ScHARe

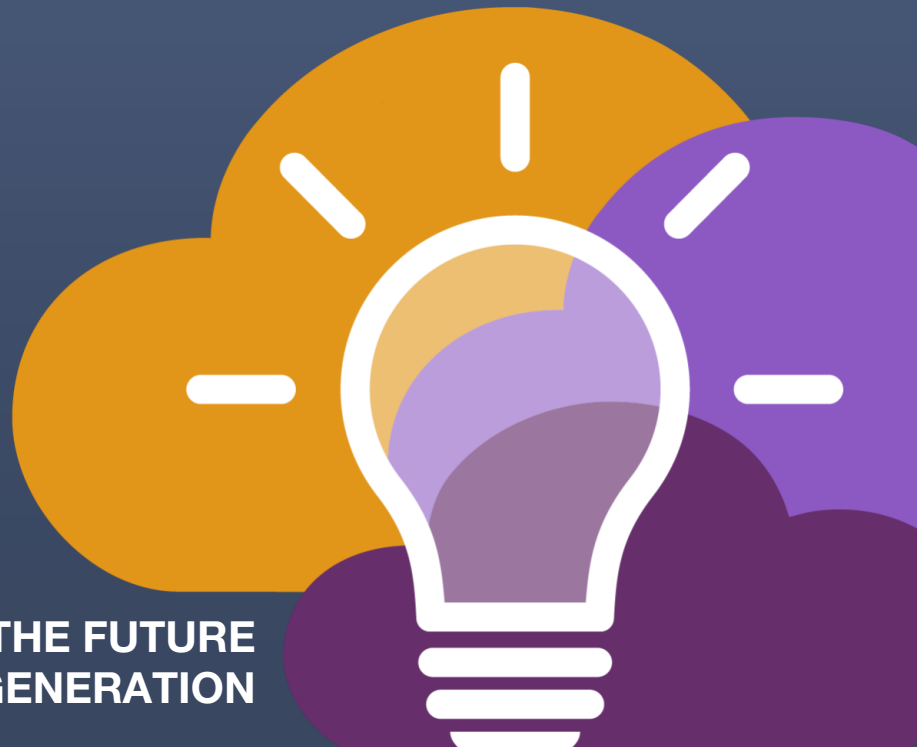
ScHARe Repository
Introduction

November 20, 2024

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

ScHARe

CDE Mapping

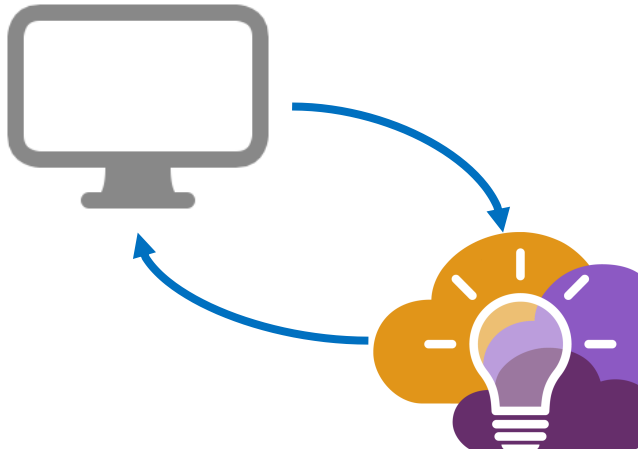


BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

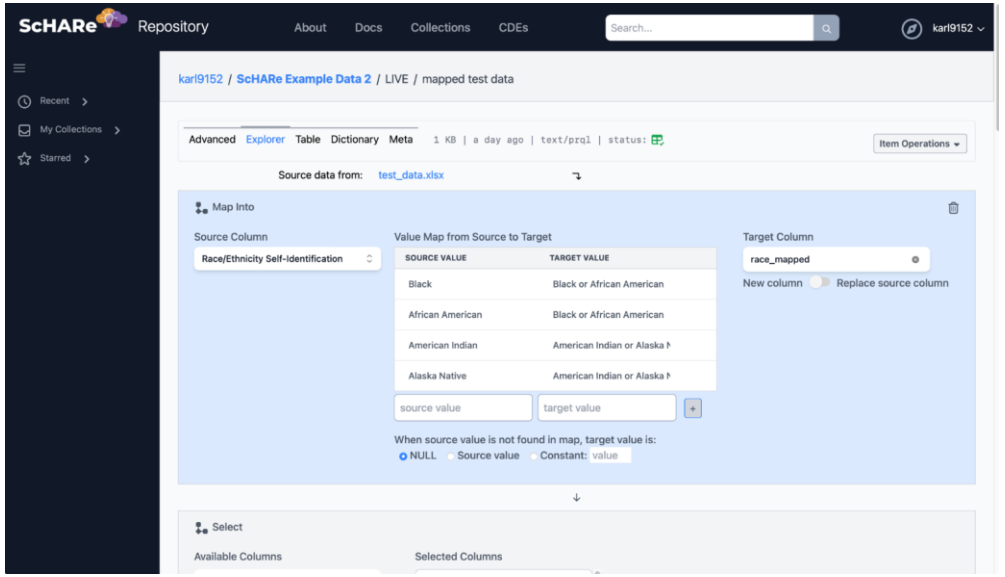
Using Dataviews to Map CDEs

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

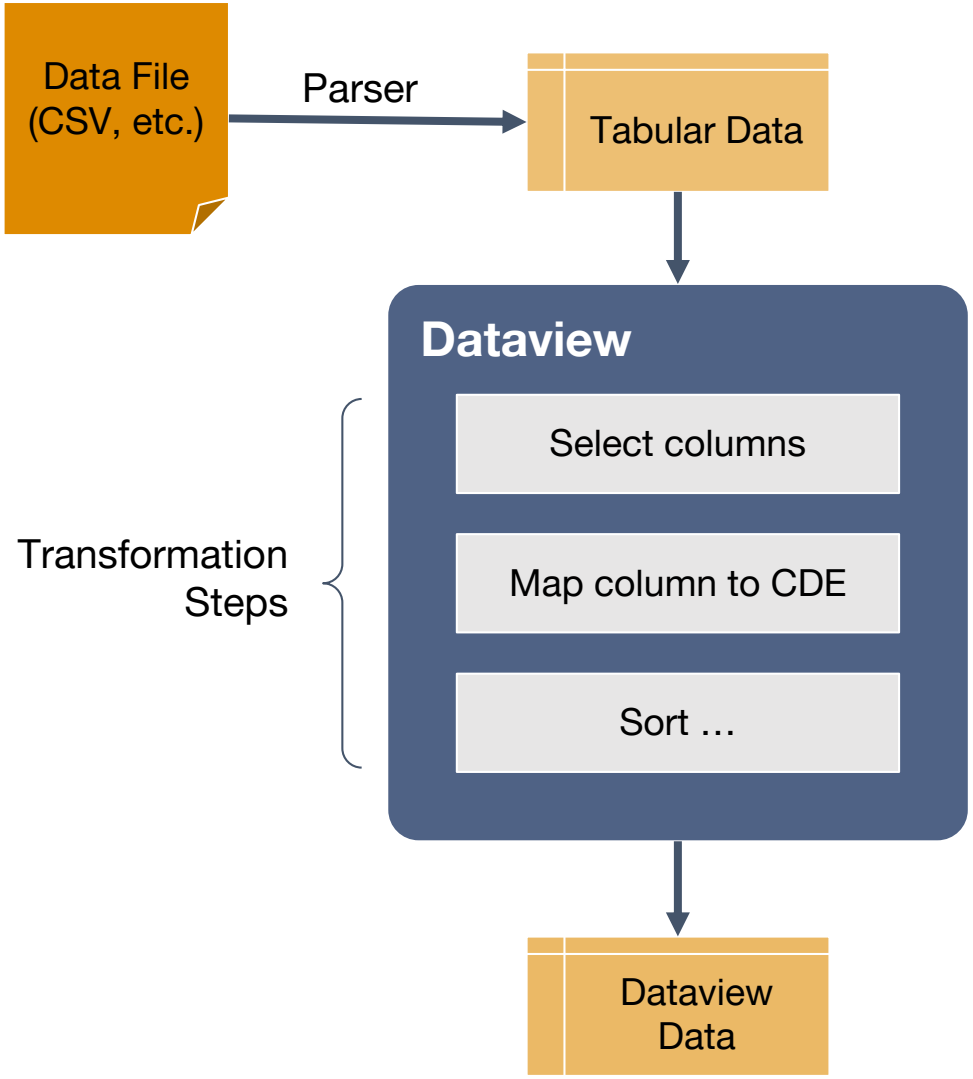
Use your own tools to adjust the data and re-upload



Use a Dataview to Map to CDEs within the Repository



What is a Dataview?

