



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

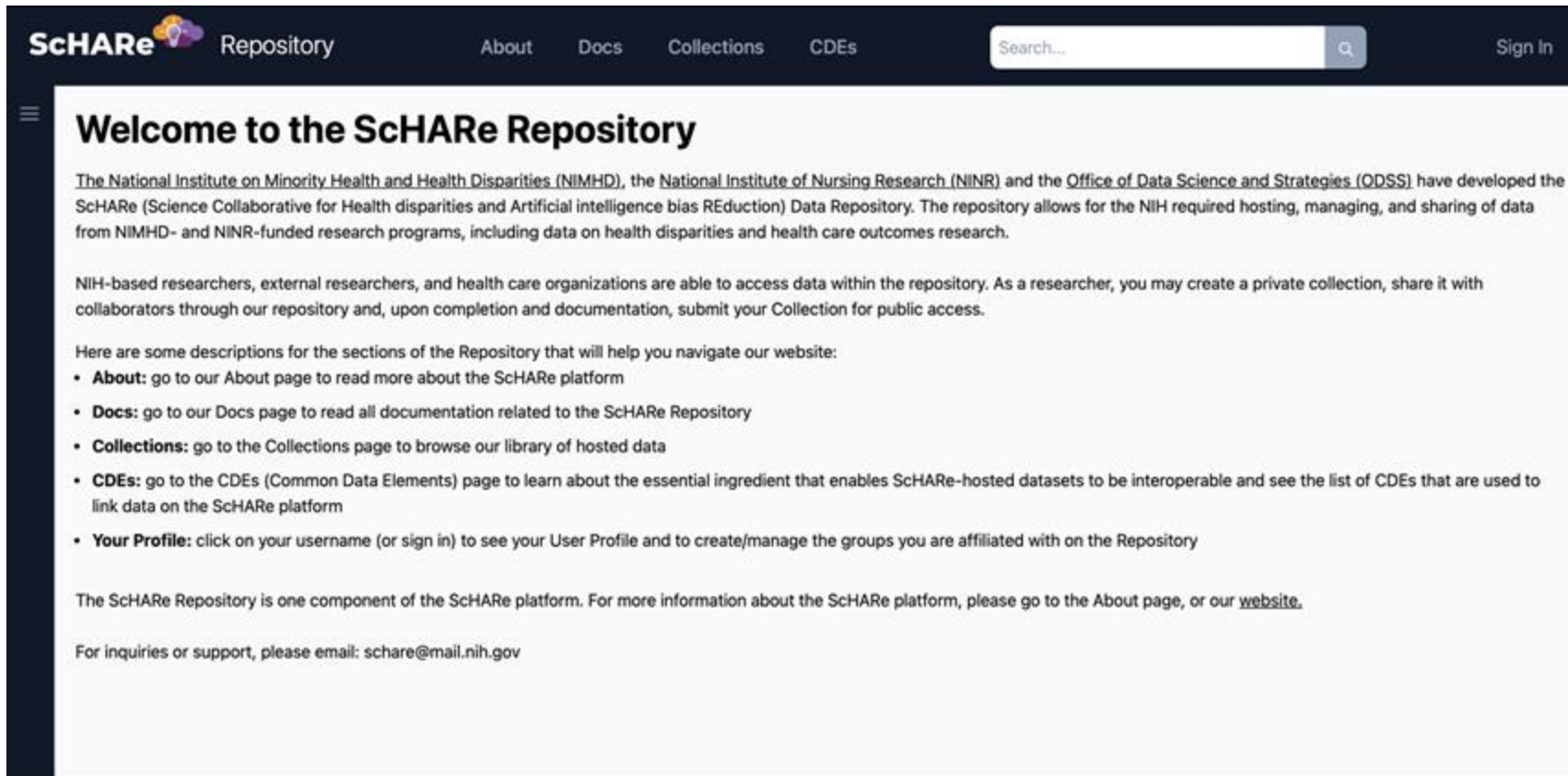
Getting Started

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Visit the ScHARe Repository!

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



The screenshot shows the homepage of the ScHARe Repository. The header includes the ScHARe logo, the word "Repository", and navigation links for "About", "Docs", "Collections", and "CDEs". A search bar and a "Sign In" link are also present. 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, a paragraph explains that NIH-based researchers, external researchers, and health care organizations can access data within the repository. 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 pages. At the bottom, there is a paragraph about the ScHARe platform and a contact email address.

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



Visit the ScHARe Repository!

