## Schare Schare Repository Introduction

November 20, 2024

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



# SCHARE

Creating a Collection

> BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

- **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



ScHARe	Repository	About	Docs	Collections	CDEs	Search	Q	scharedemo 🗸
≡								
🔇 Recent 🔉	Create Ne	w Coll	lectio	n				
My Collections 🗲								
☆ Starred >	Newly created collection	ons are private by		<u>(1</u> )				
	Collection Name					CDE Set (1)		
						ScHARe (19 CDEs)	\$	
	Abstract							

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

Domains of Influence

¢

Repository	About Docs	Collections CDEs	Search	٩	🧭 scharedemo 🗸
Create No	ew Collection	า			
ns 🗲					
Newly created collect	ions are private by default >	( <b>i</b> )			
Collection Name			CDE Set (1)		
demo-files			ScHARe (19 CDEs)	\$	
Abstract					
November 20 Think-	a-Thon Demonstration Files				
Tags			Project CDEs ()		
Enter a Tag			Levels of Influence	Select all that apply	•
			Domains of Influence	Select all that apply	
			Research Areas	Select all that apply	

ScHARe <sup>w</sup> Re	pository	About	Docs	Collections	CDEs	Search	Q	ø scharedemo v
≡ ○ Recent >	Common Da	ta Elei	ment	5				
<ul> <li>My Collections →</li> <li>Starred →</li> </ul>	A major obstacle in advancing internal reference system asso linking of datasets and a new o The ScHARe Repository leverag	health dispari ciated with co generation of l ges the followi	ities research mmon categ health dispar ing CDE sets:	is linking data bet ories in health disp ities research.	ween different studies. To surm arities data. This system is know	ount this, the ScHARe Repository m vn as Common Data Elements (CDE	nakes use of a set of shared langu Es), and the use of CDEs across th	age, definitions, and an e Repository enables the
	ScHARe 19 CDEs ScHARe core common de Age	ata elements						
	Age >						number 5 definitions 1 conce	pt PDjBiGXjO:0001
	Age Units >					string	5 definitions   1 concept   valueSet	GNVoEdFQR0:0001
	Birthplace							
	Birthplace - US >						string 5 definitions 3 concept	ts 0qP44x2oX:0001
	Birthplace - Country	Outside US →	,				string 4 definitions 2 concept	

## Tags

My Collections

☆ Starred >

Collection Name	Adolescents	
demo-files	Age	\$
Abstract	Alcohol Use	
November 20 Think-a-Thon Demonstration Files	American Indian or Alaska Native	
	Asian or Asian American	
	Black or African American	
	Children	
Tags	Chronic Diseases	
Enter a Tag	Clinical Trial Data	Select all that apply
[	Community Health Programs	Select all that apply.
	County-Level Data	verse on one oppyin
		Select all that apply 🗘
Apply relevant	tags and project-level CDEs to yo	bur dataset
	Data Science: AI/ML/LLM	
	Data from Electronic Health Records (EHRs)	
Advanced User-Defined Metadata >	Diet and Nutrition	
	Disability	

٠

Submit Cancel

Discrimination and Racism

## Tags

My Collections 🔉

숤 Starred ゝ

demo-files	ScHARe (19 CDEs)	\$	
Abstract			
November 20 Think-a-Thon Demonstration Files			
Tags	Project CDEs ()		
Cross-Sectional Study × State-Level Data ×	Levels of Influence	Select all that apply	\$
Enter a Tag	Domains of Influence	Select all that apply	0
	Descent Arres		
	Kesearch Areas	Select all that apply	Ŷ
	Research Focus	Select all that apply	\$





۰

#### **Project-Level CDEs: Levels of Influence**

Abstract			
November 20 Think-a-Thon Demonstration Files			
Tags	Project CDEs (1)		
Tags Cross-Sectional Study × State-Level Data ×	Project CDEs () Levels of Influence	Select all that apply	
Tags Cross-Sectional Study × State-Level Data × Enter a Tag	Project CDEs ① Levels of Influence Domains of Influence	Select all that apply	
Tags Cross-Sectional Study × State-Level Data × Enter a Tag	Project CDEs () Levels of Influence Domains of Influence	Select all that apply Individual Interpersonal	
Tags         Cross-Sectional Study ×       State-Level Data ×         Enter a Tag	Project CDEs Levels of Influence Domains of Influence Research Areas	Select all that apply Individual Interpersonal Community	
Tags         Cross-Sectional Study ×       State-Level Data ×         Enter a Tag	Project CDEs (1) Levels of Influence Domains of Influence Research Areas	Select all that apply Individual Interpersonal Community Societal	