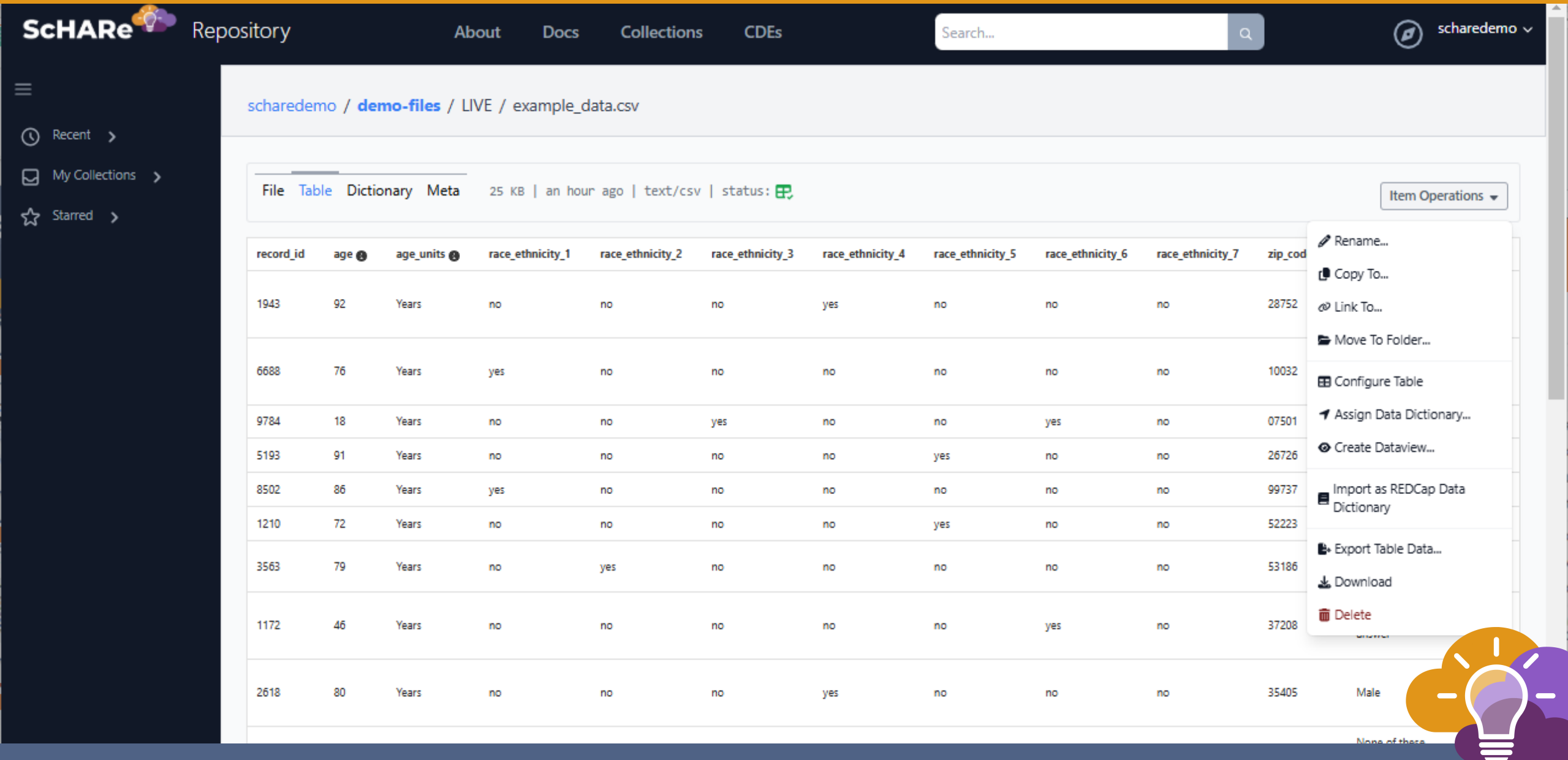


Dataviews take data from one or more sources, apply a series of transformation steps to that data *(filtering, sorting, mapping, etc.)* resulting in a new table of data as output.

Uses of Dataviews:

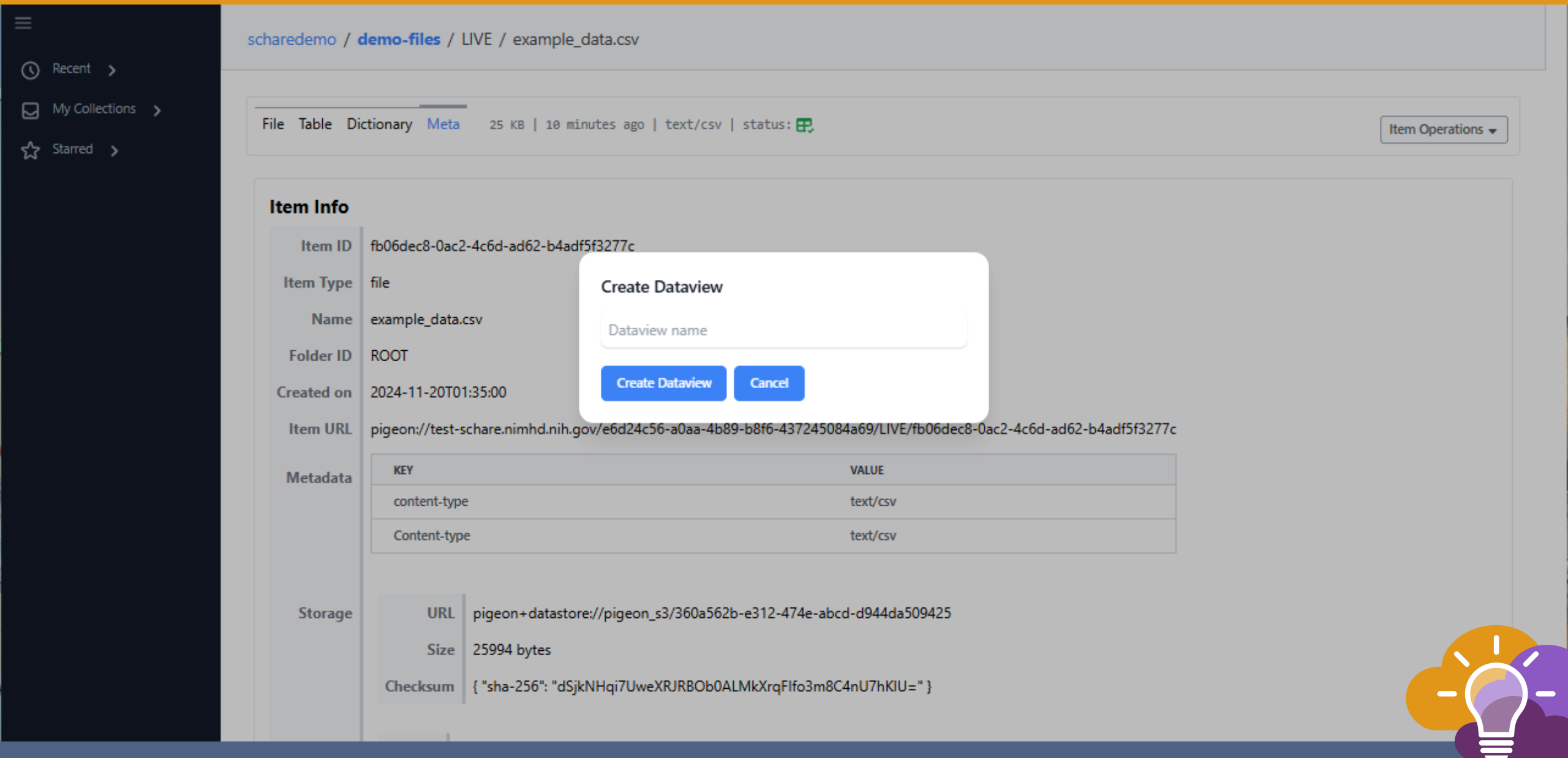
- Creating subsets of data
- Hiding PHI/PII for publishing
- Summarizing individual-level data into subsets and estimates
- Joining multiple datasets together
- **Mapping to CDEs**
- … many others!

