify		
<b>Sexual Orientation</b>		

Row 1 "American Indian" is not one of ["American I...  
Row 2 "Alaska Native" is not one of ["American Indi...  
Row 3 "Asian" is not one of ["American Indian or AI...  
Row 4 "Asian American" is not one of ["American I...  
Row 5 "Black" is not one of ["American Indian or AI...  
Row 6 "African American" is not one of ["American ...  
Row 7 "Hispanic" is not one of ["American Indian o...  
Row 8 "Latino" is not one of ["American Indian or A...  
Row 9 "Spanish" is not one of ["American Indian or...  
Row 10 "Native Hawaiian" is not one of ["American ...

Assign More Options Cancel

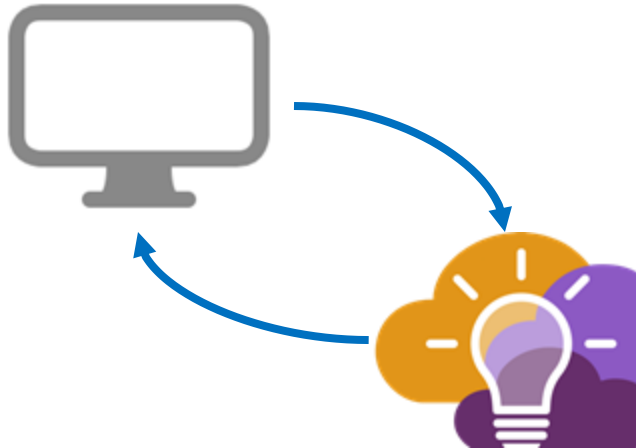
When an assigned CDE fails validation, hover over the red X to get details on why

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

# 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**



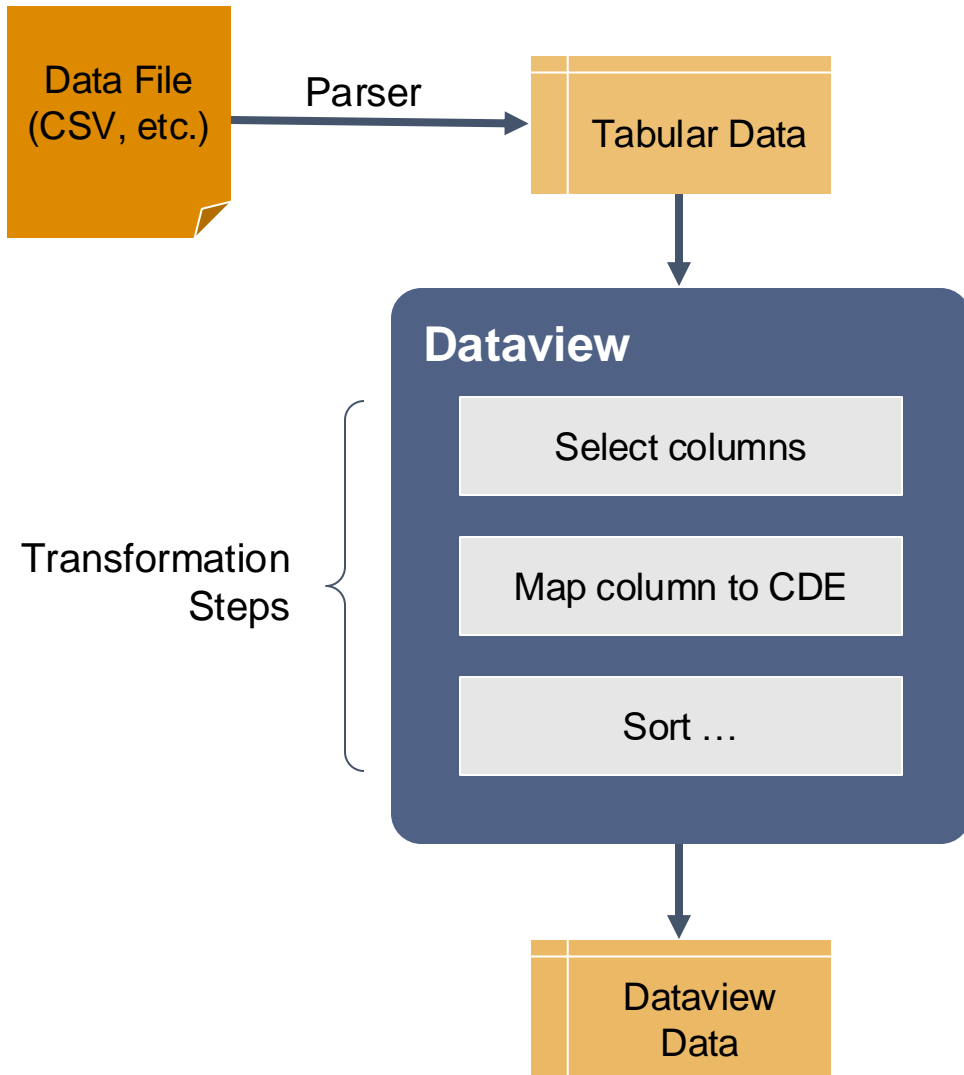
**Use a Dataview to Map to CDEs within the Repository**

The screenshot shows the SchARE Repository interface. The main content area displays a configuration for a Dataview. The source data is identified as 'test\_data.xlsx'. The 'Map Into' section shows a 'Value Map from Source to Target' table with the following entries:

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

Below the table, there are input fields for 'source value' and 'target value'. The 'Target Column' is set to 'race\_mapped'. The 'When source value is not found in map, target value is:' section has 'NULL' selected. The 'Select' section shows 'Available Columns' and 'Selected Columns'.