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

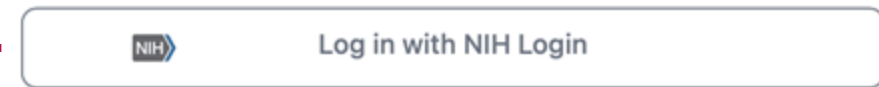
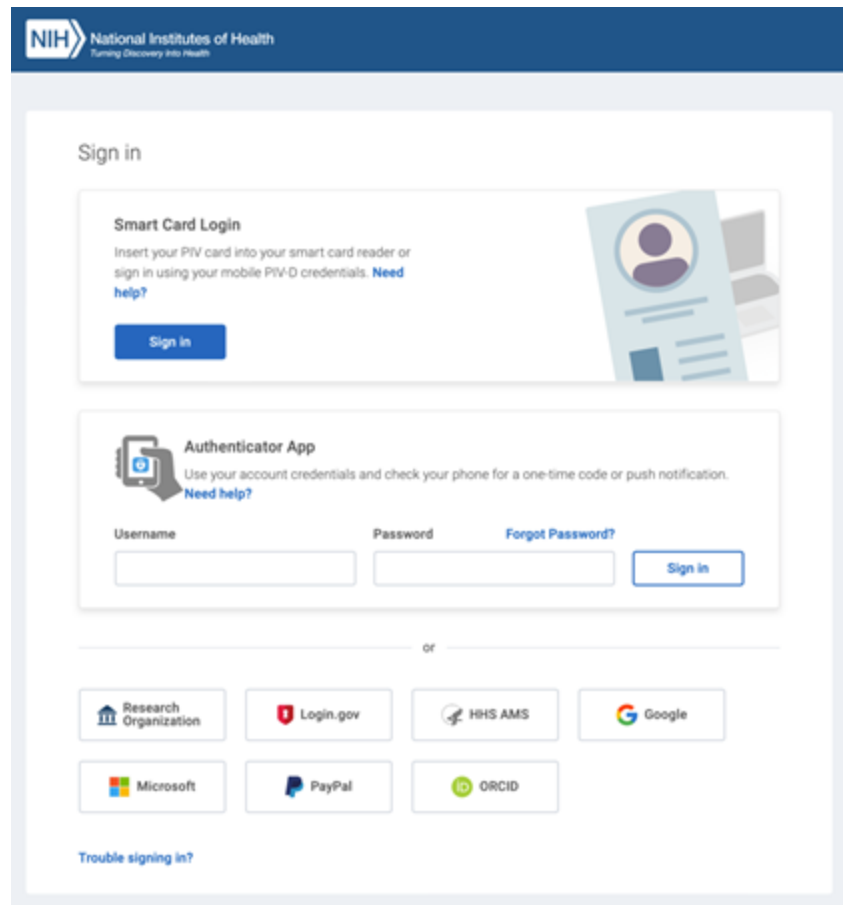
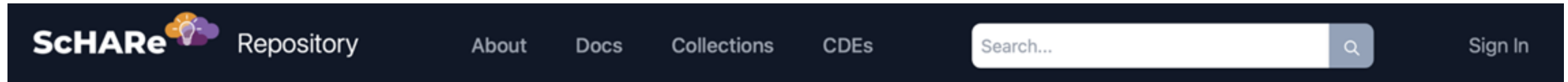
For today's training, please use our test site:

<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 is a main content area. On the left side of this area 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.

The main content area has a white background. At the top of this area is the heading 'Welcome to the SchARE Repository'. Below the heading is a paragraph of text: '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.'

Below this paragraph is another paragraph: '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.'

Below that is a paragraph: 'Here are some descriptions for the sections of the Repository that will help you navigate our website:'

Below this paragraph is a bulleted list of five items:

- **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

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

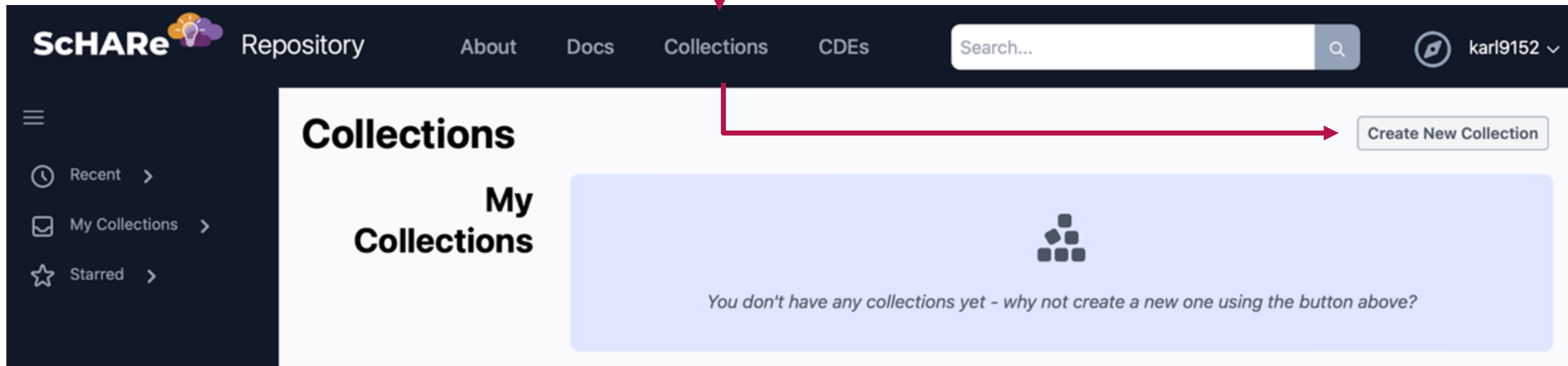
Uploading your
first Data Set

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



Creating a Collection

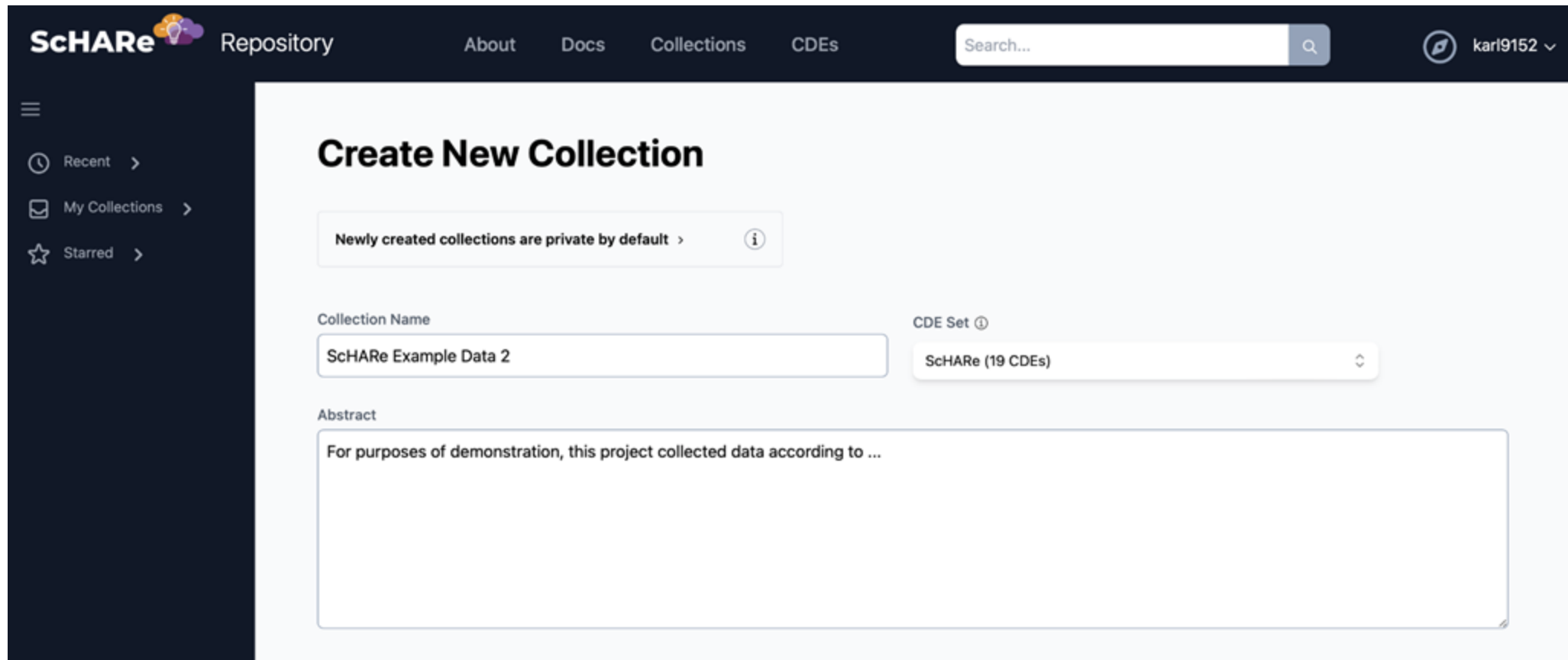
- **Collections** are a place where you can describe and store your data and any related metadata. They can be shared with colleagues and published when you're ready. You can upload any file type!
- The first step in uploading data is to create a collection



The screenshot shows the SchARE Repository interface. The top navigation bar includes the SchARE logo, 'Repository', 'About', 'Docs', 'Collections', and 'CDEs'. A search bar is on the right, and the user 'karl9152' is logged in. The left sidebar shows 'Recent', 'My Collections', and 'Starred' options. 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 to the 'Create New Collection' button.