Mapping CDEs via Dataview

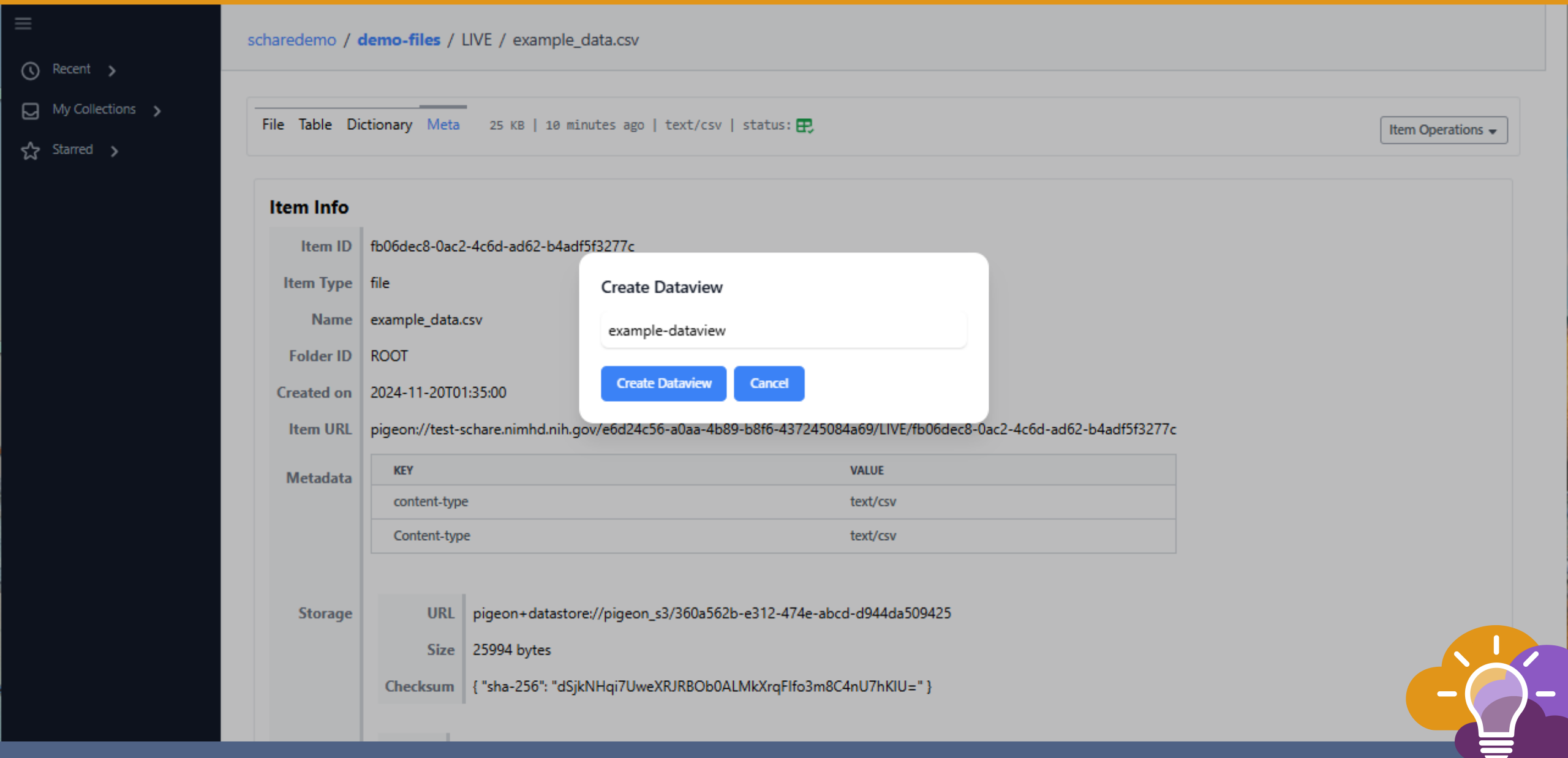


record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_cod	
1943	92	Years	no	no	no	yes	no	no	no	28752	
6688	76	Years	yes	no	no	no	no	no	no	10032	
9784	18	Years	no	no	yes	no	no	yes	no	07501	
5193	91	Years	no	no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