# What is a Dataview?

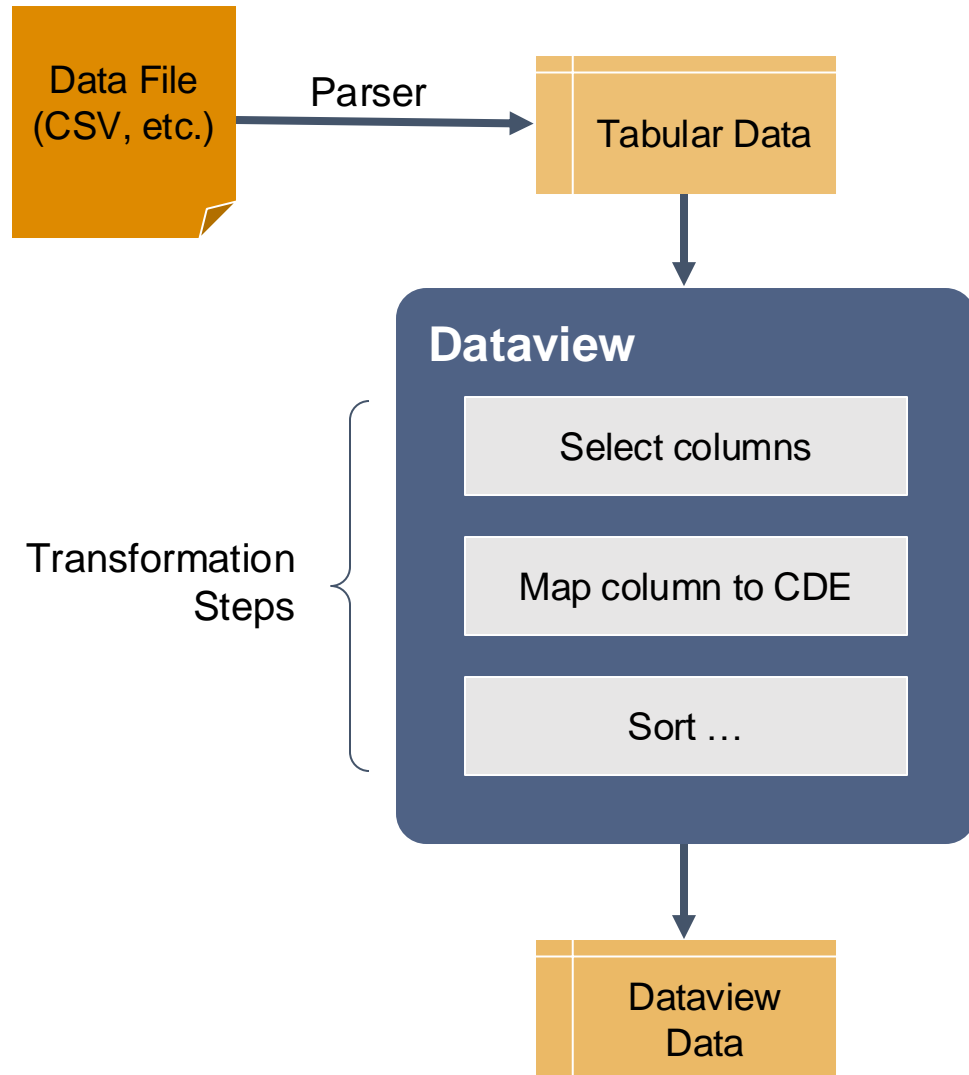


**Dataviews** take data from one or more sources, apply a series of transformation steps to that data (*filtering, sorting, mapping, etc.*) resulting in a new table of data as output.

## 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 Dataviews



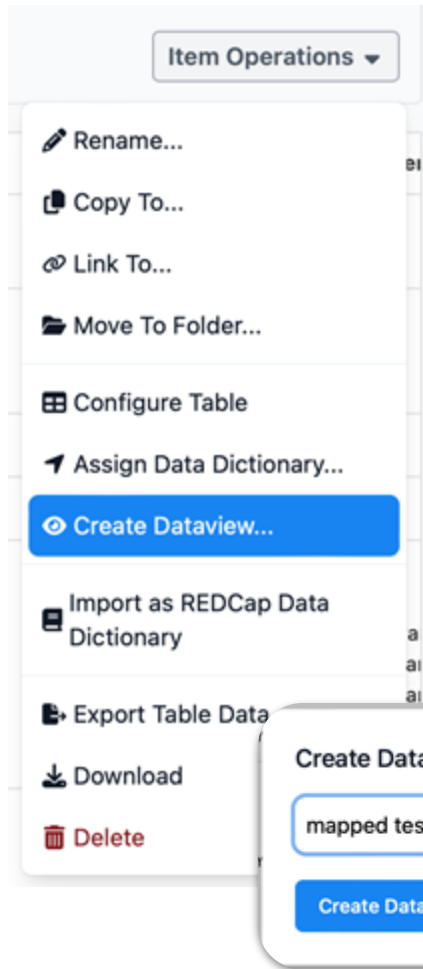
**Next, we'll walk through the process together:**

1. Create a new dataview item
2. Add mapping steps to the dataview that map your data values to align with the CDEs
3. Label the newly mapped data columns with the appropriate CDEs and check that it passes validation