#### **Project-Level CDEs: Domains of Influence**

Abstract		
November 20 Think-a-Thon Demonstration Files		
Tags	Project CDEs (1)	
Cross-Sectional Study × State-Level Data ×	Levels of Influence	Individual
Enter a Tag	Domains of Influence	
	Domains of influence	Sociocultural Environment
	Research Areas	Biological
		Behavioral
	Research Focus	Physical/Built Environments
		<ul> <li>Sociocultural Environment</li> </ul>
		Health Care Systems and Clinical Care
Advanced User-Defined Metadata >		
Submit Cancel		

#### **Project-Level CDEs- Research Areas**

Project CDEs @	
Levels of Influence	
	Individual
Domains of Influence	Sociocultural Environment
Research Areas	Select all that apply
Research Focus	Minority health study focused on a
	one race or ethnic population and not addressing a health disparity.
	Health Disparity Outcomes
	Other Health Outcomes / Healthcare
	Delivery
	Project CDEs () Levels of Influence Domains of Influence Research Areas Research Focus

#### **Project-Level CDEs: Research Focus**

Abstract		
November 20 Think-a-Thon Demonstration Files		
Tags	Project CDEs (1)	
Cross-Sectional Study × State-Level Data ×	Levels of Influence	Individual
Enter a Tag	Domains of Influence	Sociocultural Equiparment
		Sociocultural Environment
	Research Areas	Select all that apply
	Research Focus	
		Select all that apply
		Higher incidence and/or prevalence of disease, including
		earlier onset or more aggressive
Advanced User-Defined Metadata >		Premature or excessive mortality
		from specific health conditions
Submit Cancel		Greater global burden of disease,

#### **The Collection Main Page**



#### **The Collection Main Page**



## Slido Poll

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

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

# SCHARE

## Uploading your first Data Set

BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

Upload any document, presentation, etc.

Link to papers through PubMed scharedemo / demo-files / LIVE 5.7 Operations -Ø Access Level ① demo-files A Private Abstract November 20 Think-a-Thon Demon Add Document or Link Levels of Influence Do **Analysis Readiness** Upload Pubmed Search Link Individual Soc 🛽 Not Ready 🔰 TYPE Select one... 0 Links and Documents  $\sim$ UPLOAD **CDE Compliance -**Choose File No file chosen ScHARe ① TITLE 🗙 0 / 19 CDEs assigned Title... Add New DESCRIPTION Tags > Data Items # Topics tagged in this collection > Metadata and Statistics Cross-Sectional St Add Cancel State-Level Da

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...

HARe <sup>®</sup> Reposito	ry About	Docs Collections	S CDEs	Search	م scharedemo ب
scharedemo / demo-files	/ LIVE				
demo-files Abstract November 20 Thin	-a-Thon Demonstration F	iles			Access Level ① Private
Levels of Influence Individual	Domains of In Sociocultural I	nfluence Environment			Analysis Readiness Not Ready
<ul> <li>Links and Docum</li> <li>Data Items</li> </ul>	ents	<b>↑</b>			CDE Compliance - ScHARe ① ② 0 / 19 CDEs assigned
> Metadata and St	Drag and	Drop or <u>Browse Files</u> to Uplo	bad		Tags # Topics tagged in this collection
, metaata ana se					State-Level Data



Recer

🛛 My

숬 Starr

UEIIU-IIES		Access Level (1)
Abstract		Private
November 20 Think-a-Th	on Demonstration Files	
Levels of Influence Individual	Domains of Influence Sociocultural Environment	Analysis Readiness
		🛽 Not Ready 🔉
		Tags
	Drag and Drop or <u>Browse Files</u> to Upload	<b>#</b> Topics tagged in this collection
	Upload Files	Cross-Sectional Study State-Level Da

scharedemo / demo-files / LIVE Operations -0 Э 53 Recent > My Collections > Starred > Access Level ① demo-files A Private Abstract File Upload Complete November 20 Think-a-Thon Demonstrati The ScHARe Repository uses the ScHARe common Levels of Influence Don data elements to enable data discovery, linking, and Analysis Readiness Individual Soci interoperability. To make your data more Not Ready > interoperable and ready for publication, we will guide you through the process of assigning common data Links and Documents elements to your data column headings. Click OK to > continue. CDE Compliance - ScHARe (1) Data Items V. X 0 / 19 CDEs assigned TIP: You can cancel now and complete the process later using the "Configure CDEs" tool from the collection Operations STATUS NAME menu. SIZE ⊞ example data.csv 25 KB Tags Cancel OK # Topics tagged in this collection Page 1 of 1 Cross-Sectional Study State-Level Data Ü Drag and Drop or Browse Files to Upload