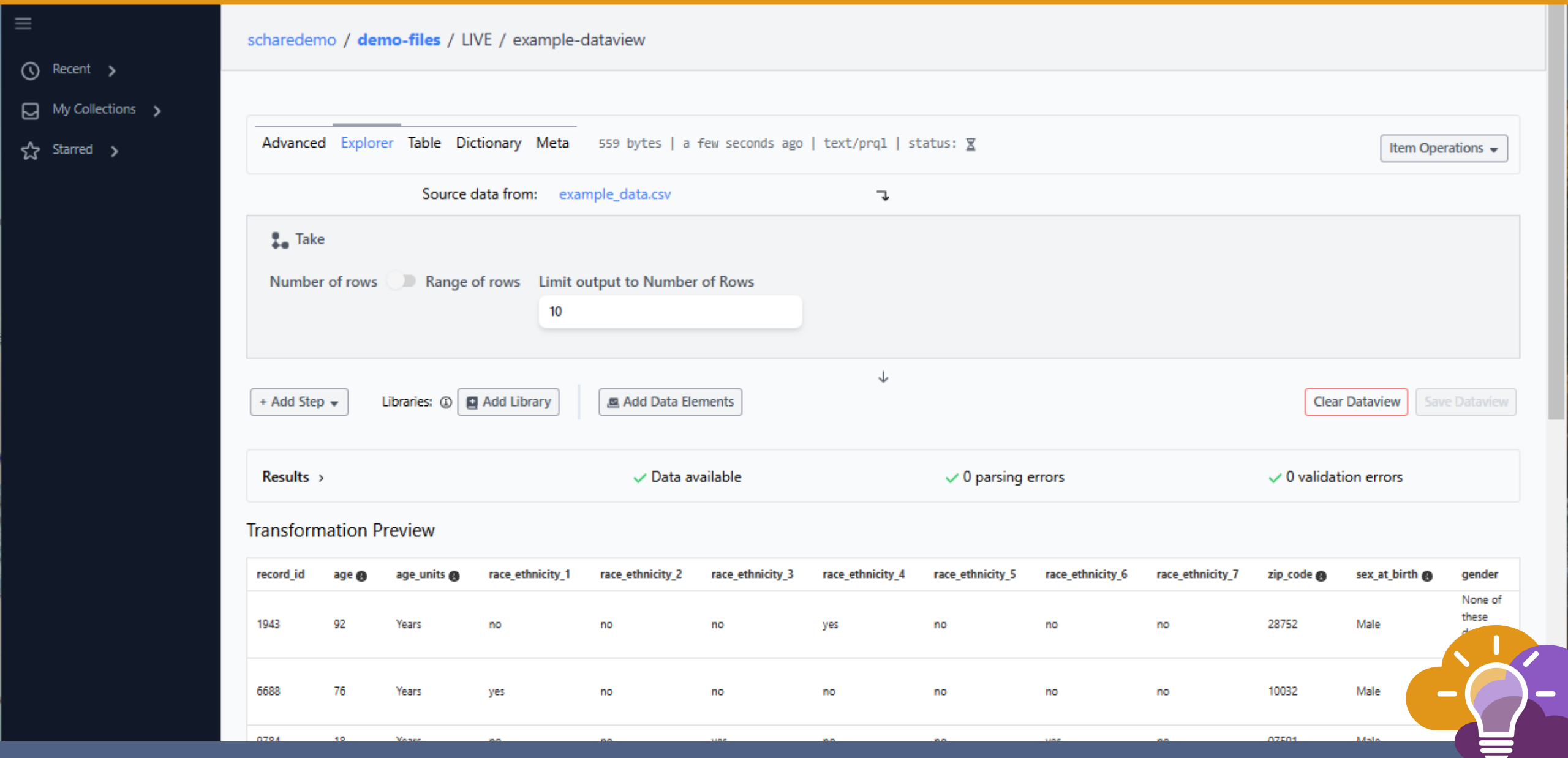
Step 1: Create a new Dataview



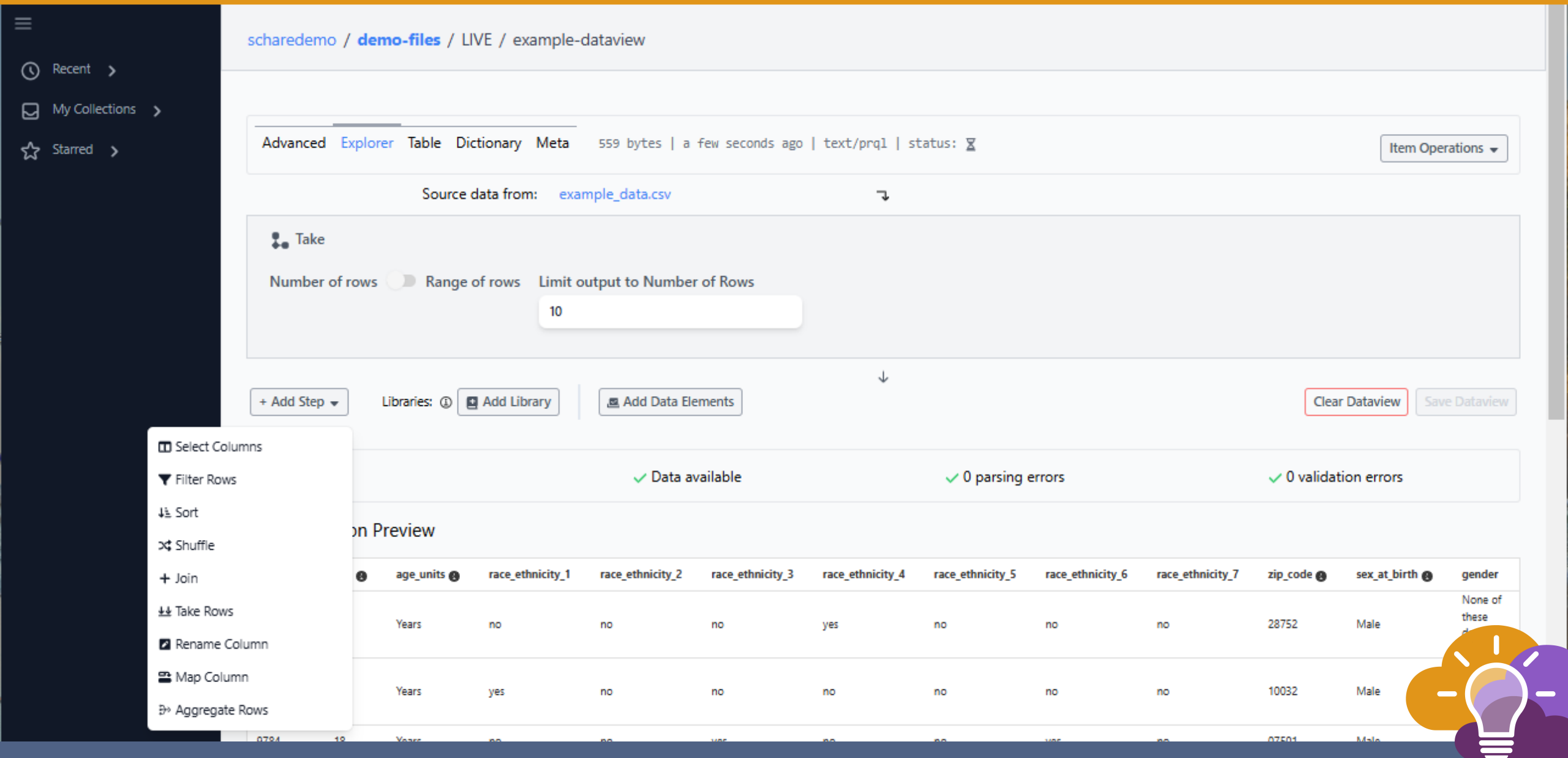
Step 1: Create a new Dataview



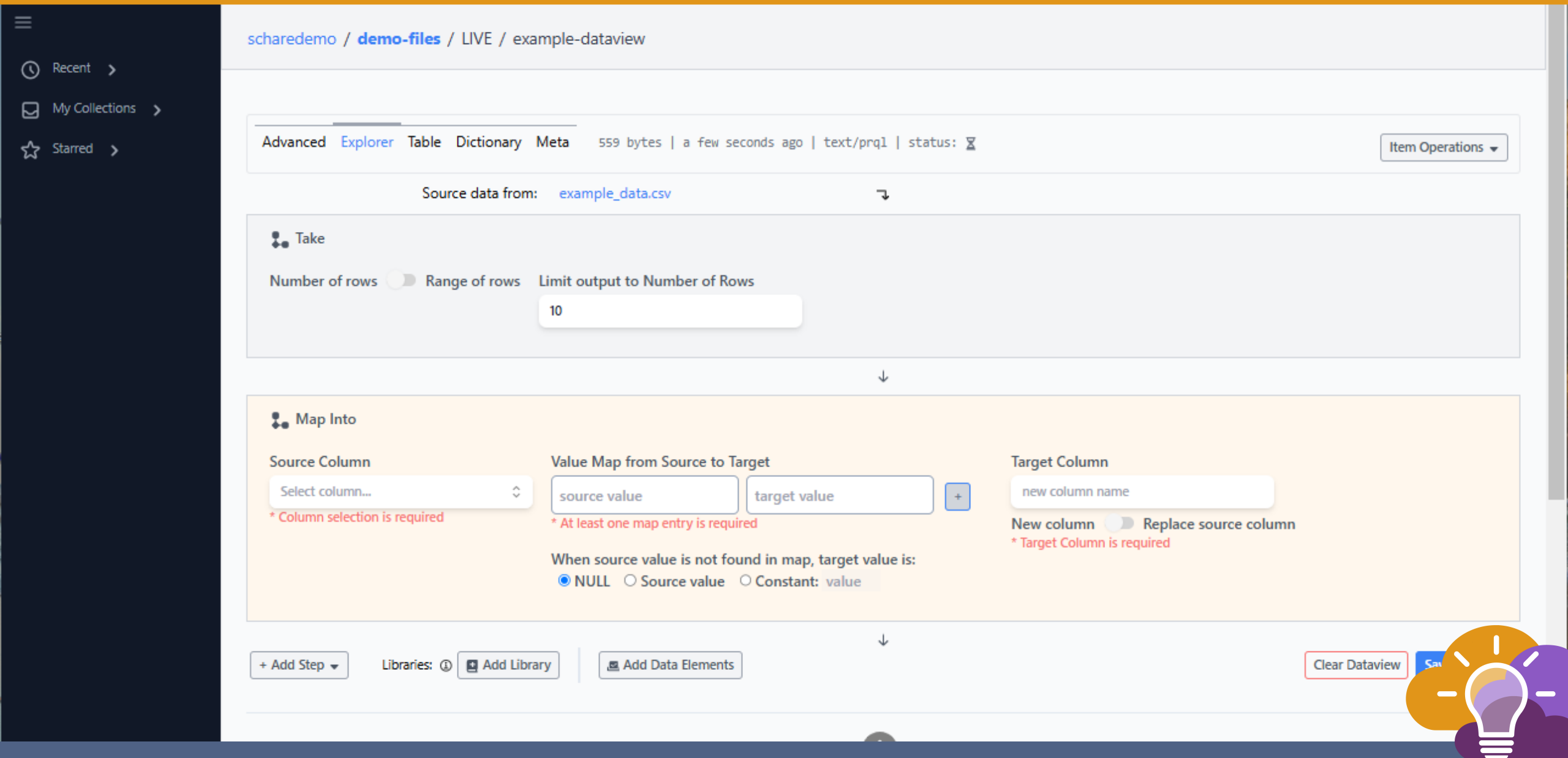
Step 2: Add Mapping Steps to your Dataview



Step 2: Add Mapping Steps to your Dataview



Step 2: Add Mapping Steps to your Dataview



Step 2: Add Mapping Steps to your Dataview

Map Into

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

source value | target value | +

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

NULL | Source value | Constant: value

Target Column

gender_mapped

New column | Replace source column

+ Add Step | Libraries: | Add Library | Add Data Elements | Clear Dataview | Save Dataview

Results › | ✓ Data available | ✓ 0 parsing errors | ✓ 0 validation errors

Transformation Preview

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

Step 2: Add Mapping Steps to your Dataview

Map Into

Source Column

gender

Value Map from Source to Target

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

source value | target value | +

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

Target Column

gender_mapped

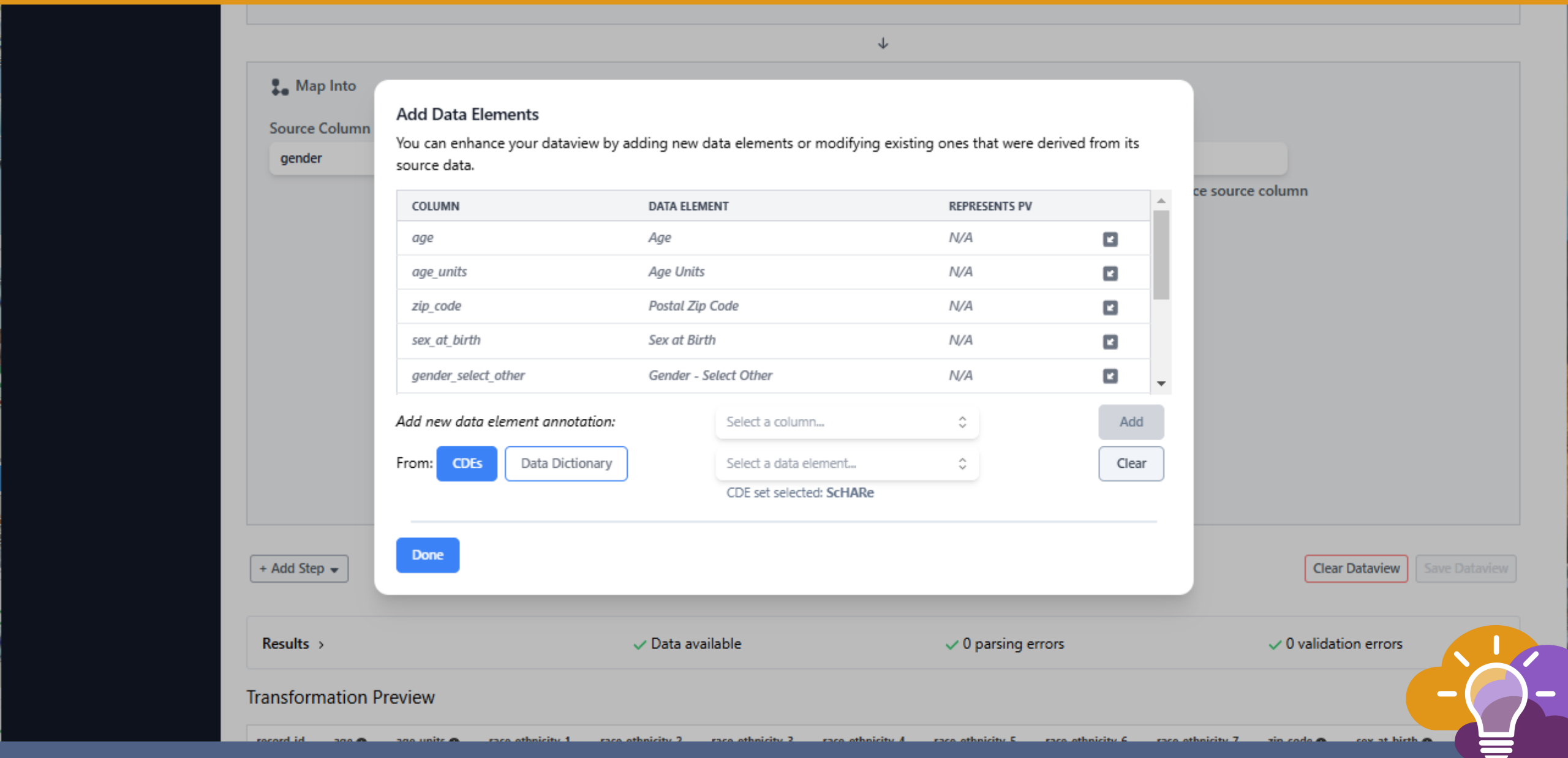
New column | Replace source column

+ Add Step | Libraries: | Add Library | Add Data Elements | Clear Dataview | Save Dataview

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

Transformation Preview

Step 3: Assign CDEs to Data Columns



Step 3: Assign CDEs to Data Columns

Map Into

Source Column

gender

Add Data Elements

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

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

Add new data element annotation: gender_mapped — Add

From: CDEs | Data Dictionary — Gender — Clear

CDE set selected: ScHARe

OPTIONAL: Select a PV...

Done

ce source column

+ Add Step

Clear Dataview | Save Dataview

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

Transformation Preview

Step 3: Assign CDEs to Data Columns

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

No errors to report!

Transformation Preview

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

rows per page: 25 1–10 of

Step 3: Assign CDEs to Data Columns

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

Target Column

gender_mapped

New column | Replace source column

Add Data Elements

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

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

Add new data element annotation:

Select a column...

Add

From: CDEs | Data Dictionary

Select a data element...