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, there is a sidebar with a menu icon and links for 'Recent', 'My Collections', and 'Starred'. The main content area is titled 'Create New Collection' and contains a notification box stating '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' with the text 'For purposes of demonstration, this project collected data according to ...'.

ScHARe Repository

About Docs Collections CDEs

Search...

karl9152

Create New Collection

Newly created collections are private by default

Collection Name

ScHARe Example Data 2

CDE Set

ScHARe (19 CDEs)

Abstract

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

The screenshot displays the SchARE Repository interface for a collection named 'SCHARE Example Data 2'. The top navigation bar includes 'About', 'Docs', 'Collections', and 'CDEs', along with a search bar and a user profile 'karl9152'. The left sidebar shows 'Recent', 'My Collections', and 'Starred' items. The main content area features the collection title 'SCHARE Example Data 2' and an abstract. Below the abstract are sections for 'Links and Documents', 'Data Items', and 'Metadata and Statistics'. On the right, a sidebar provides key information: 'Access Level' is set to 'Private', 'Analysis Readiness' is 'Not Ready', and 'CDE Compliance - SchARE' shows '0 / 19 CDEs assigned'. There are also 'Tags' for 'Co-Morbidities' and 'Disability'. An 'Operations' dropdown menu is visible in the top right of the collection page.

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.)

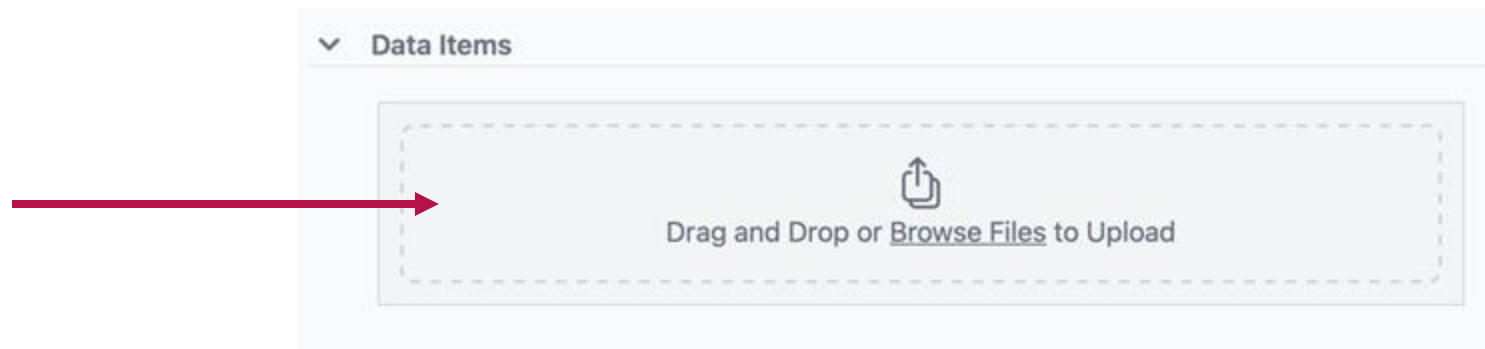
Some file formats can be previewed in the web interface

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 item 'example-data.csv' is highlighted. A context menu is open over the item, listing various actions. A red arrow points to the 'Download' button at the bottom of the 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 shows the ScHARe Repository interface. At the top, there is a navigation bar with 'ScHARe Repository', 'About', 'Docs', 'Collections', 'CDEs', and a search bar. The main content area displays the item 'karl9152 / ScHARe Example Data 2 / LIVE'. The item details include an abstract, levels of influence (Individual, Community), and domains of influence (Biological, Behavioral). A table of data items is shown below, with one item 'test_data.xlsx' listed. The right sidebar contains metadata such as 'Access Level: Private', 'Analysis Readiness: Ready', and 'CDE Compliance - ScHARe: 2 / 19 CDEs assigned'. Callouts provide additional context: one points to the item name, another to the status icon in the table, and a third to the 'Private' access level.

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 breadcrumb path is: karl9152 / ScHARe Example Data 2 / LIVE / test_data.xlsx. Below the breadcrumb, there are tabs for File, Table, Dictionary, and Meta. The 'Table' tab is active, showing a 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
Definitions
What is the person's age (in years if more than 24 months old or months if 24 months or younger)?
https://cde.nlm.nih.gov/api/de/PDjBiGXjO
The number of years (if more than 24 months old) or months (if 24 months or younger) that the person has been alive.
National Longitudinal Survey
https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/household/age
age, person, demographics