ScHARe Repository
Introduction

October 16, 2024

Deborah Duran, PhD · NIMHD
Luca Calzoni, MD MS PhD Cand. · NIMHD
Elif Dede Yildirim, PhD · NIMHD

ScHARe

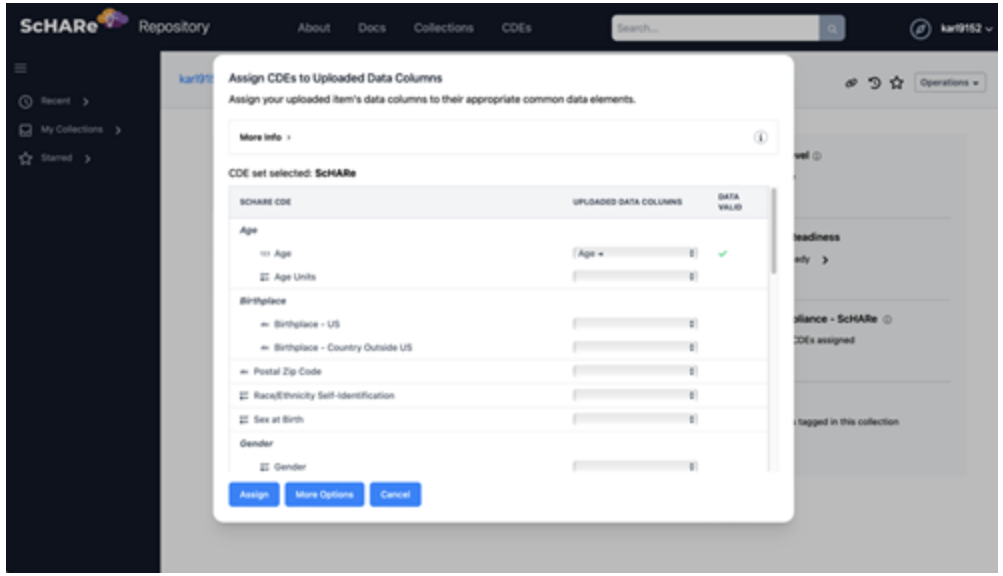
CDE Mapping

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

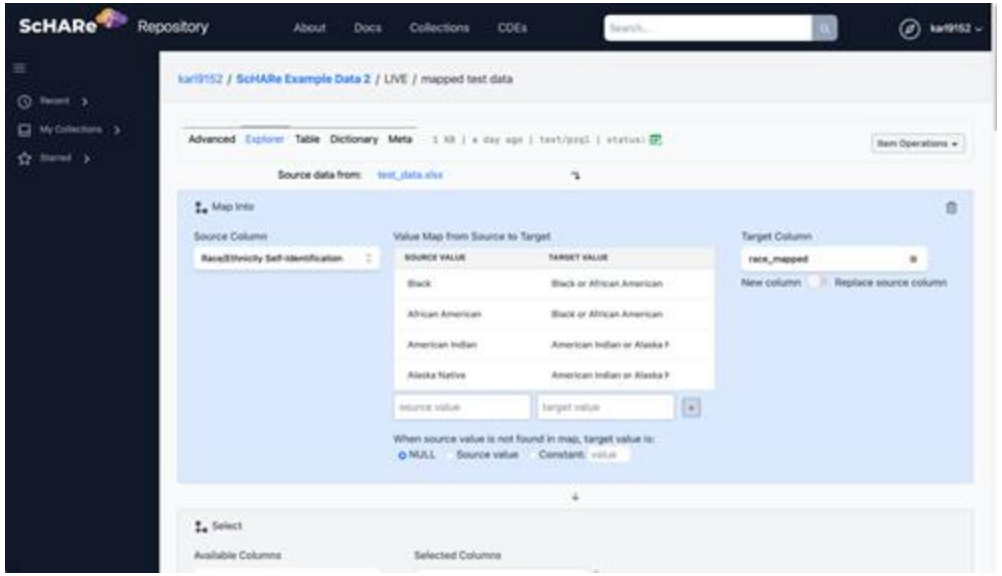
Getting Started with Mapping CDEs

- 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