Clear

CDE set selected: ScHARe

Done

+ Add Step

Clear Dataview | Save Dataview

Results

0 validation errors

No errors to report

Transformation Preview

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

Step 3: Assign CDEs to Data Columns

Source Column: gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE

Target Column: gender_mapped

...ce source column

Add Data Elements

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

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

Add new data element annotation: race_ethnicity_1 — Add

From: CDEs | Data Dictionary — Race/Ethnicity Self-Identification — Clear

CDE set selected: ScHARe

American Indian or Alaska Native

Done

+ Add Step

Clear Dataview | Save Dataview

Results

✓ 0 validation errors

No errors to report

Transformation Preview

...iency	measure_0	measure_1	measure_2	measure_3	measure_4	measure_5	measure_6	measure_7	measure_8	measure_9	gen...
	62	0.5455549357622266	1	0.5884900154230546	0.12911211858353722	57	33	0.7170385313006806	0.8788787389282339	0.5467863499436898	addition

Step 3: Assign CDEs to Data Columns

Source Column

gender

Value Map from Source to Target

SOURCE VALUE	TARGET VALUE
Male	Man

Target Column

gender_mapped

New column | Replace source column

Add Data Elements

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

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

Add new data element annotation:

From: CDEs | Data Dictionary

Select a column...

Select a data element...

CDE set selected: ScHARe

Add

Clear

Done

+ Add Step

Clear Dataview | Save Dataview

Results

0 validation errors

No errors to report

Transformation Preview

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

Step 3: Assign CDEs to Data Columns

+ Add Step | Libraries: | Add Library | Add Data Elements | Clear Dataview | Save Dataview

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

No errors to report!

Transformation Preview

record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth	gender
1943	92	Years	no					no	no	28752	Male	None of these describe me
6688	76	Years	yes					no	no	10032	Male	None of these describe me
9784	18	Years	no					yes	no	07501	Male	Trans
5193	91	Years	no					no	no	26726	None of these describe me	Trans
1210	72	Years	no					no	no	52223	Female	Male
3563	79	Years	no					no	no	53186	None of these describe me	Trans
1172	46	Years	no	no	no	no	no	yes	no	37208	Prefer not to answer	
6809	58	Years	yes	no	no	no	no	no	no	11769	None of	
8502	86	Years	yes	no	no	no	no	no	no	99737	Prefer not to	

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

Concepts

C41259: American Indian or Alaska Native	*valueDomain*
C17049: Race	*dataElement*
C16564: Ethnic Group	*dataElement*
C74528: Self-Report	*dataElement*

Definitions ›

Step 3: Assign CDEs to Data Columns

Trans	Transgender

Add Data Elements

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

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

Add new data element annotation:

From: CDEs | Data Dictionary

Select a column...

Select a data element...

CDE set selected: ScHARe

Add

Clear

Done

+ Add Step

Clear Dataview | Save Dataview

Results

✓ 0 validation errors

No errors to report

Transformation

record_id	age								race_ethnicity_7	zip_code	sex_at_birth
1943	92								no	28752	Male
6688	76	Years	yes	no	no	no	no	no	no	10032	
9784	18	Years	no	no	yes	no	no	yes	no	07501	
5193	91	Years	no	no	no	no	yes	no	no	26726	

Step 3: Assign CDEs to Data Columns

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

No errors to report!

Transformation Preview

record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth
1943	92	Years	no	no					no	28752	Male
6688	76	Years	yes	no					no	10032	Male
9784	18	Years	no	no					no	07501	Male
5193	91	Years	no	no					no	26726	None of thes describe me
1210	72	Years	no	no					no	52223	Female
3563	79	Years	no	yes					no	53186	None of thes describe me
1172	46	Years	no	no	no	no	no	yes	no	37208	Prefer not to answer
6809	58	Years	yes	no	no	no	no	no	no	11769	None of thes describe me
8502	86	Years	yes	no	no	no	no	no	no	99737	Prefer not to answer
2618	80	Years	no	no	no	yes	no	no	no	35405	

Race/Ethnicity Self-Identification : Asian or Asian American
(string)

Concepts

C41260: Asian or Asian American	*valueDomain*
C17049: Race	*dataElement*
C16564: Ethnic Group	*dataElement*
C74528: Self-Report	*dataElement*