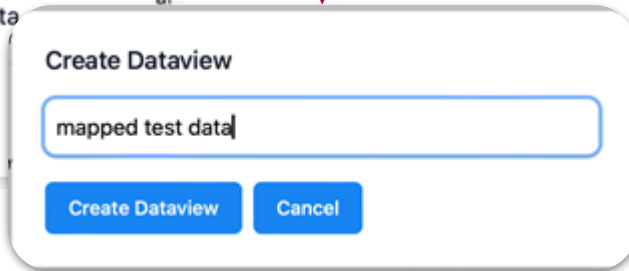
# Mapping CDEs via Dataviews

## Step 1: Create a new Dataview

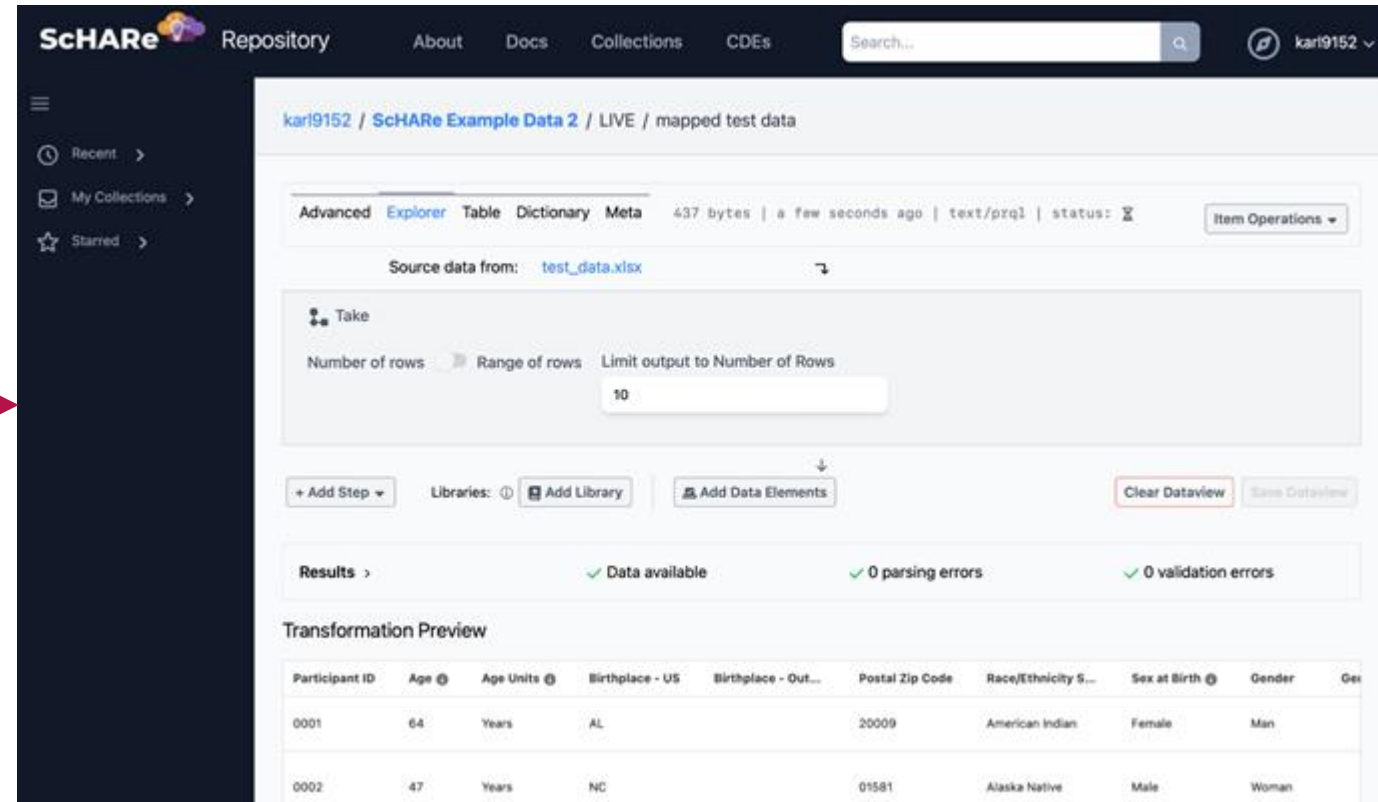
1. Select **Create Dataview...** from the item's menu
2. Give it a name
3. It will open the Explorer, where you can begin building



The screenshot shows the 'Item Operations' menu for a data item. The 'Create Dataview...' option is highlighted in blue. Other options include 'Rename...', 'Copy To...', 'Link To...', 'Move To Folder...', 'Configure Table', 'Assign Data Dictionary...', 'Import as REDCap Data Dictionary', 'Export Table Data', 'Download', and 'Delete'.



The 'Create Dataview' dialog box is shown with the name 'mapped test data' entered in the text field. There are 'Create Dataview' and 'Cancel' buttons at the bottom.



The screenshot shows the ScHARe Repository Explorer interface. The breadcrumb path is 'kari9152 / ScHARe Example Data 2 / LIVE / mapped test data'. The interface shows the 'Advanced' view with options for 'Table', 'Dictionary', and 'Meta'. The source data is 'test\_data.xlsx'. The 'Take' section shows 'Number of rows' set to 10. The 'Results' section shows 'Data available', '0 parsing errors', and '0 validation errors'. The 'Transformation Preview' table is shown below.

Participant ID	Age	Age Units	Birthplace - US	Birthplace - Out...	Postal Zip Code	Race/Ethnicity S...	Sex at Birth	Gender	Get
0001	64	Years	AL		20009	American Indian	Female	Man	
0002	47	Years	NC		01581	Alaska Native	Male	Woman	

# Mapping CDEs via Dataviews

## A Quick Tour of the Dataview Explorer screen

Source data comes in at the top

Add more steps or add a library of additional steps

The result is previewed at the bottom

The screenshot shows the SchARE Dataview Explorer interface. At the top, the source data is identified as 'test\_data.xlsx'. A 'Take' step is configured with 'Number of rows' set to 10. Below the step, there are buttons for '+ Add Step', 'Libraries', 'Add Library', and 'Add Data Elements'. The 'Results' section shows 'Data available', '0 parsing errors', and '0 validation errors'. A 'Transformation Preview' table is displayed at the bottom.

Participant ID	Age	Age Units	Birthplace - US	Birthplace - Out...	Postal Zip Code	Race/Ethnicity S...	Sex at Birth	Gender	Ge...
0001	64	Years	AL		20009	American Indian	Female	Man	
0002	47	Years	NC		01581	Alaska Native	Male	Woman	

Data flows vertically down the pipeline of transformation steps

Save your changes using the Save Dataview button

Once you're done, use the **Table** tab to view the full results

# Mapping CDEs via Dataviews

## Step 2: Add Mapping Steps to your Dataview

1. Using the **Add Step** menu, add a **Map Column** step

+ Add Step ▾

- Select Columns
- Filter Rows
- Sort
- Shuffle
- Join
- Take Rows
- Rename Column
- Map Column**
- Aggregate Rows

2.

Fill in the step fields (source column, target column, value map)

3.

Review the results, adjust as needed

4.

Repeat for all columns that need to be mapped

Map Into

Source Column: Race/Ethnicity Self-Identification

Value Map from Source to Target

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

New column  Replace source column

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


Race/Ethnicity S...	race_mapped
American Indian	American Indian or Alaska Native
Alaska Native	American Indian or Alaska Native
Black	Black or African American
African American	Black or African American



# Mapping CDEs via Dataviews

## Step 3: Assign CDEs to Data Columns

1. Click **Add Data Elements**
2. Select the column and data element, click **Add**
3. Verify the data element is on the column and the data validates successfully

 Add Data Elements



**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	<input checked="" type="checkbox"/>
Age Units	Age Units	N/A	<input checked="" type="checkbox"/>
Sex at Birth	Sex at Birth	N/A	<input checked="" type="checkbox"/>


Add new data element annotation:

From:


CDE set selected: ScHARe

OPTIONAL: Select a PV...