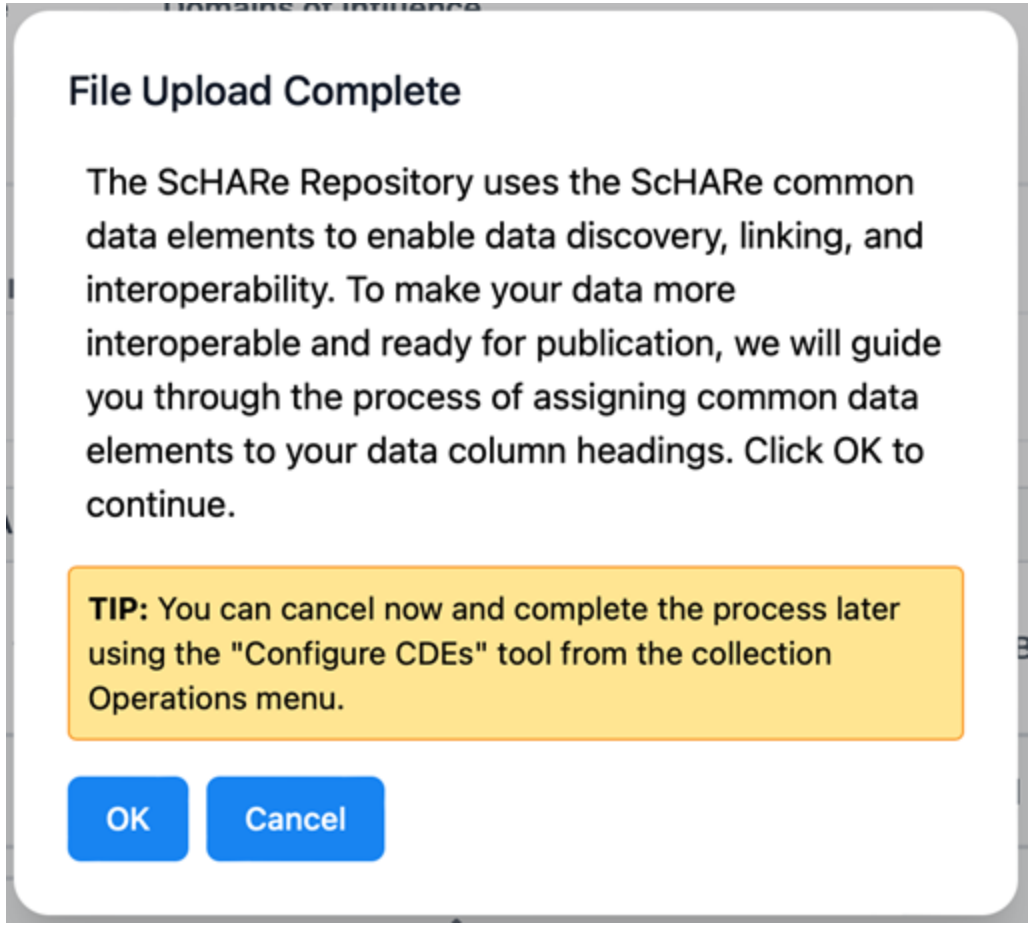
Assign CDEs at Upload



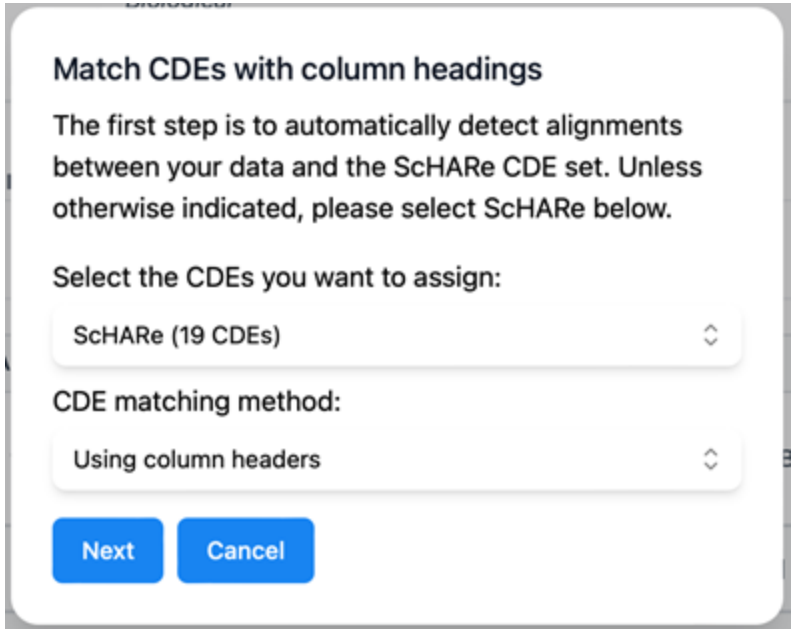
Map CDEs using Dataviews



Assigning CDEs at Upload



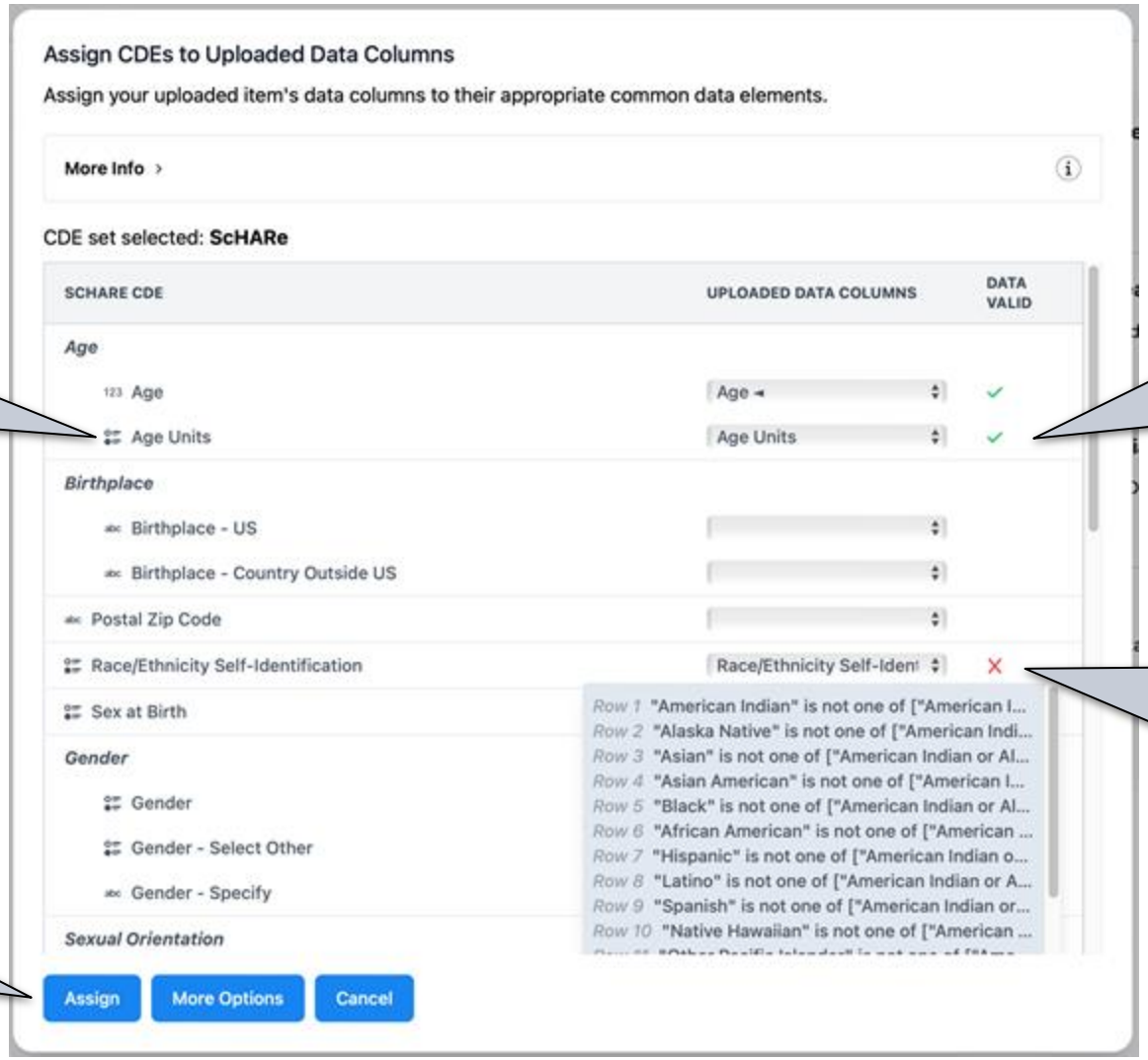
- After you upload a tabular data file, the system will recommend that you begin the process of assigning CDEs to your data column headings.