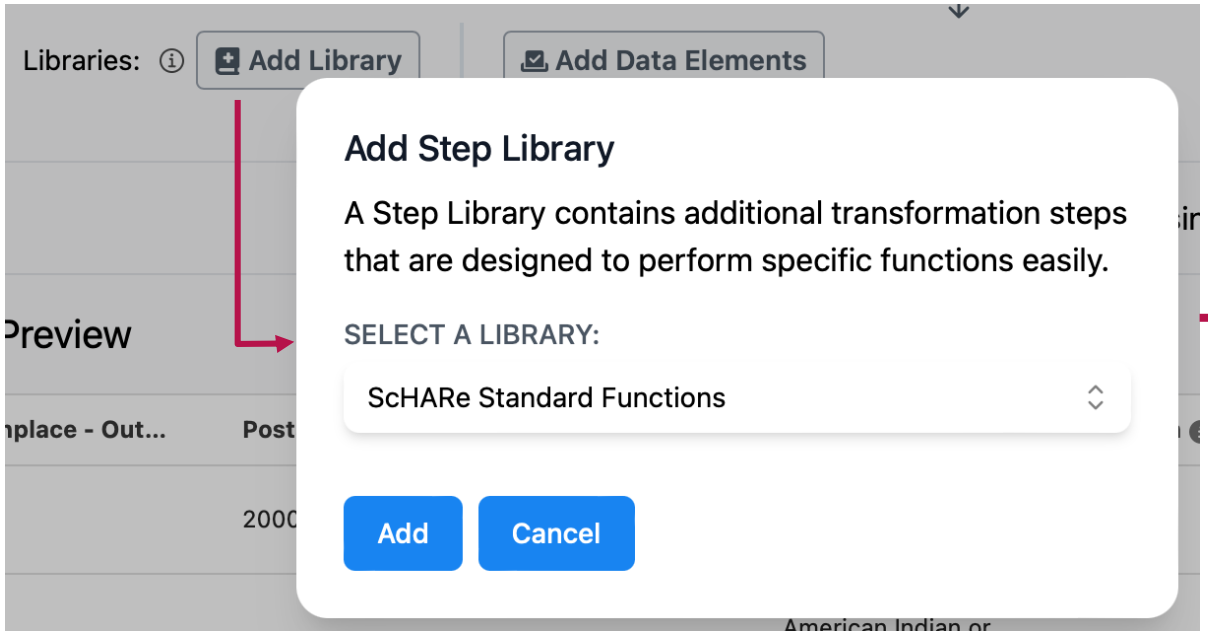
Definitions ›

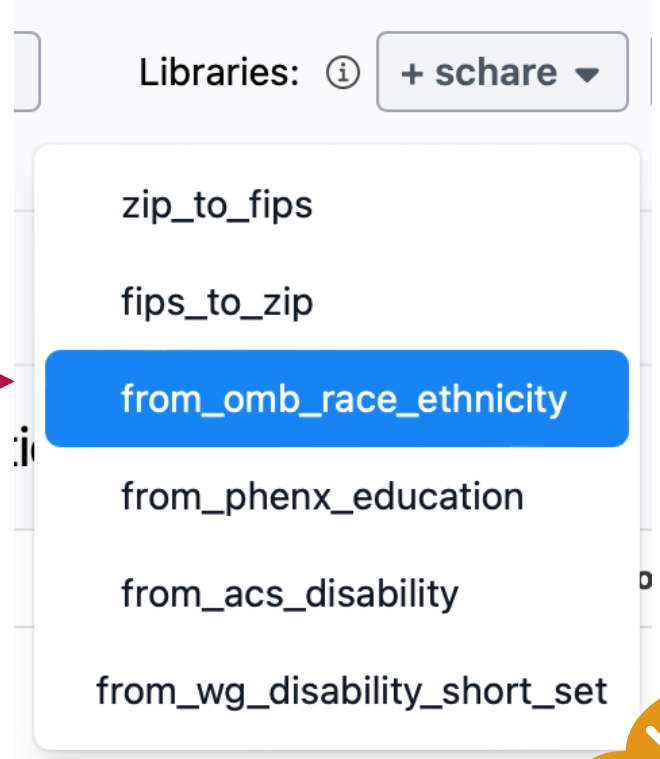
rows per page: 25 1–10 of

Mapping CDEs via Dataview

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

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





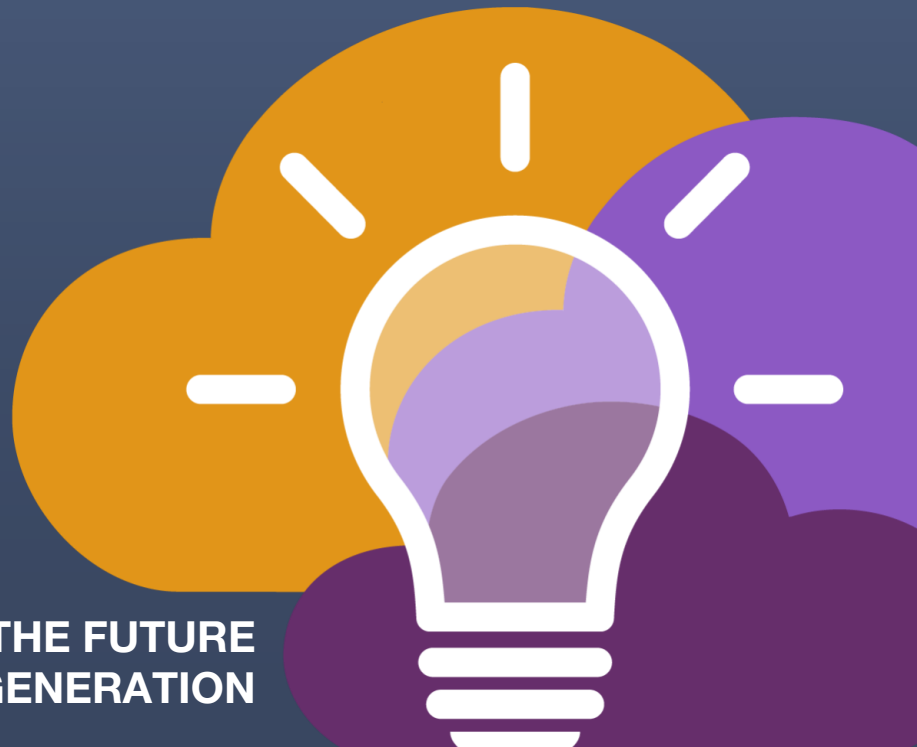
Slido Poll

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

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

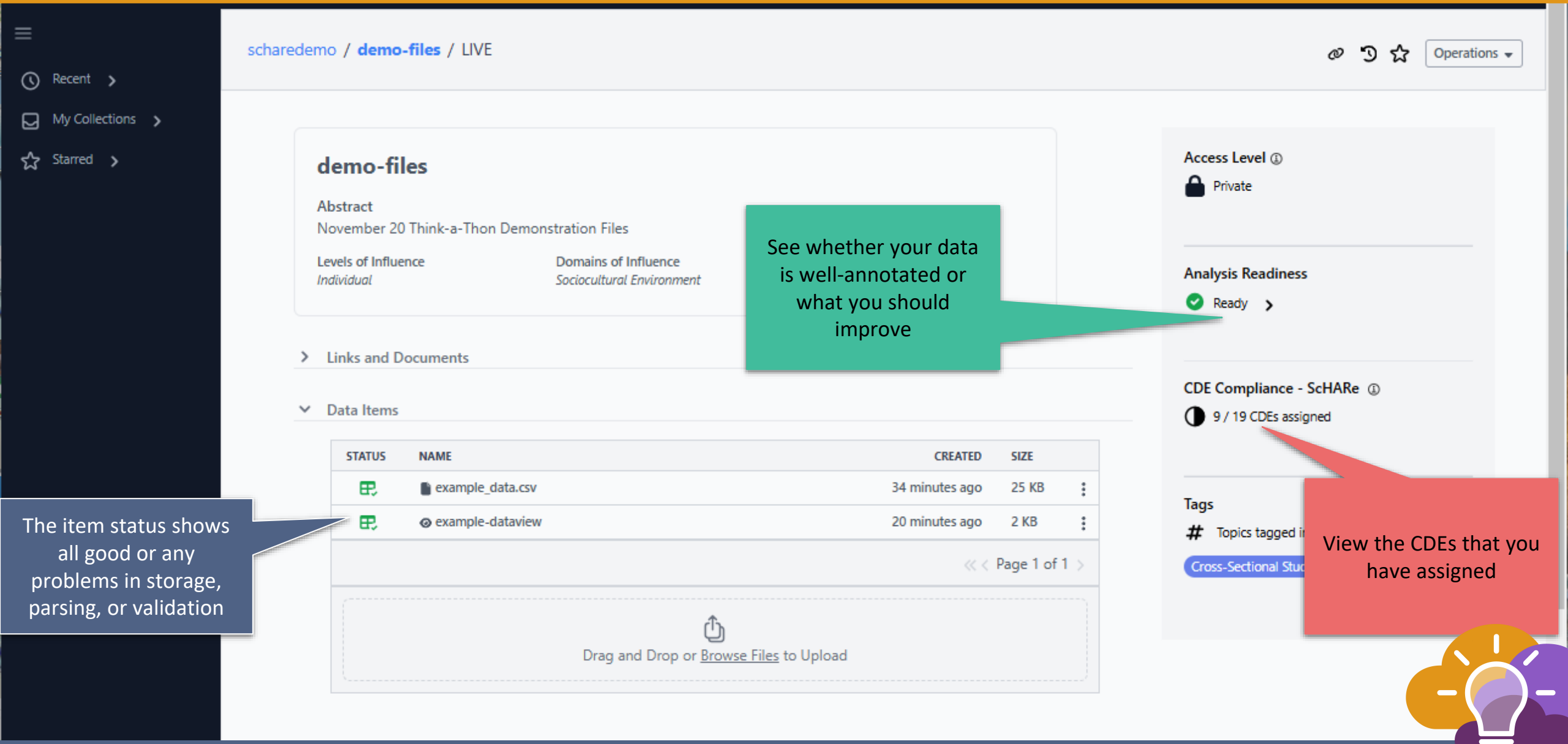
ScHARe

Sharing Data



BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

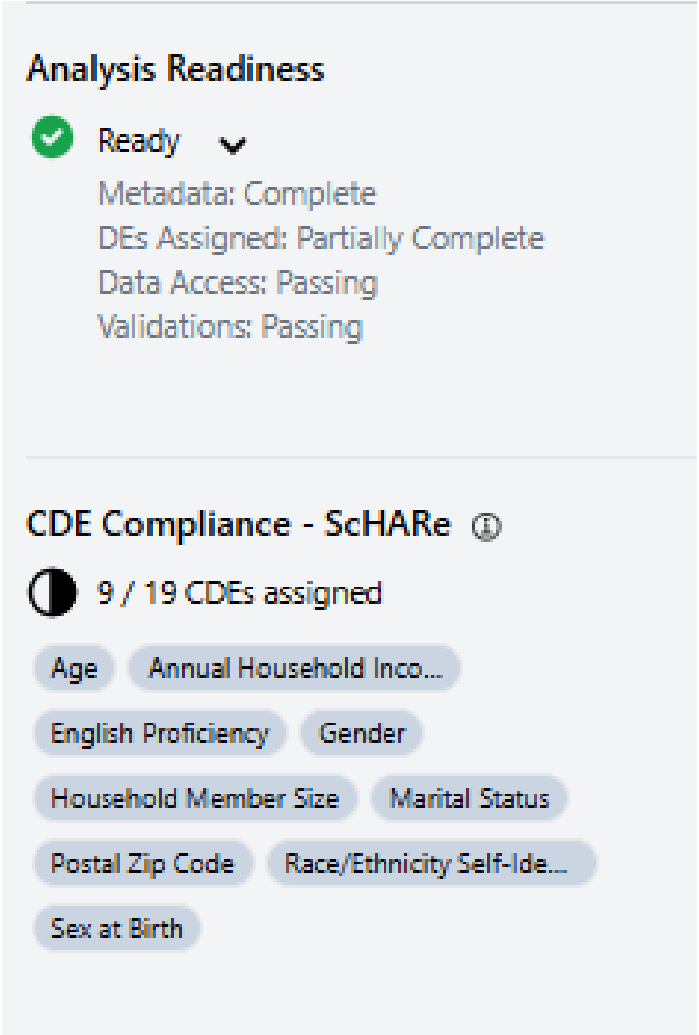
Viewing your Collection



CDE Compliance and Analysis Readiness

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

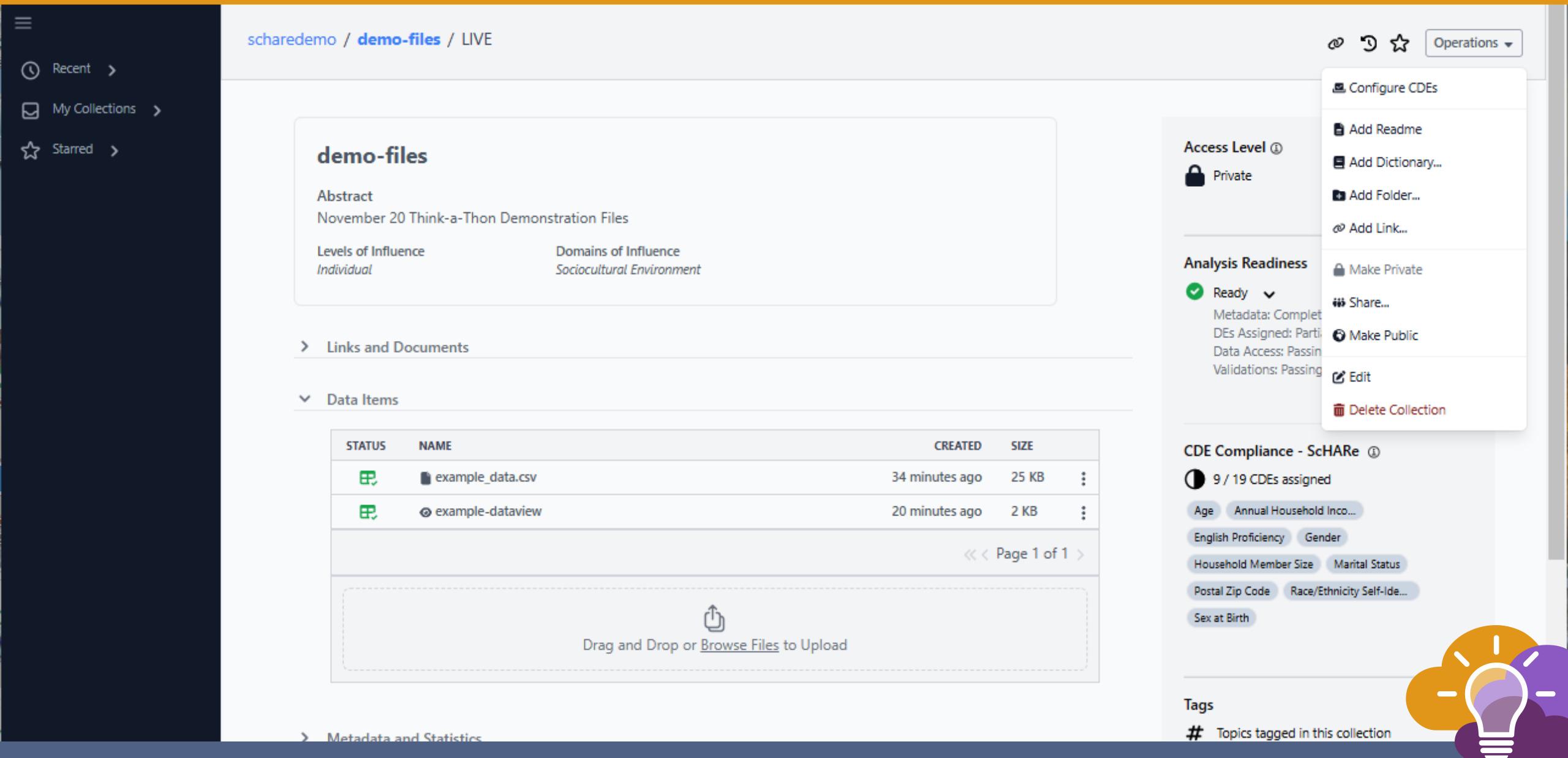
1. Did you assign **metadata** to your collection (tags, project-level CDEs)?
2. Have you partially or fully **assigned CDEs** to your data?
3. Is the data **accessible** to the system (no broken links)?
4. Does the data pass **validation** according to the assigned CDEs?