 0 validation errors




race\_mapped 

American Indian Alaska Native Race/Ethnicity Self-Identification (string)

Concepts 

- C17049: Race *dataElement*
- C16564: Ethnic Group *dataElement*
- C74528: Self-Report *dataElement*

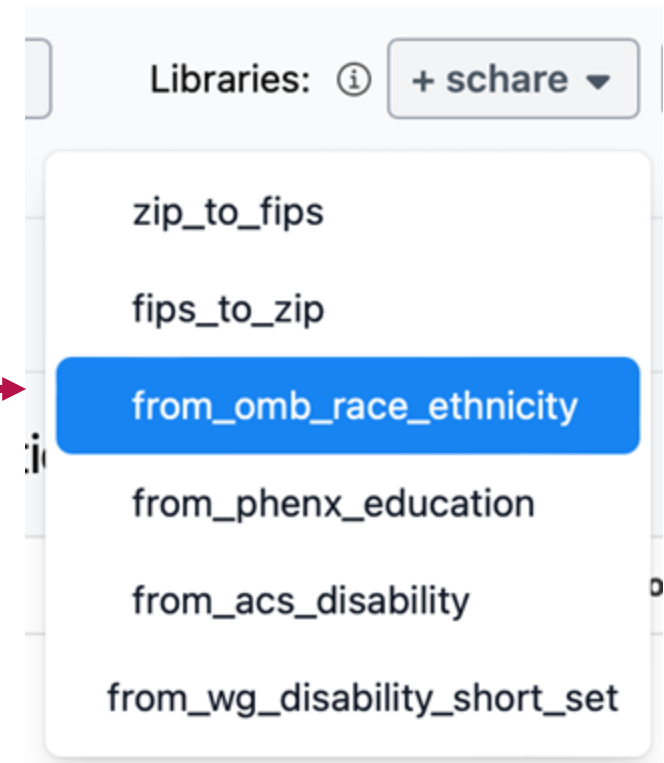
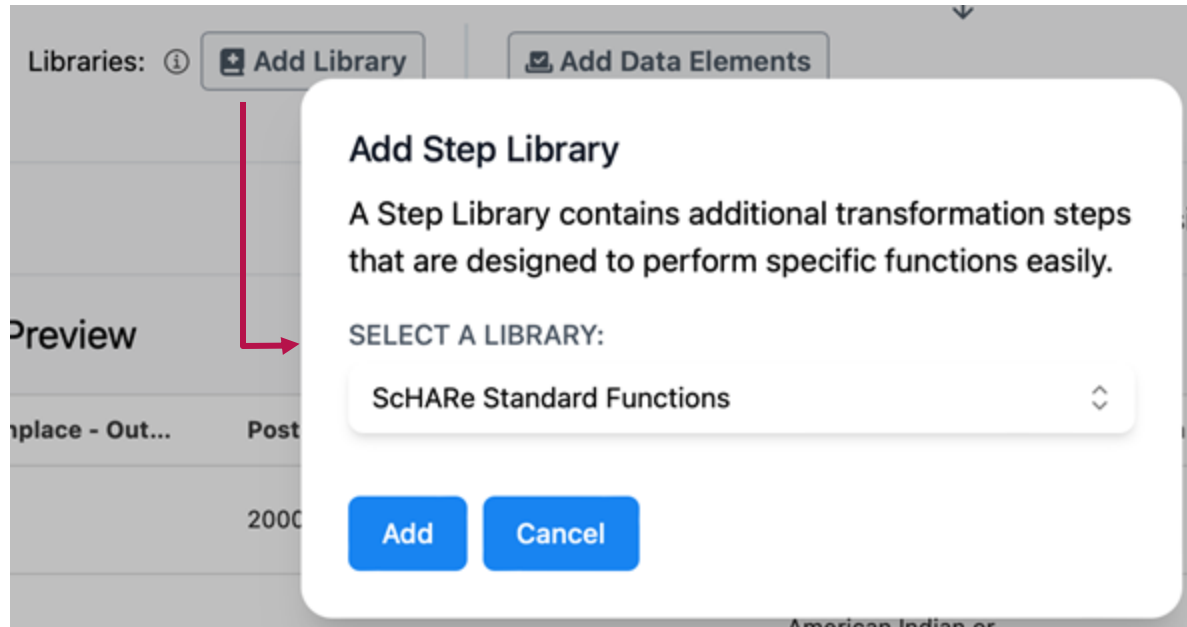
Definitions 

Black or African American

# Mapping CDEs via Dataviews

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!**



# You're done uploading your first data set!

What we've accomplished:

- Created a private Collection to hold all of your project data in one place
- Uploaded a data dictionary, documentation, linked to papers, etc.
- Uploaded your tabular data and viewed it
- Assigned and mapped your data to CDEs

## Questions so far?

# ScHARe

Sharing Data



BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION

# Viewing your Collection

**ScHARe** Repository    About    Docs    Collections    CDEs    Search...

karl9152 / **ScHARe Example Data 2** / LIVE

### ScHARe Example Data 2

**Abstract**  
For purposes of de... according to ...

**Levels of Influence**  
Individual  
Community

**Access Level**  
Private

**Analysis Readiness**  
Ready >

**CDE Compliance - ScHARe**  
2 / 19 CDEs assigned

**Tags**  
# Topics tagged in this collection  
Co-Morbidities    Disability

**Links and Documents**

**Data Items**

STATUS	NAME	CREATED	SIZE
	test_data.xlsx	an hour ago	14 KB

Page 1 of 1

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

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

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

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

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?