Assigning CDEs at Upload

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

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



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

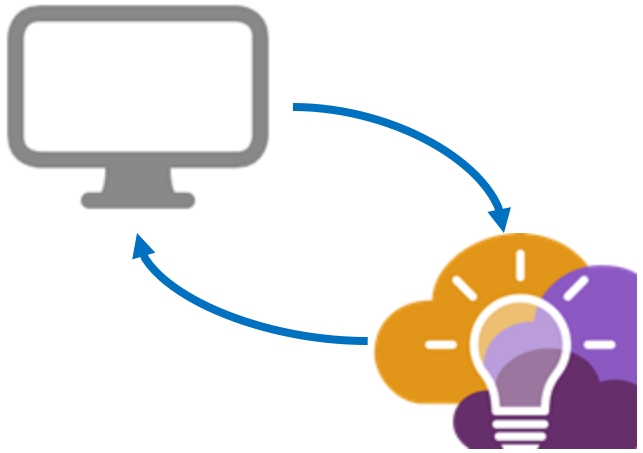
When an assigned CDE fails validation, hover over the red X to get details on why

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

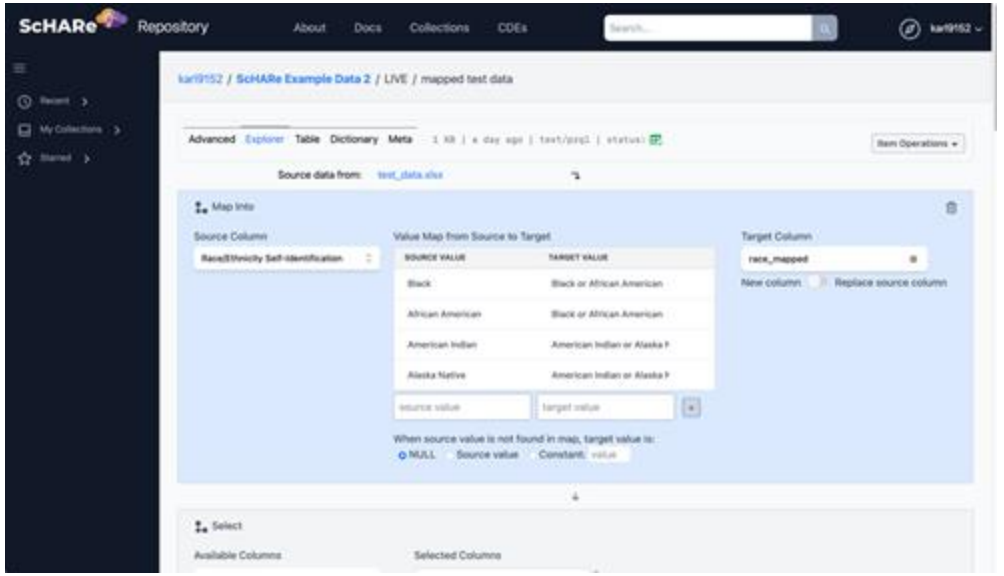
Using Dataviews to Map CDEs

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

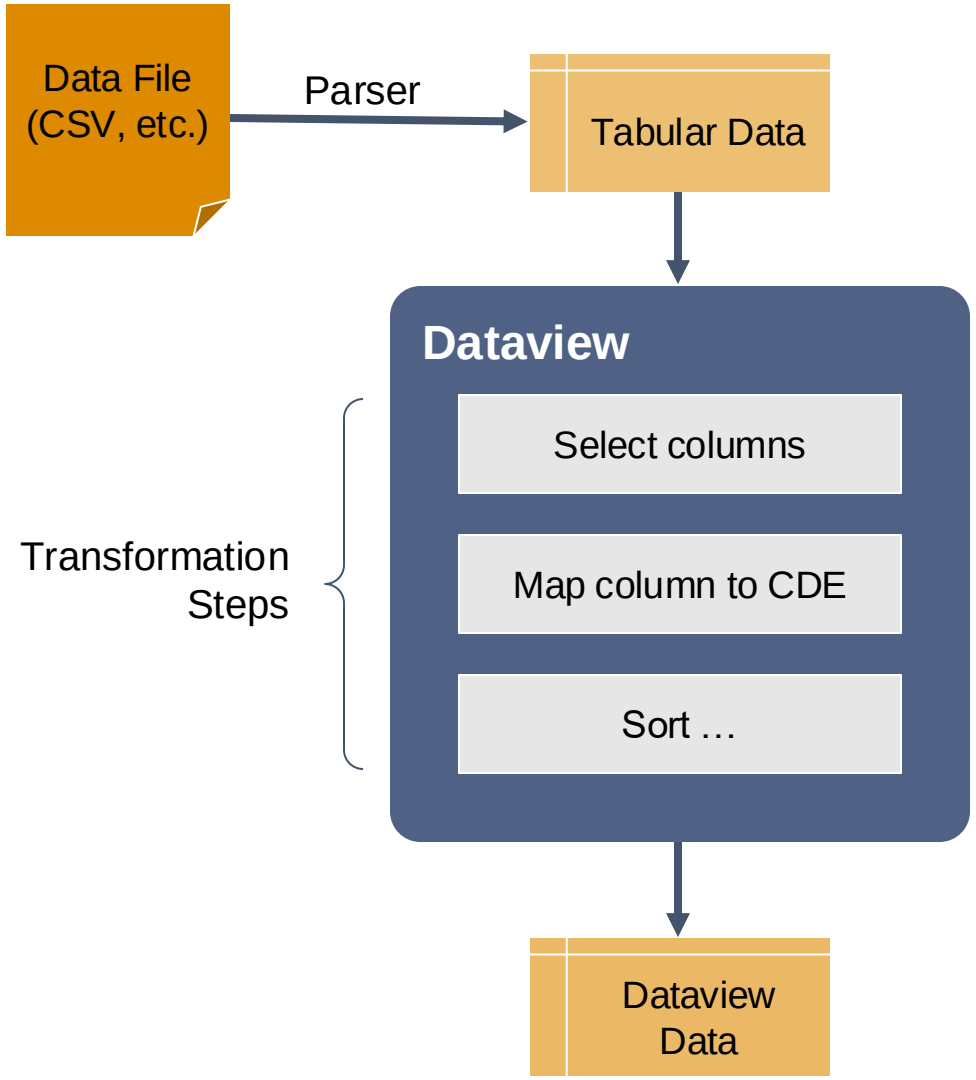
Use your own tools to adjust the data and re-upload



Use a Dataview to Map to CDEs within the Repository



What is a Dataview?