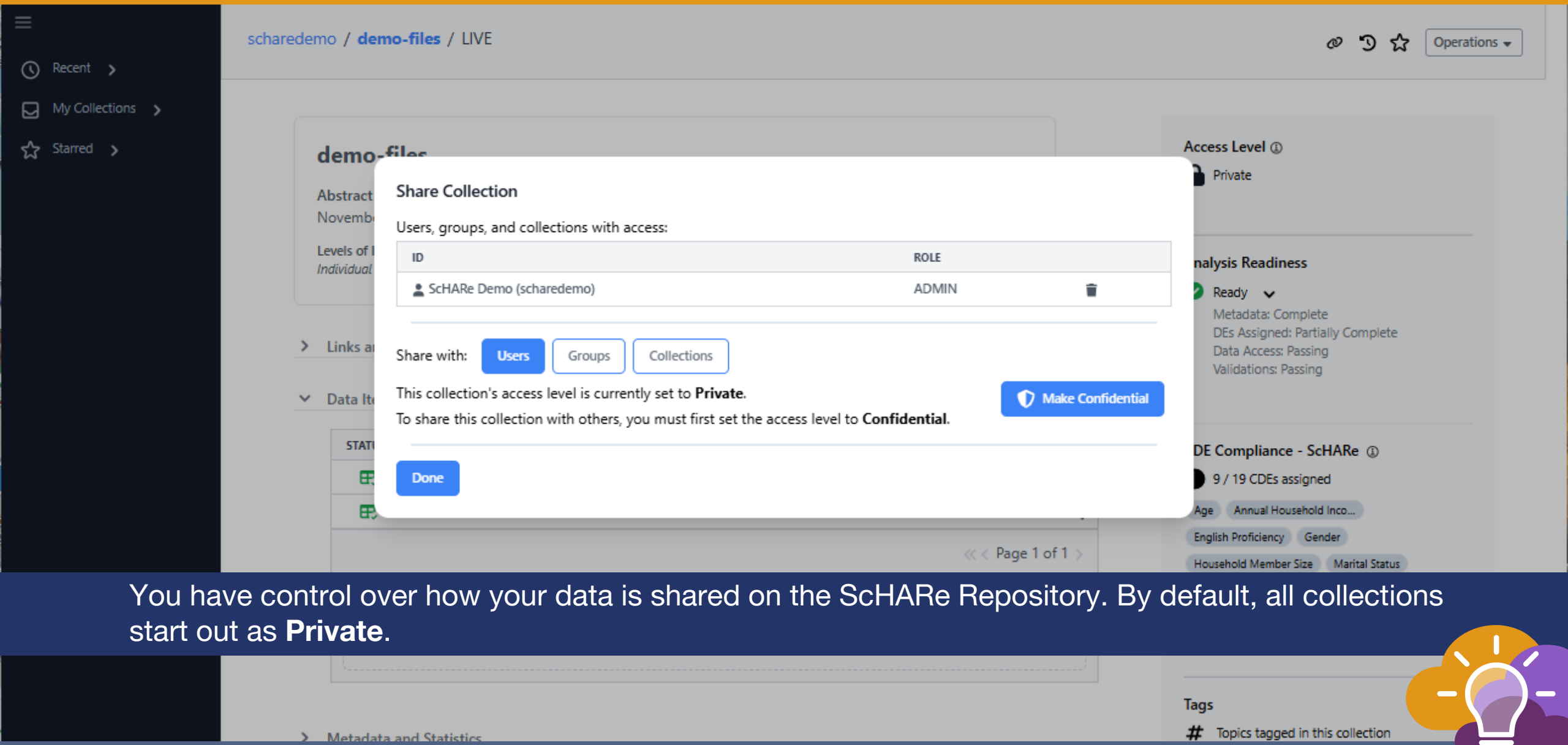
CDE Compliance - Showing how many CDEs have been assigned across the data in the collection