# SCHARE

## CDE Mapping: Option 1

BE A PART OF THE FUTURE OF KNOWLEDGE GENERATION

#### **Getting Started with Mapping CDEs**

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

#### Assign CDEs at Upload

ScHARe Repository	About Docs Collections	CDEs Search		Q (art9152 ~
karl91!	Assign CDEs to Uploaded Data Columns			@ 5 5 Operations +
S Recent >	Assign your uploaded item's data columns to their appropr	iate common data elements.		
My Collections >	More Info >		(i)	
숫 Starred >	CDE set selected: ScHARe			vel D
	SCHARE CDE	UPLOADED DATA COLUMNS	DATA VALID	
	Age			teadiness
	123 Age	Age → \$	~	ady 🔉
	💭 Age Units	;		
	Birthplace			pliance - ScHARe @
	Birthplace - US		2	CDEs assigned
	Beinplace - County Outside 05		,	
	## Race/Ethnicity Self-Identification		1	
	≌≓ Sex at Birth	:	1	a tagged in this collection
	Gender			
	₩ Gender	:	1	
	Assign More Options Cancel			
			_	

#### Map CDEs using Dataviews

ARe <sup>®</sup> Re	epository	About Docs	Collections	CDEs Search		Q	Ø karl915
	karl9152 / ScHARe E	xample Data 2 / I	LIVE / mapped test	data			
	Advanced Explorer	Table Dictionary	Meta 1 KB   a d	ay ago   text/prql   status: 🗜		10.	om Operations -
						10	
	Sou	urce data from: te	st_data.xlsx	7			
	🖏 Map Into						Û
	Source Column		Value Map from Sou	irce to Target	Target Colu	ımn	
	Race/Ethnicity Self-Id	dentification 0	SOURCE VALUE	TARGET VALUE	race_map	bed	0
			Black	Black or African American	New colum	n 🔵 Replace so	urce column
			African American	Black or African American			
			American Indian	American Indian or Alaska N			
			Alaska Native	American Indian or Alaska N			
			source value	target value	+		
			When source value • NULL • Source	is not found in map, target value is: e value Constant: value			
				¥			
	Select						
	Available Columns		Selected Colum	ins			



scharedemo / demo-files / LIVE

🖉 🌀 😭 Operations 🗸

#### My Collections >

숬 Starred >



( Recent )

र्द्र Starred >

My Collections >

#### scharedemo / demo-files / LIVE ØD Operations -☆ Access Level ① demo-files A Private Abstract November 20 Think-a-Thon Demonstrati Match CDEs with column headings Levels of Influence The first step is to automatically detect alignments Don **Analysis Readiness** Individual Soci between your data and the ScHARe CDE set. Unless Not Ready > otherwise indicated, please select ScHARe below. Select the CDEs you want to assign: > Links and Documents ScHARe (19 CDEs) ÷ CDE Compliance - ScHARe (1) CDE matching method: Data Items $\sim$ 🗙 0 / 19 CDEs assigned Using column headers ÷ STATUS NAME SIZE Using column headers ⊞ example data.csv 25 KB Tags Manual assignment (no search) # Topics tagged in this collection Page 1 of 1 > State-Level Data Cross-Sectional Study Drag and Drop or Browse Files to Upload