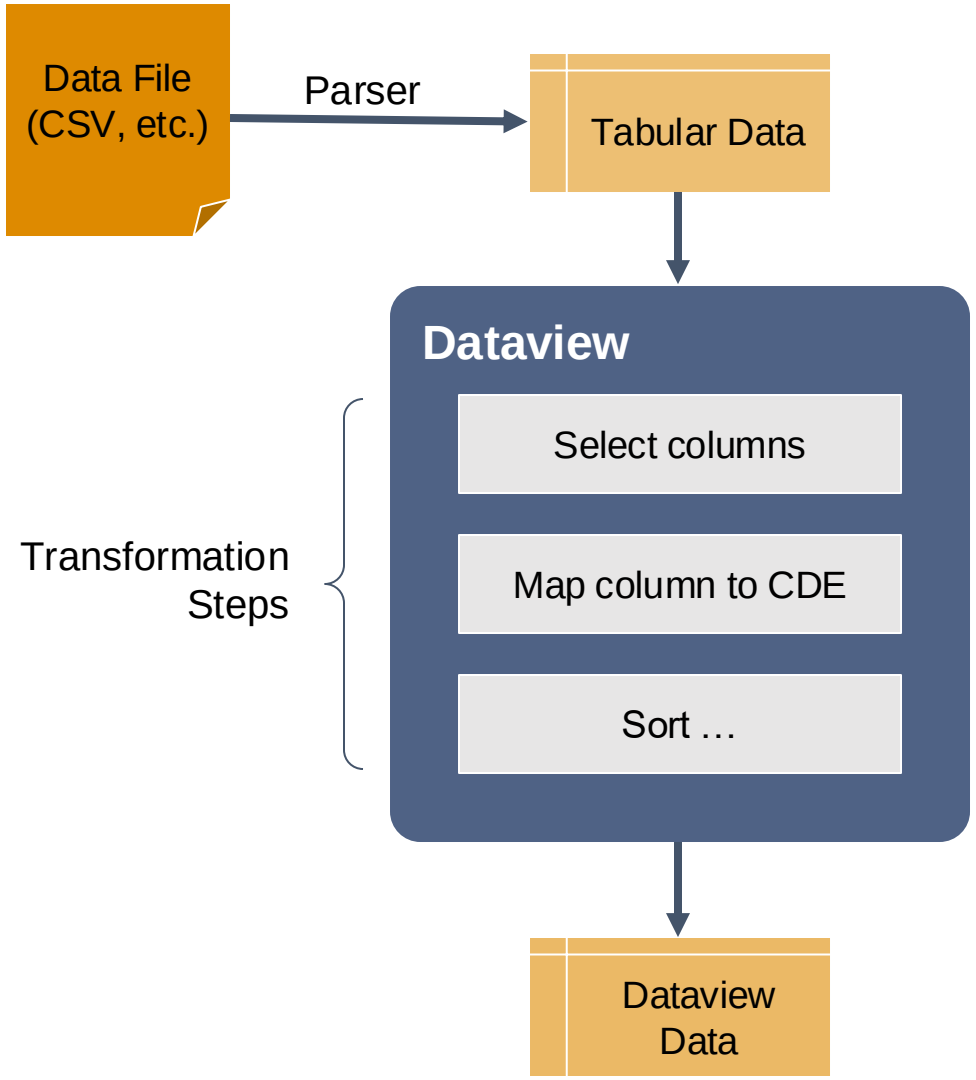


Dataviews take data from one or more sources, apply a series of transformation steps to that data *(filtering, sorting, mapping, etc.)* resulting in a new table of data as output.

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

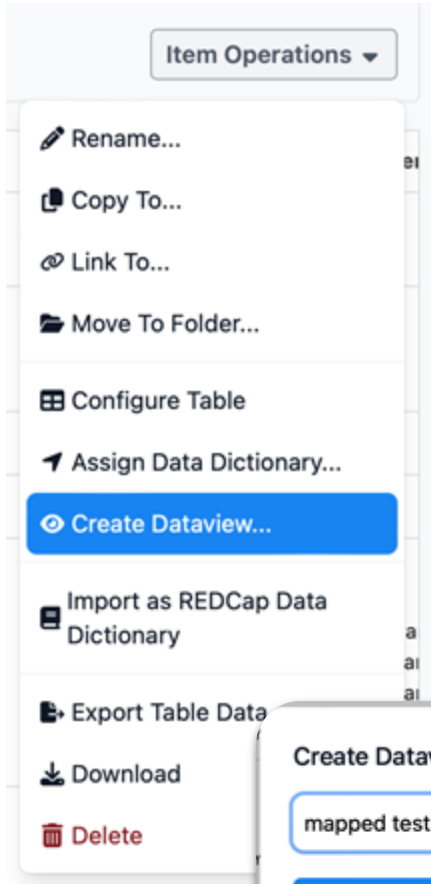


Next, we'll walk through the process together:

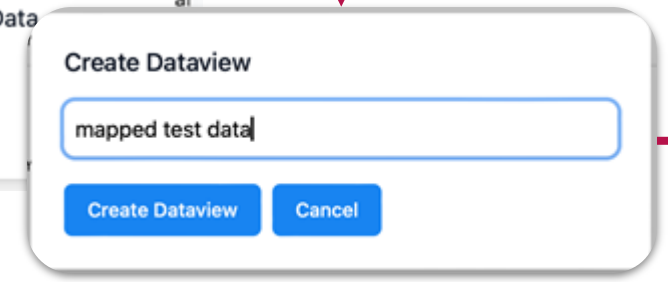
1. Create a new dataview item
2. Add mapping steps to the dataview that map your data values to align with the CDEs
3. Label the newly mapped data columns with the appropriate CDEs and check that it passes validation