The screenshot shows a user interface for data analysis readiness and compliance. At the top, under the heading "Analysis Readiness", there is a green checkmark icon followed by the word "Ready" and a downward arrow. Below this, four status indicators are listed: "Metadata: Complete", "DEs Assigned: Partially Complete", "Data Access: Passing", and "Validations: Passing". A horizontal line separates this section from the one below. The second section is titled "CDE Compliance - ScHARe" with an information icon. It features a progress indicator showing a pie chart with 4 segments filled out of 19, labeled "4 / 19 CDEs assigned". Below the progress indicator, there are four buttons representing assigned CDEs: "Age", "Postal Zip Code", "Race/Ethnicity S...", and "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

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

### Share Collection

Users, groups, and collections with access:

ID	ROLE	
 Karl Gutwin (karl9152)	ADMIN	

Share with:

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**.

## 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

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

**Share Collection**

Users, groups, and collections with access:

ID	ROLE
Karl Gutwin (karl9152)	ADMIN

Share with: **Users** Groups Collections

John

Role: Reader

**Share**

**John Jacquay (john)**

**Done**

## Access Levels

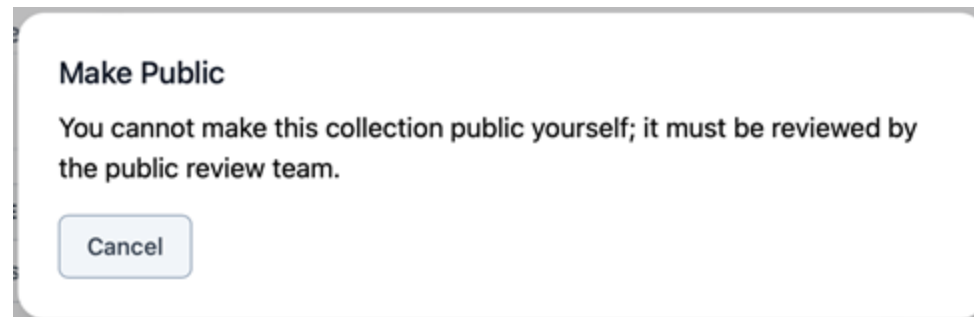
The access level of a collection defines the maximum permissions that can be used to share it with others. The following access levels are supported:

- **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

# 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

Data Aggregation  
and 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



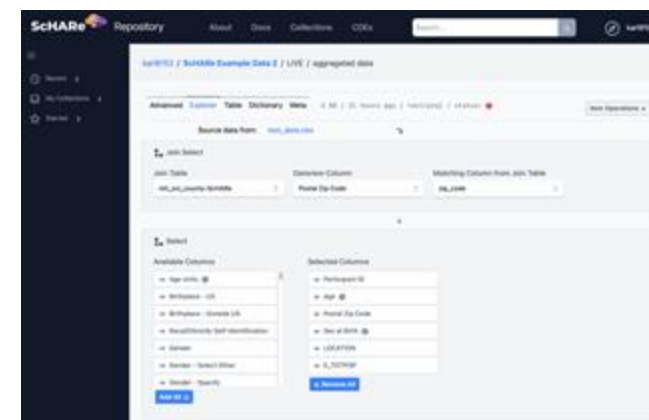
The screenshot shows a Jupyter notebook environment titled "WORKSPACES" with the URL "schare-gdr-demo.pytib". The notebook contains several code cells. The final cell displays a data table with columns for various identifiers and numerical values.

CHIRID	EST	STARTDATE	ENDDATE	PUBLICITY	PERCENTAGE	SCORE	ALIGNMENT	STATE	E_TOTOP	E_M	E_P	E_BOTTOM	E_PD	E_BOTTOM
1	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0
2	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0
3	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0
4	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0
5	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0
6	20100000	AL	1	1.0	0.0	1.0	1.0	10	ALABAMA	466680.0	562260.0	657880.0	747860.0	837840.0

## 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)

The screenshot shows a Terra workspace interface with a green header. The workspace name is 'schare-gde-demo.ipynb'. The notebook content is as follows:

```
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')

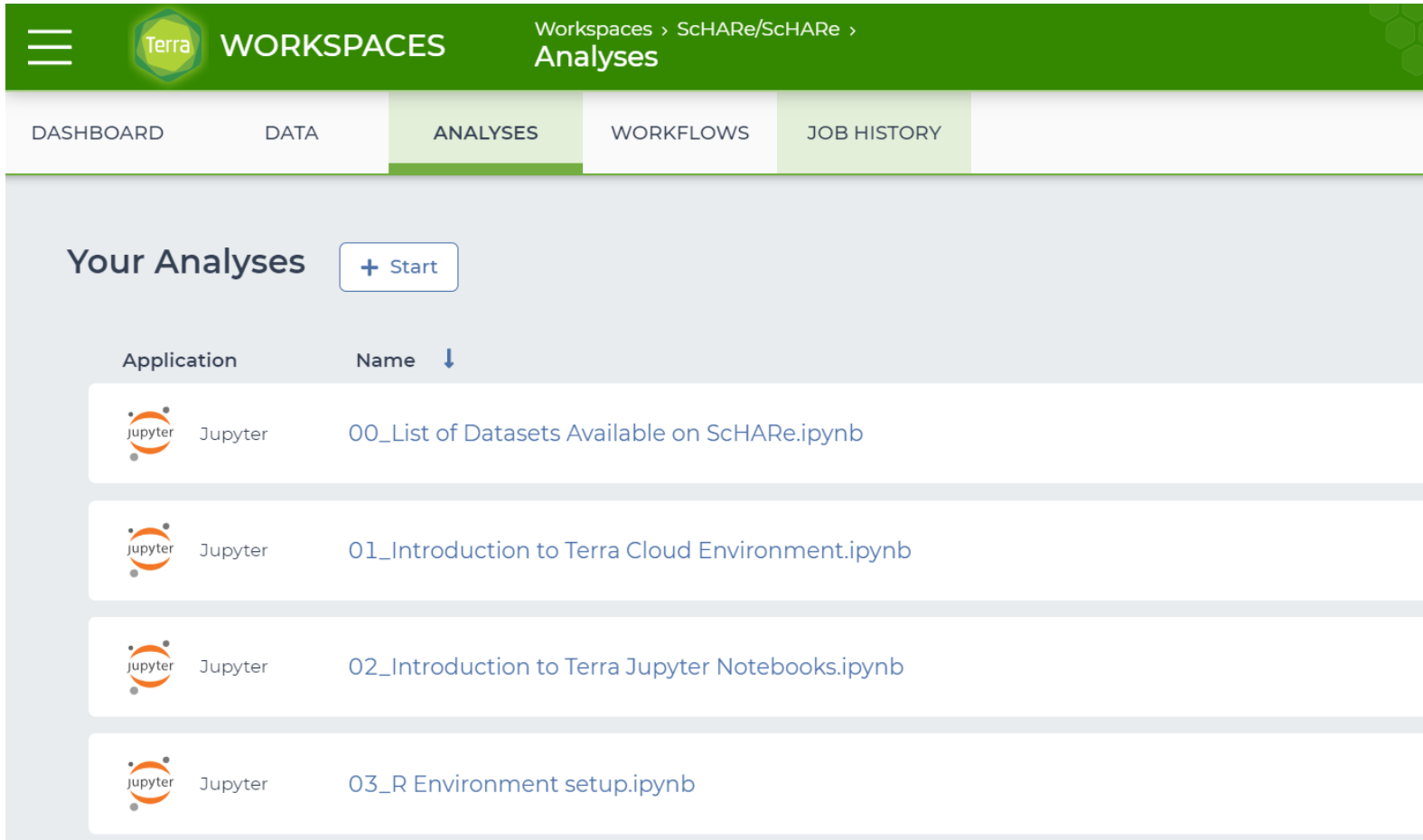
In [4]: nmhss = collection.get_table('nmhss-cbt-facilities')
Loading nmhss-cbt-facilities: 0it [00:00, ?it/s]

In [5]: nmhss
```