C Recent >

My Collection

숬 Starred

		Assign CDEs to Uploaded Data Columns Assign your uploaded item's data columns to their	appropriate commo	n data elements.					
	demo	More Info >					í	ccess Lev	el (j)
		CDE set selected: ScHARe						Private	
	Abstract	🔤 Postal Zip Code		zip_code 🗠	~	~			When an assigned CDE
	laure of t	Race/Ethnicity Self-Identification		race_ethnicity_1 <=	~	X			fails validation hover
	Individual	s≡ Sex at Birth	Row 1 "no" Row 2 "yes	is not one of ["American " is not one of ["American	Indian or n Indian o	Alaska Ni r Alas			over the red X to get
		Gender	Row 3 "no"	is not one of ["American	Indian or	Alask			details on why
>	Links a	≌∓ Gender	Row 4 "no" Row 5 "yes Row 6 "no"	is not one of ["American " is not one of ["American is not one of ["American	n Indian o n Indian o n Indian o	r Alask r Alas			NOTE: CDE validation
		📰 Gender - Select Other	Row 7 "no"	is not one of ["American	Indian or	Alask		DE Com	errors can be corrected
~	Data Ite	🐲 Gender - Specify	Row 8 "no" Row 9 "no"	is not one of ["American is not one of ["American	n Indian o Indian o	Alask		3 0/19	using dataviews (coming
	STATI	Sexual Orientation	Row 10 "ye Row 11 "no	s" is not one of ["America " is not one of ["America	an Indian n Indian o	or Ala r Alas			up later)
	Œ	Sexual Orientation	<			- F		105	
		Sexual Orientation - Select Other			~			t Topics	tagged in this collection
		ec Sexual Orientation - Specify			~			Tross Secti	
	(	📰 Marital Status		marital_status 🛪	~	~		Bruss-Betti	Char Study State-Lever Data
		or Education					-		
		Assign More Options Cancel							

Recent >

My Collections

숬 Starred ゝ

	Assign CDEs to Uploaded Data Columns Assign your uploaded item's data columns to their appro	opriate common data elements.		
demo	More Info >		i	ccess Level (1)
uemo	CDE set selected: ScHARe			Private
Abstract Novemb	🐲 Postal Zip Code	zip_code ⊲ 🗸 🗸	×	
Levels of I	ST Race/Ethnicity Self-Identification	race_ethnicity_1 ⊲ ►	×	
Individual	≌≓ Sex at Birth	<u>^</u>	~	nalysis Readiness
Links ar	Gender	record_id age	×	Not Ready >
✓ Data Ite	s∰ Gender - Select Other ⊯c Gender - Specify	race_ethnicity_1 → race_ethnicity_2 →	Č.	DE Compliance - ScHARe () 0 / 19 CDEs assigned
E	Sexual Orientation         Image: Sexual Orientation         Image: Sexual Orientation - Select Other         Image: Marital Status         Image: Education         Assign       More Options         Cancel	race_ethnicity_3 <> race_ethnicity_4 <> race_ethnicity_5 <> race_ethnicity_6 <> race_ethnicity_7 <> zip_code sex_at_birth gender	✓ ▼	ags Topics tagged in this collection Cross-Sectional Study State-Level Data

scharedemo / de	Assign CDEs to Uploaded Data Columns Assign your uploaded item's data columns to their a	ppropriate common data elements.				Ø ් Coperations -
tions >	More Info >				i	ccess Level ①
demo						Private
Abstract	CDE set selected: ScHARe					
Novemb	*** Postal ZIP Code	zip_code <=	~	×	-	
Levels of I	ST Race/Ethnicity Self-Identification		~			
Individual	🚛 Sex at Birth	sex_at_birth 🔫	~	~		nalysis Readiness
	Gender					Not Ready
> Links a	📰 Gender	gender	~	×		
M. D. L.	📰 Gender - Select Other	Row 1 "None of these describe me Row 2 "None of these describe me	e" is not on e" is not on	e of ["Man e of ["	<b>^</b>	DE Compliance - ScHARe (1)
✓ Data ite	🕸 Gender - Specify	Row 3 "Trans" is not one of ["Man	","Woman"	,"Non		0 / 19 CDEs assigned
STATI	Sexual Orientation	Row 6 "Male" is not one of [ Man	", woman ","Woman",	, Non ,"Non		
F	Sexual Orientation	Row 7 "Trans" is not one of ["Man Row 8 "Female" is not one of ["Ma	","Woman" an","Woma	,"Non n","No		
	Sexual Orientation - Select Other Row 10 "Trans" is not one of ["Man", "Woman", "N Bow 11 "Male" is not one of ["Man" "Woman", "N			","Non		igs
	** Sexual Orientation - Specify	Row 13 "Male" is not one of ["Mar Row 14 "Trans" is not one of ["Mar	n","Woman n" "Woman	","Non	_	Cross-Sectional Study State-Level Data
(*************************************	🚝 Marital Status	4		•	•	
	of Education					
	Assign More Options Cancel					