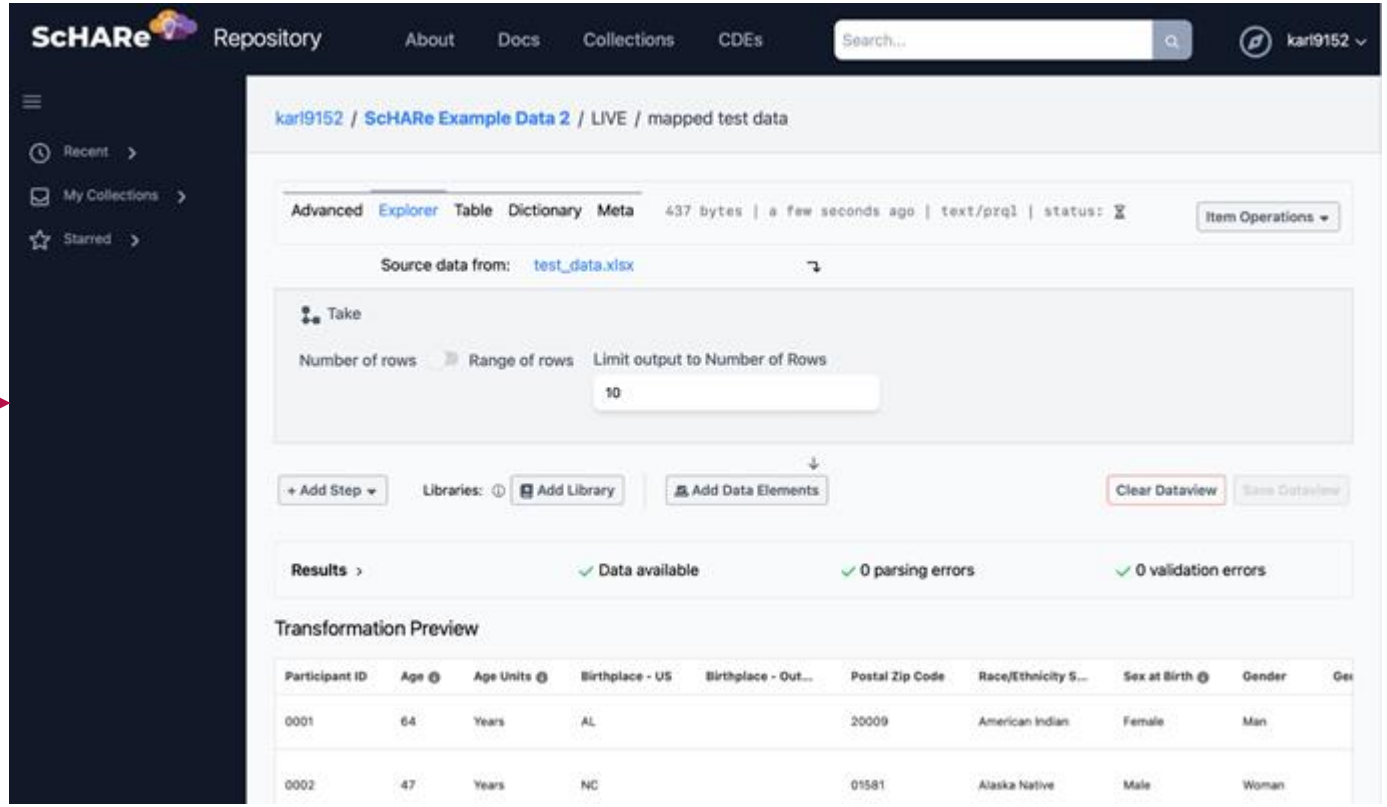
Mapping CDEs via Dataviews

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





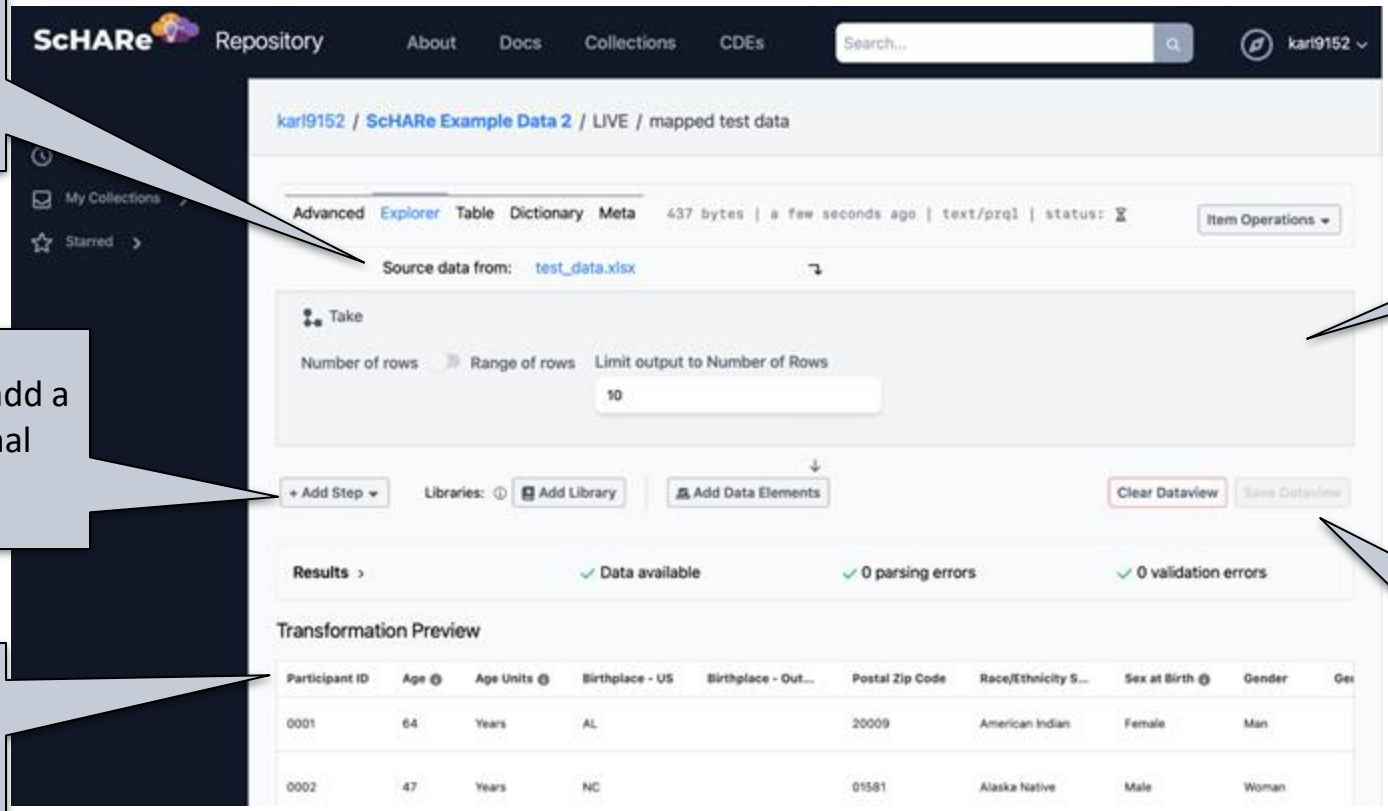
Mapping CDEs via Dataviews

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



Data flows vertically down the pipeline of transformation steps

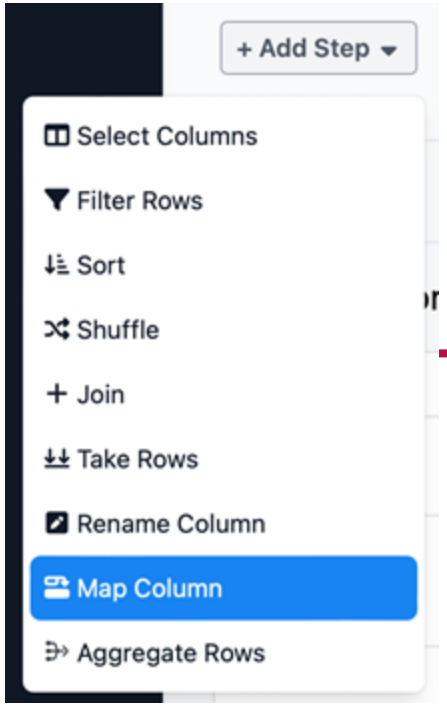
Save your changes using the Save Dataview button