The output of the notebook is a table with 17 columns: CASEID, LST, MHINTAKE, OWNERSHP, PUBLICAGENCY, TREATCOGTHRPY, SENIORS, ALZHDementia, STATE, E\_TOTPOP, E\_HH, E\_POV, E\_UNEMP, E\_PCI, and E\_NOHSDP. The table contains 17 rows of data, with some rows truncated with '...'.





	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

# Analyzing Data on Terra (Jupyter Notebooks)

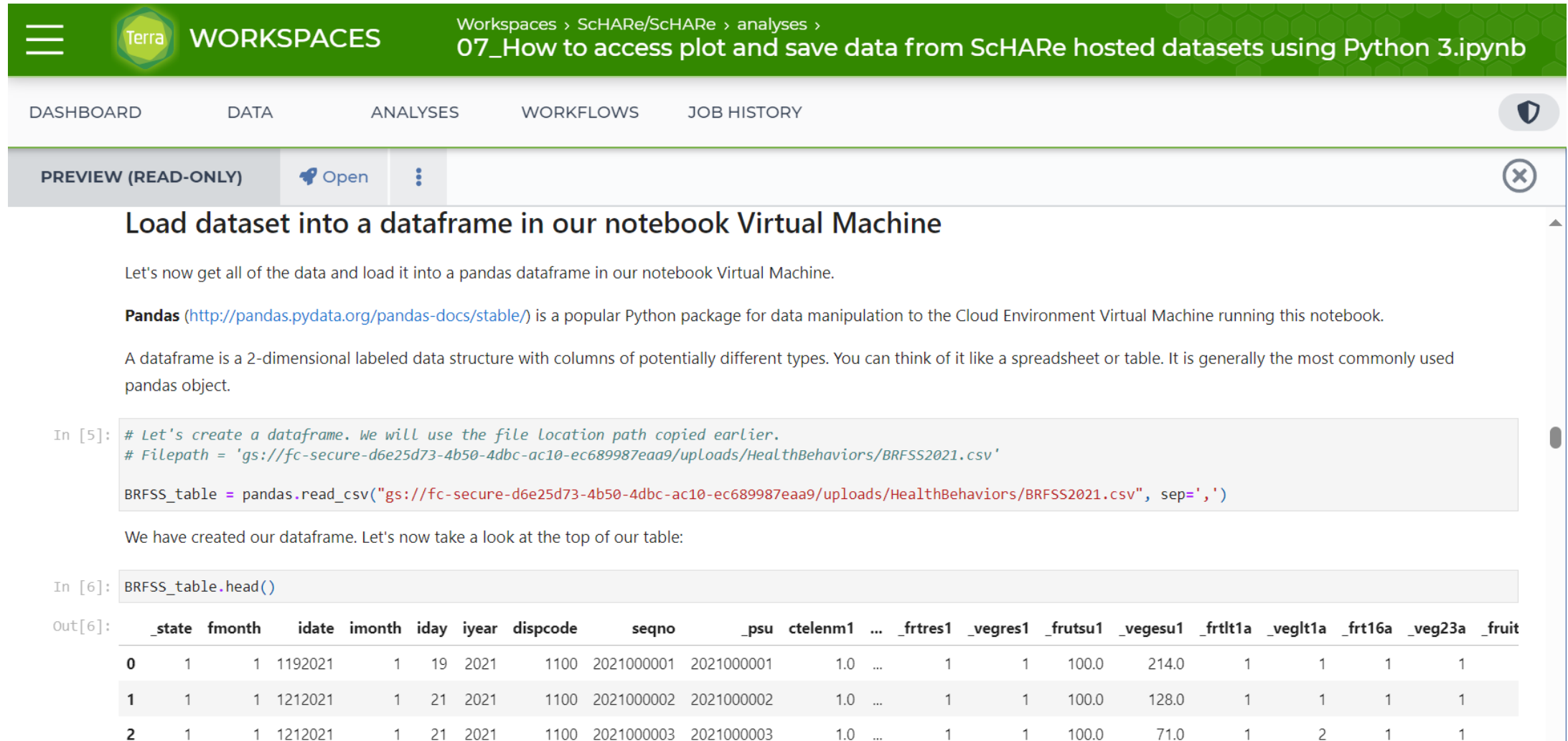


The screenshot displays the Terra WORKSPACES interface. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES", and a breadcrumb trail: "Workspaces > ScHARe/ScHARe > Analyses". Below this is a horizontal menu with tabs for "DASHBOARD", "DATA", "ANALYSES" (which is highlighted), "WORKFLOWS", and "JOB HISTORY".