- 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

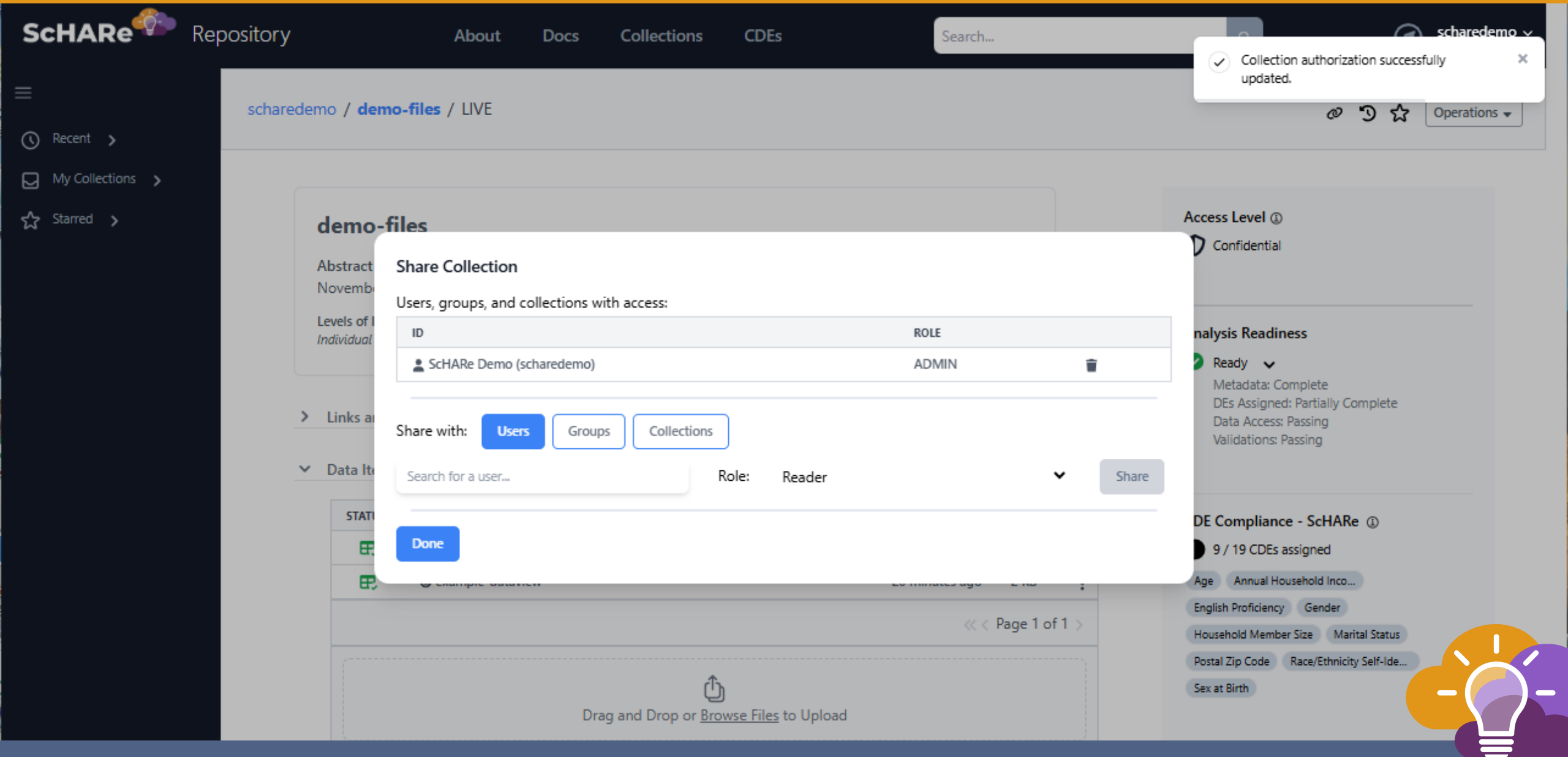


Access Levels and Sharing Data

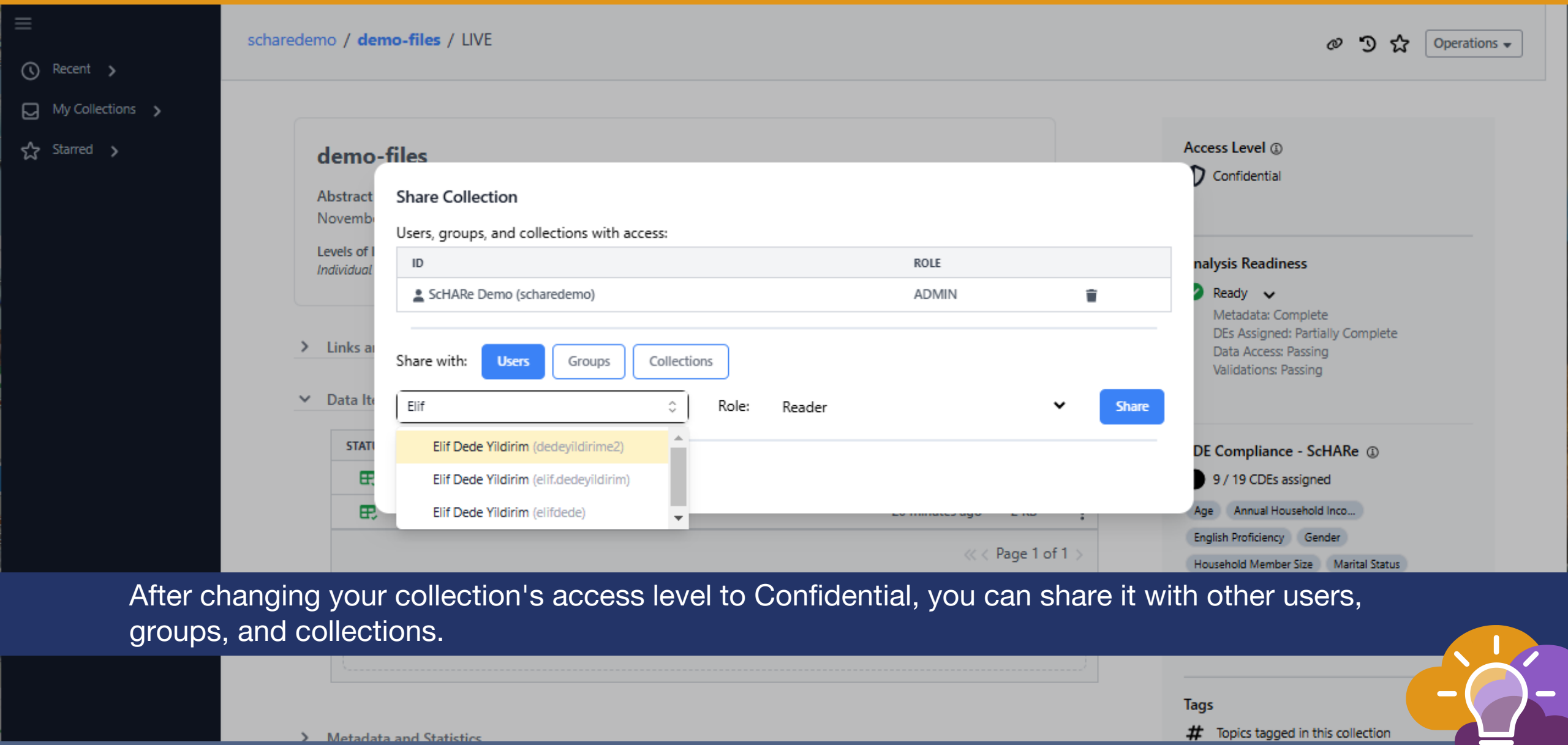


You have control over how your data is shared on the ScHARe Repository. By default, all collections start out as **Private**.

Access Levels and Sharing Data

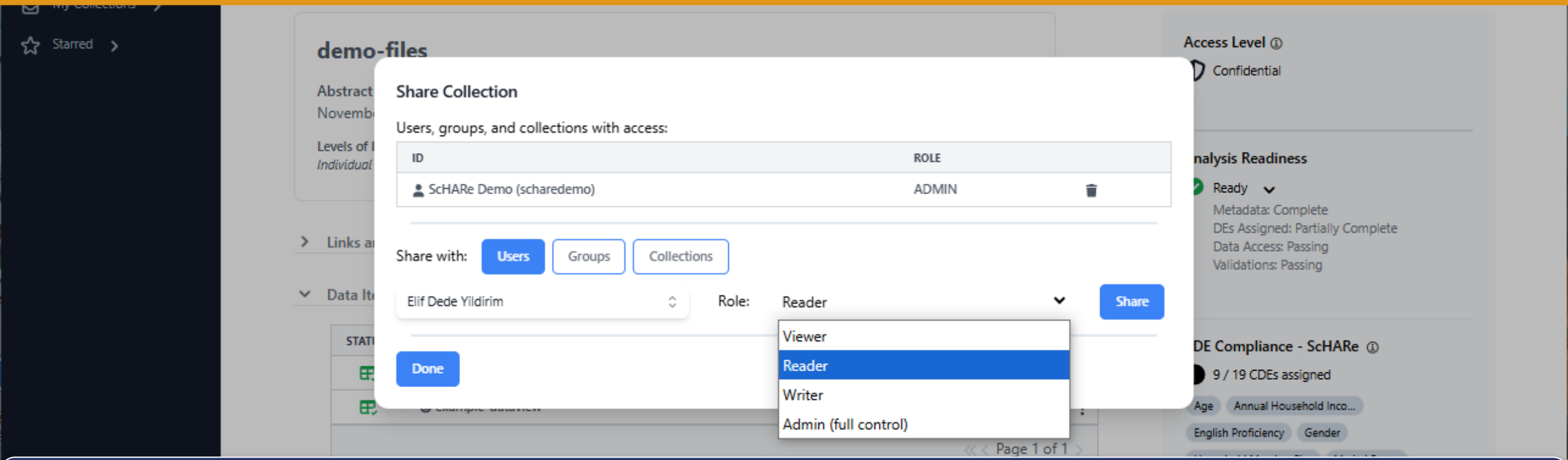


Access Levels and Sharing Data



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

Access Levels and Sharing Data

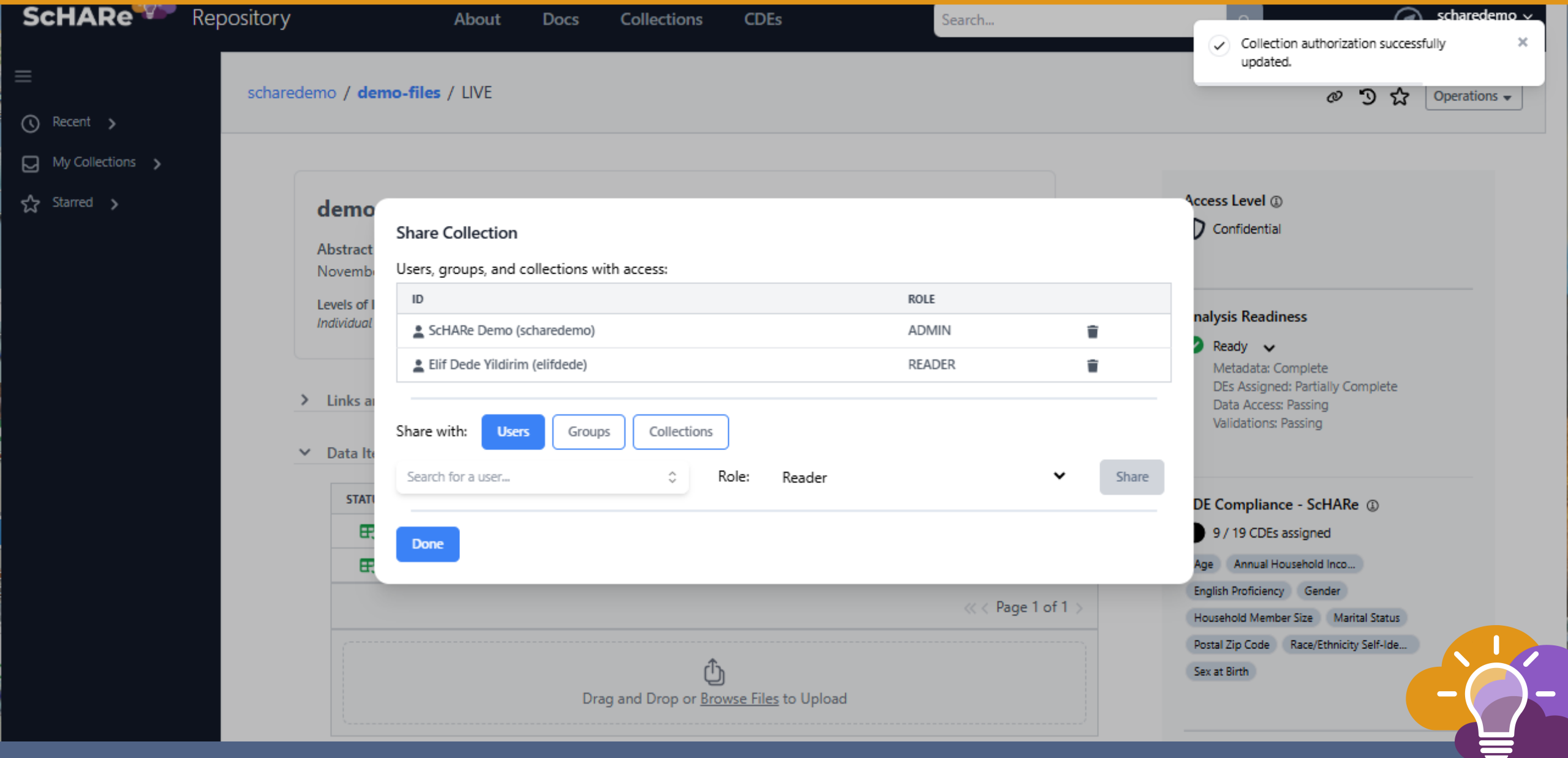


Access Levels

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



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