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

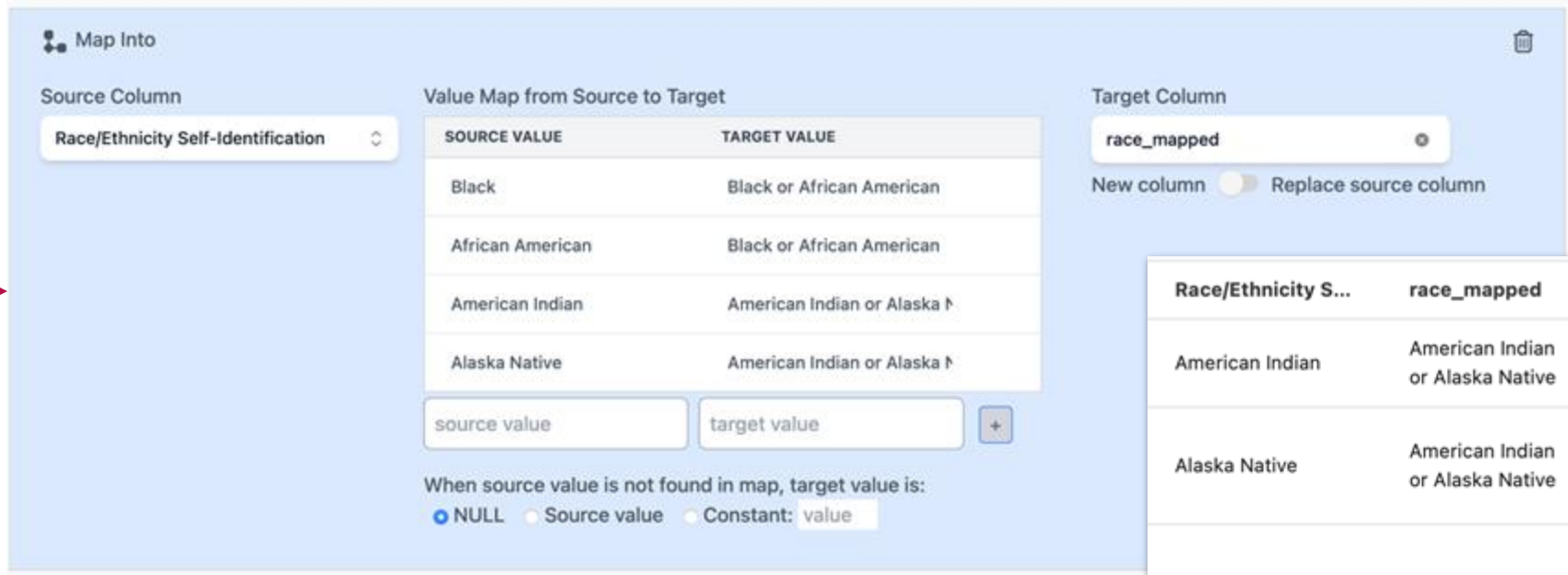
Mapping CDEs via Dataviews

Step 2: Add Mapping Steps to your Dataview

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

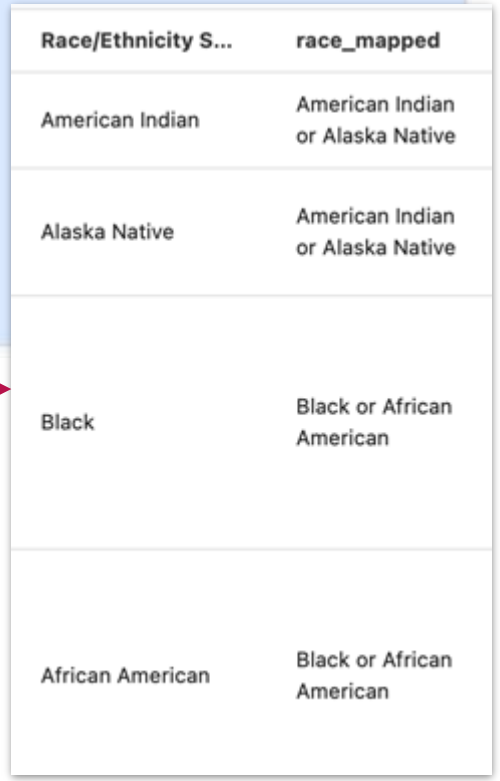


2. Fill in the step fields (source column, target column, value map)



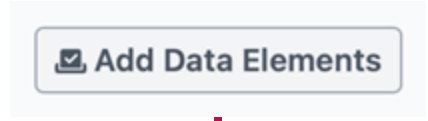
3. Review the results, adjust as needed

4. Repeat for all columns that need to be mapped