The main content area is titled "Your Analyses" and includes a "+ Start" button. Below this is a table listing the analyses:

Application	Name ↓
 Jupyter	00_List of Datasets Available on ScHARe.ipynb
 Jupyter	01_Introduction to Terra Cloud Environment.ipynb
 Jupyter	02_Introduction to Terra Jupyter Notebooks.ipynb
 Jupyter	03_R Environment setup.ipynb

# Analyzing Data on Terra (Jupyter Notebooks)



**Terra 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	1	1212021	1	21	2021	1100	2021000002	2021000002	1.0	...	1	1	100.0	128.0	1	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	1



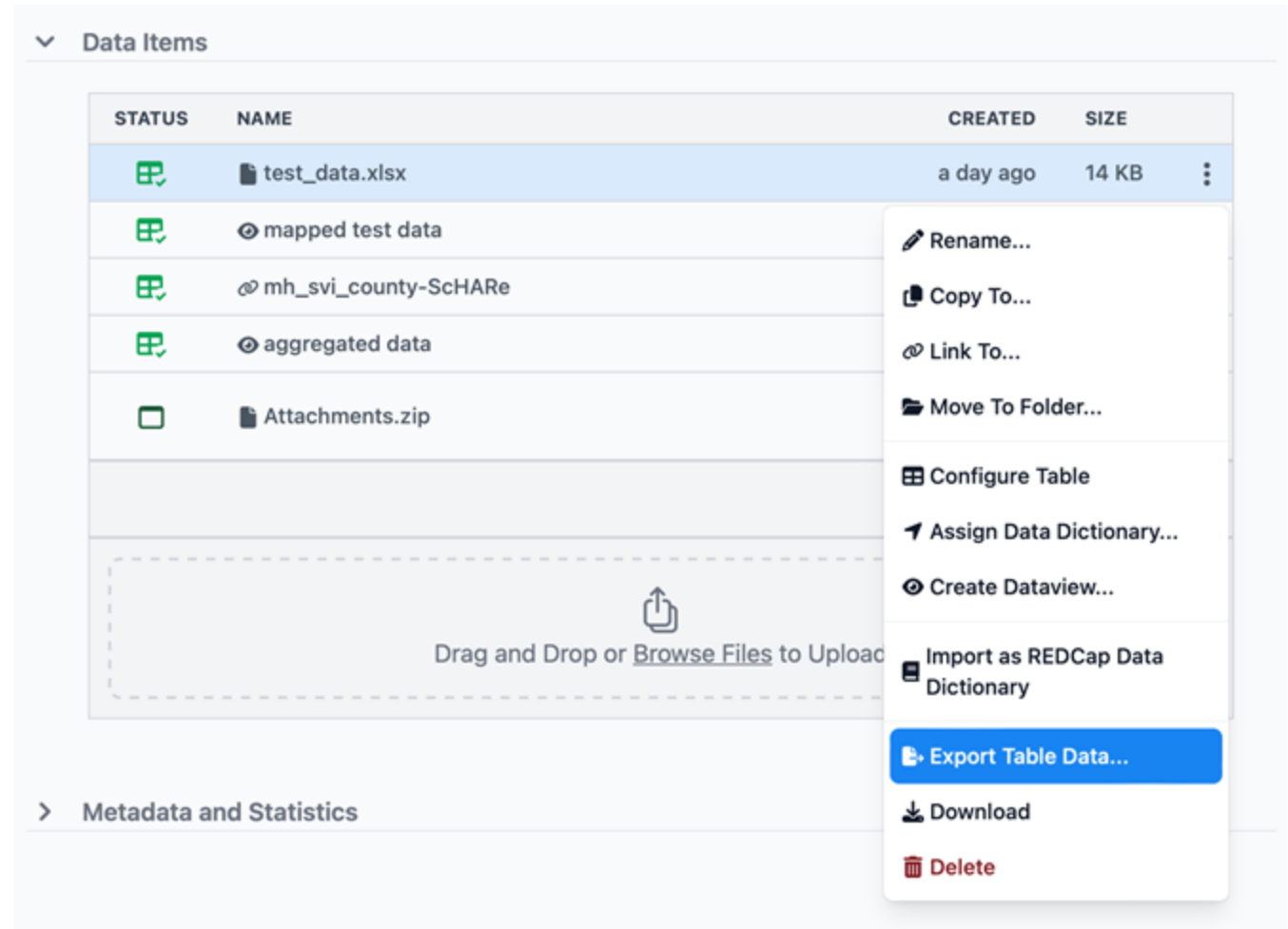
# Downloading data to your own computer

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 shows a 'Data Items' section with a table listing various data files. A context menu is open over the 'test\_data.xlsx' item, displaying several actions. The 'Export Table Data...' option is highlighted in blue.

STATUS	NAME	CREATED	SIZE
	test_data.xlsx	a day ago	14 KB
	mapped test data		
	mh_svi_county-SCHARe		
	aggregated data		
	Attachments.zip		