Recent <b>&gt;</b>	scharedemo / de	Assign CDEs to Uploaded Data Columns Assign your uploaded item's data columns to their approp	riate common data elements.				Ø ් Coperations •
My Collections		More Info >				í	coars Level ①
, statien >	<b>demo</b> Abstract Novemb	CDE set selected: ScHARe <sup>abc</sup> Postal Zip Code	zip_code ⊲	~	~	•	Private
	Levels of I Individual	Image: Sex at Birth         Gender	sex_at_birth ◄	* *	~	4	nalysis Readiness Not Ready >
	<ul> <li>Links an</li> <li>Data Ite</li> </ul>	s∰ Gender s∰ Gender - Select Other	gender_select_other ◄	*	~		DE Compliance - ScHARe ()
	STATI	Sexual Orientation  Sexual Orientation  Sexual Orientation - Select Other  Sexual Orientation - Specify	gender_specify ~	*	~		ags <b>t</b> Topics tagged in this collection
		Marital Status     Assign More Options Cancel	marital_status ◄	•	~	•	Cross-Sectional Study State-Level Data

#### **Viewing Uploaded Data** All data is initially scharedemo / demo-files / LIVE private (only you can access it). Recent > My Collections > Access Level demo-files A Private Abstract Click the item name to November 20 Think-a-Thon Demonstration Files view the item's data and Levels of Influence Domains of Influence Analysis Readiness more information Sociocultural Environment Individual Ready > **Viewing Uploaded Data** nd Documents CDE Compliance - ScHARe (1) $\sim$ Data Items 8 / 19 CDEs assigned STATUS NAME CREATED SIZE æ example\_data.csv 5 minutes ago 25 KB Tags << Page 1 of 1 > # Topics ta View the CDEs that you Cross-Sectio have assigned (coming up next) Drag and Drop or Browse Files to Upload > Metadata and Statistics

### **Viewing Uploaded Data**

 $\bigcirc$ 

☆

Bernet A	charedemo / <b>demo-files</b> / LIVE			の う ☆ Operations 🗸
Recent >				
My Collections >				
Starred >	demo-files Abstract			Access Level (1) Private
	November 20 Think-a-Thon Demonstra	tion Files		
	Levels of Influence Do Individual So	omains of Influence ciocultural Environment		Analysis Readiness
	> Links and Documents			
	✓ Data Items			8 / 19 CDEs assigned
	STATUS NAME	CREATED	) SIZE	
	Storage: $\checkmark ok$ Parser: $\checkmark ok$ Validator: $\checkmark ok$	5 minutes ago	< Page 1 of 1 >	Tags # Topics tagged in this collection
The item status show all good or any problems in storage parsing, or validatio	/s , n	Drag and Drop or <u>Browse Files</u> to Upload		Cross-Sectional Study State-Level Data
	<ul> <li>Metadata and Statistics</li> </ul>			

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

#### **Viewing Uploaded Data**

Click the collection name to return to the collection main page

scharedemo / demo-files / LIVE / example\_data.csv

#### Recent >

My Collections >

숬 Starred 🔉

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

	e Dictio	onary weta	25 KB   10 MIN	utes ago   text/	(CSV   Status: 🗄	5					Item O	perations 👻
record_id	age 🚯	age_units 💽	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_et	thnicity_7 zip_code	g sex_at_birt	gender
1943	92	Years	10	no	no	yes	no	no	no	28752	Male	None o these describ me
6688	76	Years	)*	no	no	no	no	no	no	10032		None these descrit me
9784	18	Years	<b>no</b>	no	VAS	no	no	yes	no	07501		Trans
5193	91	Years	For colum	nns with CDE	Es	no	yes	no	no	Perform ope	rations on	Trans
8502	86	Years	assigned,	click the inf	o	no	no	no	no	the d	ata	Prefer to ans
1210	72	Years	popup to	view the CD	E	no	yes	no	no	(download,	rename,	Male
3563	79	Years	d	etails		no	no	no	no	move, dele	ete, etc.)	Trans
1172	46	Years	no	no	no	no	no	yes	no	37208	Prefer not to answer	Femal
2618	80	Years	no	no	no	yes	no	no	no	35405	Male	Prefer
6809	58	Years	yes	no	no	no	no	no	no	11769	None of the describe n	-(