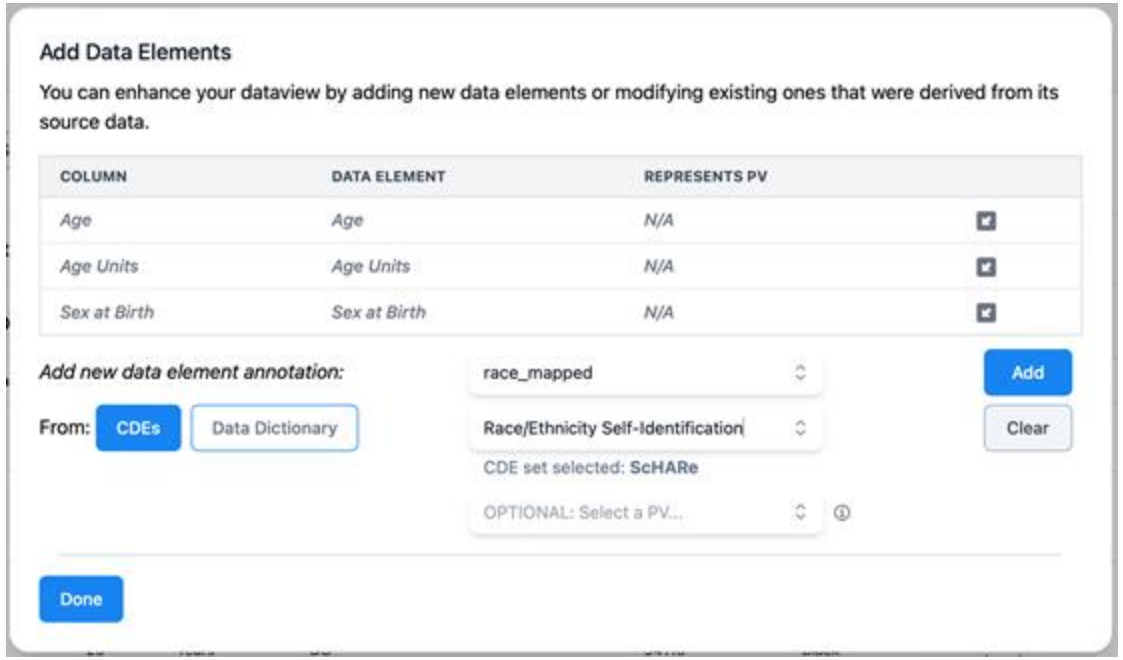


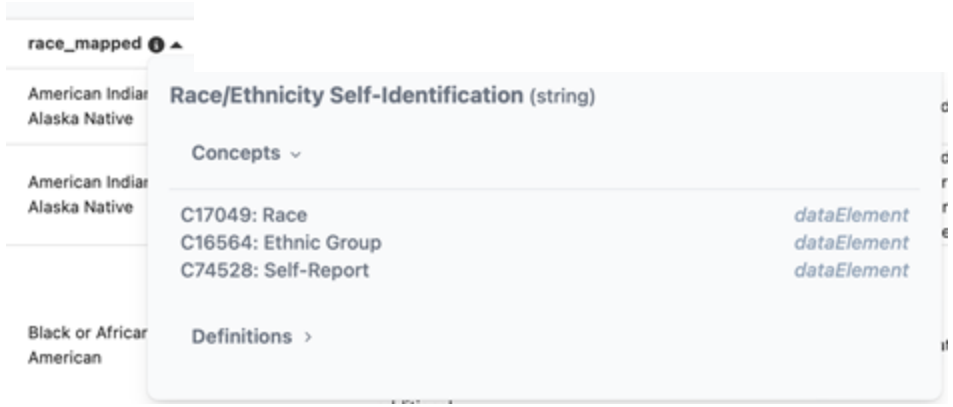
Mapping CDEs via Dataviews

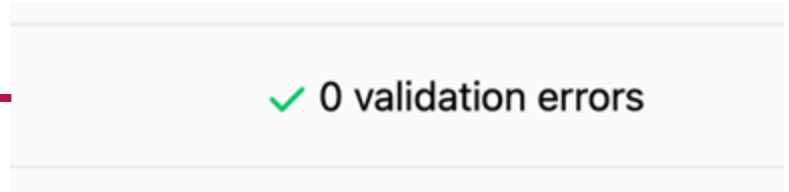
Step 3: Assign CDEs to Data Columns



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



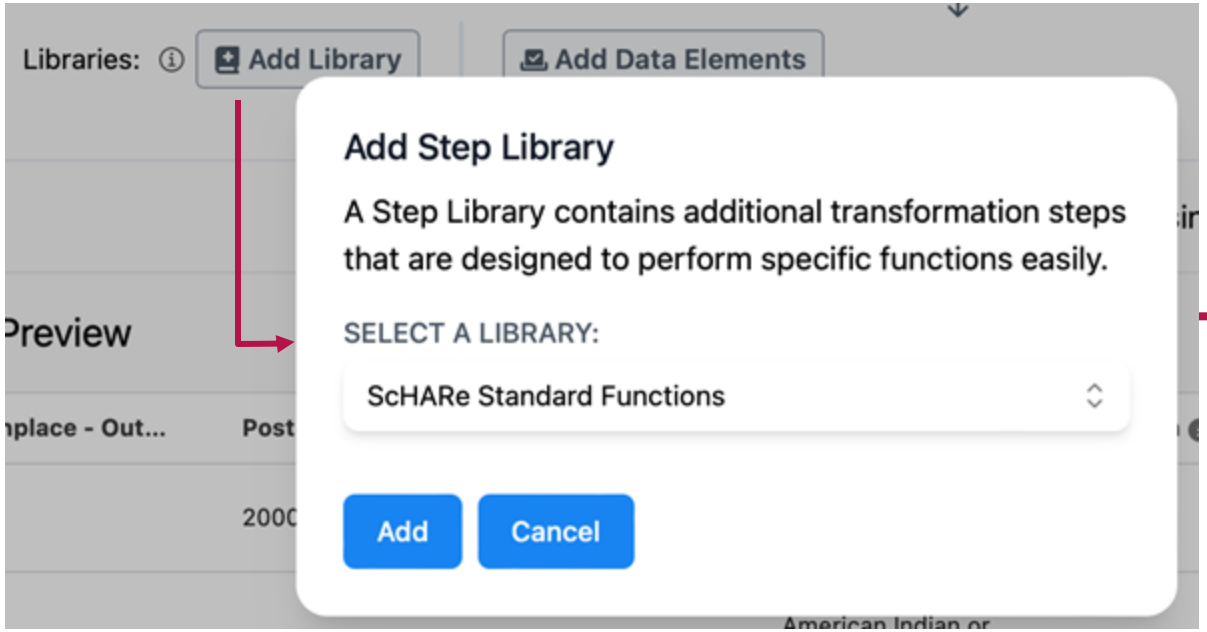


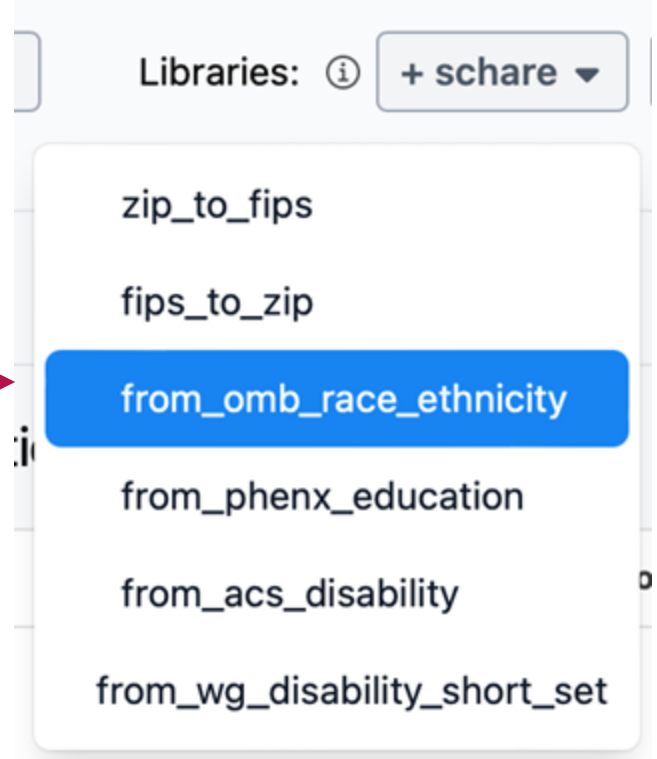


Mapping CDEs via Dataviews

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