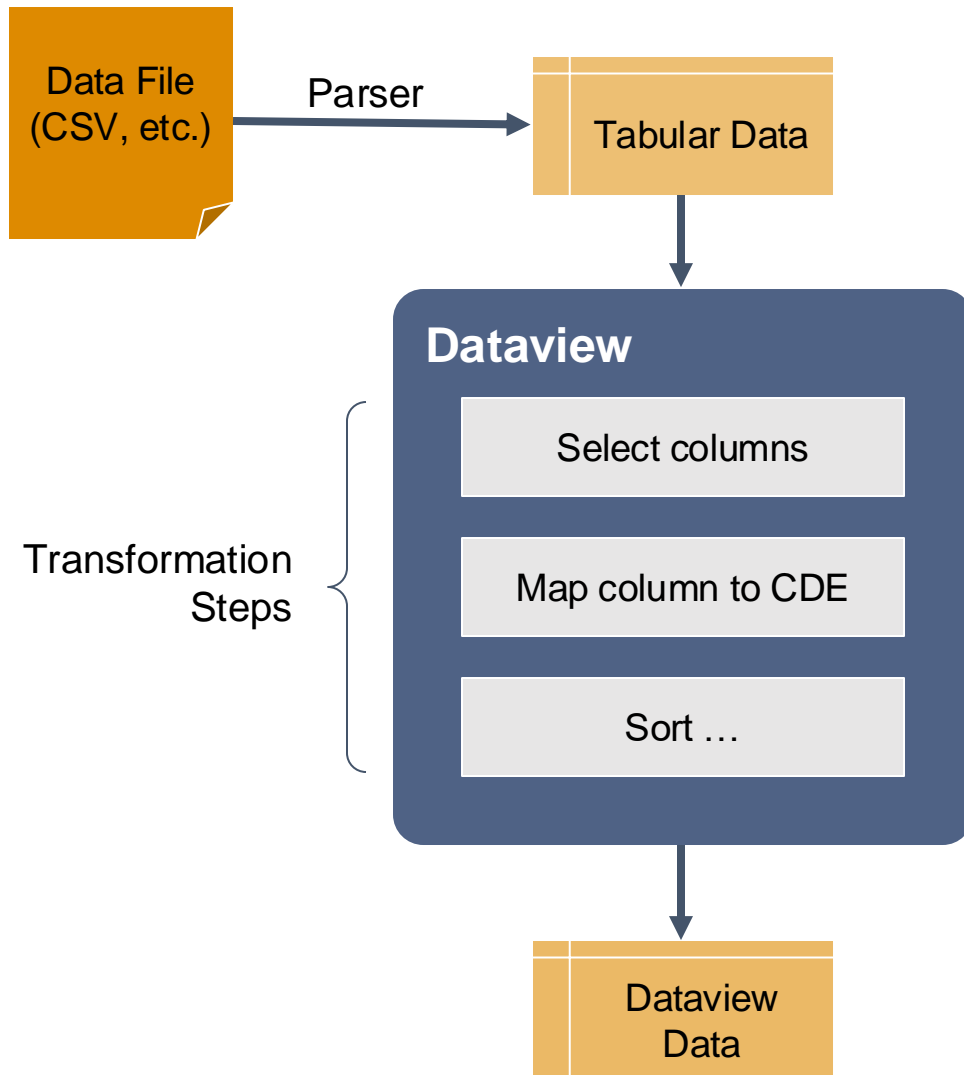
Drag and Drop or [Browse Files](#) to Upload

Metadata and Statistics

- 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

# 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

The screenshot shows the SchARE Repository interface. At the top, there is a navigation bar with 'About', 'Docs', 'Collections', and 'CDEs' links, a search bar, and a user profile 'karl9152'. The main content area is titled 'Collections' and includes a 'Create New Collection' button. It is divided into three sections: 'My Collections' with one item 'SchARE Example Data 2' (LIVE, 4 hours ago); 'Shared with Me' with a message 'You don't have any collections shared with you right now!'; and 'Public Collections grouped by: Domains of Influence'. The domains shown are 'Biological', 'Health Care Systems and Clinical Care', and 'Sociocultural Environment'. Under 'Health Care Systems and Clinical Care', there are two items: 'Example NMHSS Analysis' (LIVE, 5 months ago) and 'Minority Health SVI' (LIVE, 7 months ago).

The screenshot shows the search results page for 'nmhss'. The search bar at the top contains 'nmhss'. The page is titled 'Search' and features a 'Filters' sidebar on the left with categories: 'metadata' (0 / 90), 'content type' (0 / 51), 'type' (0 / 6), and 'tags' (0 / 7). The main content area displays the top result, 'Example NMHSS Analysis', which is 5 months old. The description states: 'This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.' Below the description are three tags: 'nmhss-puf-2018-csv...', 'nmhss-cbt-facilities', and 'nmhss-dictionary-pa...'. An 'Apply Filters' button is located at the bottom of the filters sidebar.

*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

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

The screenshot displays the ScHARE Repository interface. At the top, the navigation bar includes 'ScHARE Repository', 'About', 'Docs', 'Collections', 'CDEs', a search bar, and a user profile 'kari9152'. The main content area shows a breadcrumb path: 'kari9152 / ScHARE Example Data 2 / LIVE / aggregated data'. Below this, there are tabs for 'Advanced', 'Explorer', 'Table', 'Dictionary', and 'Meta'. The 'Advanced' tab is active, showing a 'Source data from: test\_data.xlsx' and an 'Item Operations' dropdown. The 'Join Select' section is configured with 'Join Table' set to 'mh\_svl\_county-ScHARE', 'Dataview Column' set to 'Postal Zip Code', and 'Matching Column from Join Table' set to 'zip\_code'. Below this is the 'Select' section, which has two columns: 'Available Columns' and 'Selected Columns'. The 'Available Columns' list includes 'Age Units', 'Birthplace - US', 'Birthplace - Outside US', 'Race/Ethnicity Self-Identification', 'Gender', 'Gender - Select Other', and 'Gender - Specify'. The 'Selected Columns' list includes 'Participant ID', 'Age', 'Postal Zip Code', 'Sex at Birth', 'LOCATION', and 'E\_TOTPOP'. At the bottom of the 'Select' section, there are buttons for 'Add All' and 'Remove All'. Below the 'Select' section, there are buttons for '+ Add Step', 'Libraries: Add Library', and 'Add Data Elements'. On the right side of this row, there are buttons for 'Clear Dataview' and 'Save Dataview'. The 'Results' section shows 'Data available', '0 parsing errors', and '0 validation errors'. The 'Transformation Preview' section shows a table with the following data:

Participant ID	Age	Postal Zip Code	Sex at Birth	LOCATION	E_TOTPOP
0001	64	20009	Female	District of Columbia, District of Columbia	684498
0002	47	01581	Male	Worcester County, Massachusetts	822280

# SCHARe

Conclusion

BE A PART OF THE FUTURE  
OF KNOWLEDGE GENERATION



# ScHARe Repository

## In summary:

- 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](mailto:schare@mail.nih.gov)

*Response time: within 24 hours*

**For Technical Support inquiries, contact:**

 [schare-repository-support@bioteam.net](mailto:schare-repository-support@bioteam.net)

 Office Hours: Wednesdays, 4-5 PM US Eastern

## Questions?

# 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](https://bit.ly/think-a-thons)

Register for ScHARe:



[bit.ly/join-schare](https://bit.ly/join-schare)

 [schare@mail.nih.gov](mailto:schare@mail.nih.gov)