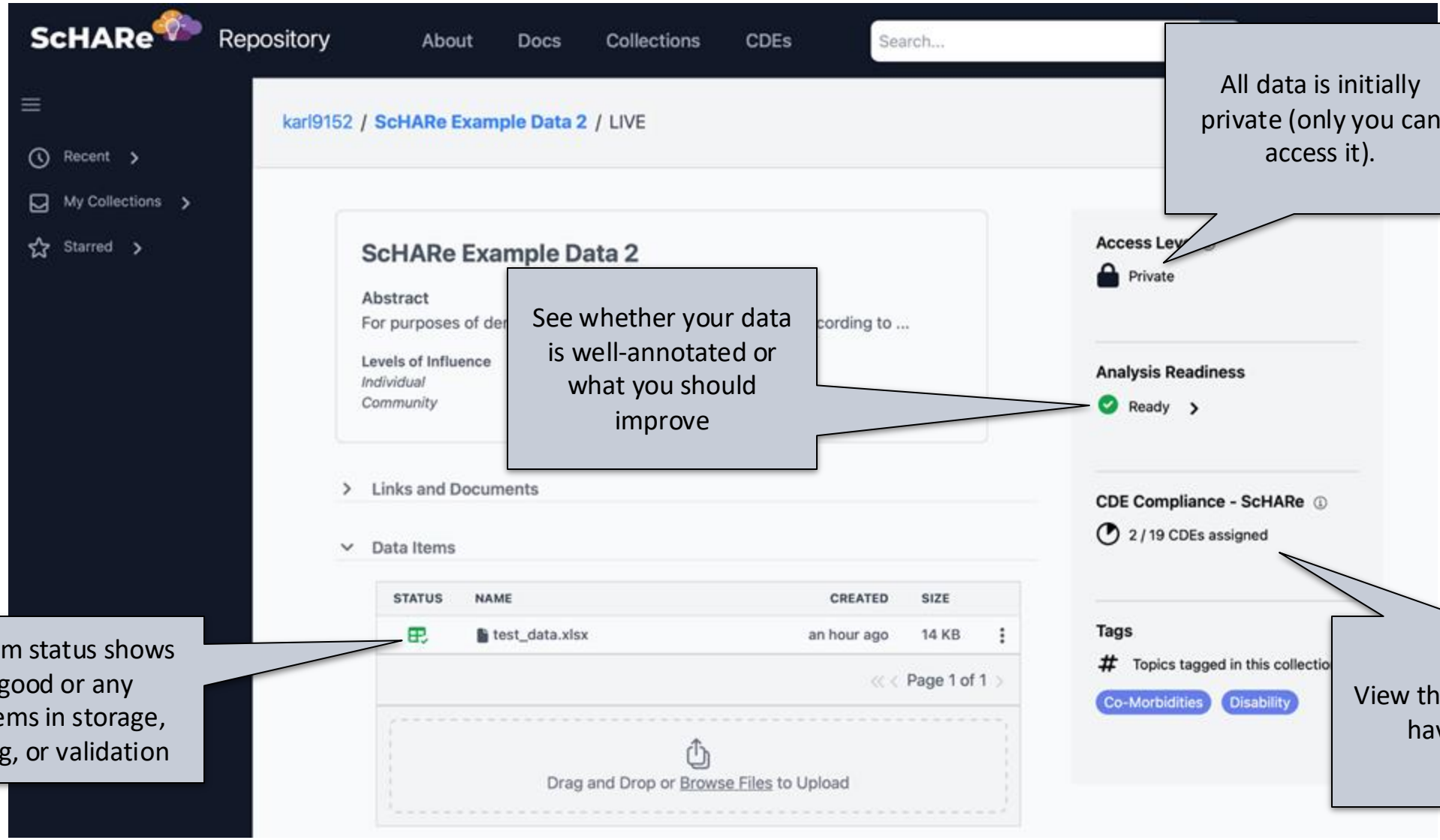
Questions so far?

ScHARe

Sharing Data

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

Viewing your Collection



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

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

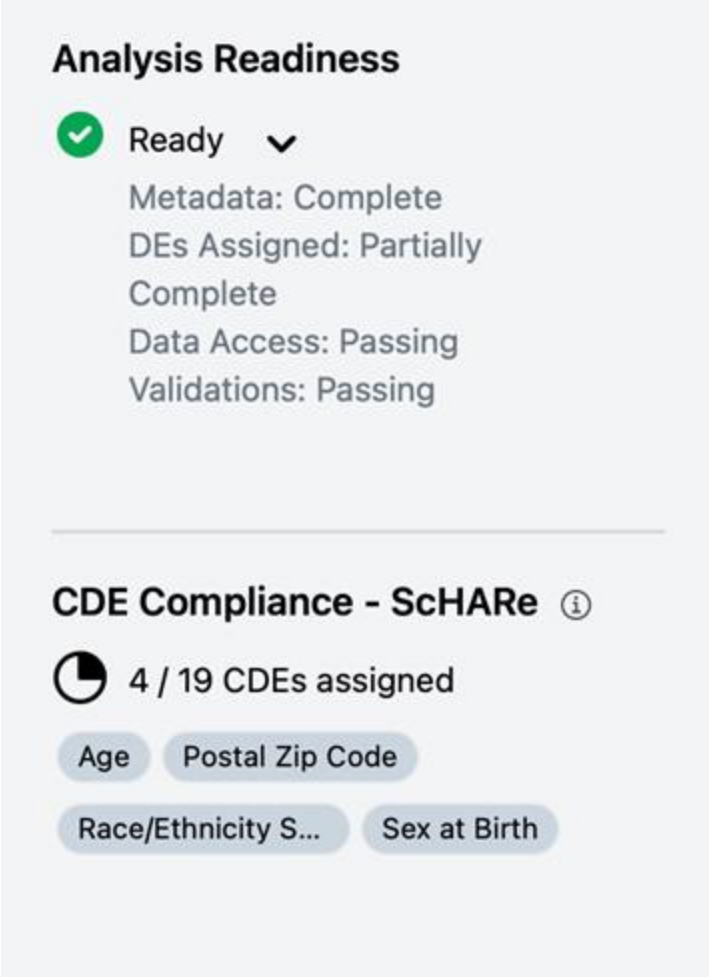
The item status shows all good or any problems in storage, parsing, or validation

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?

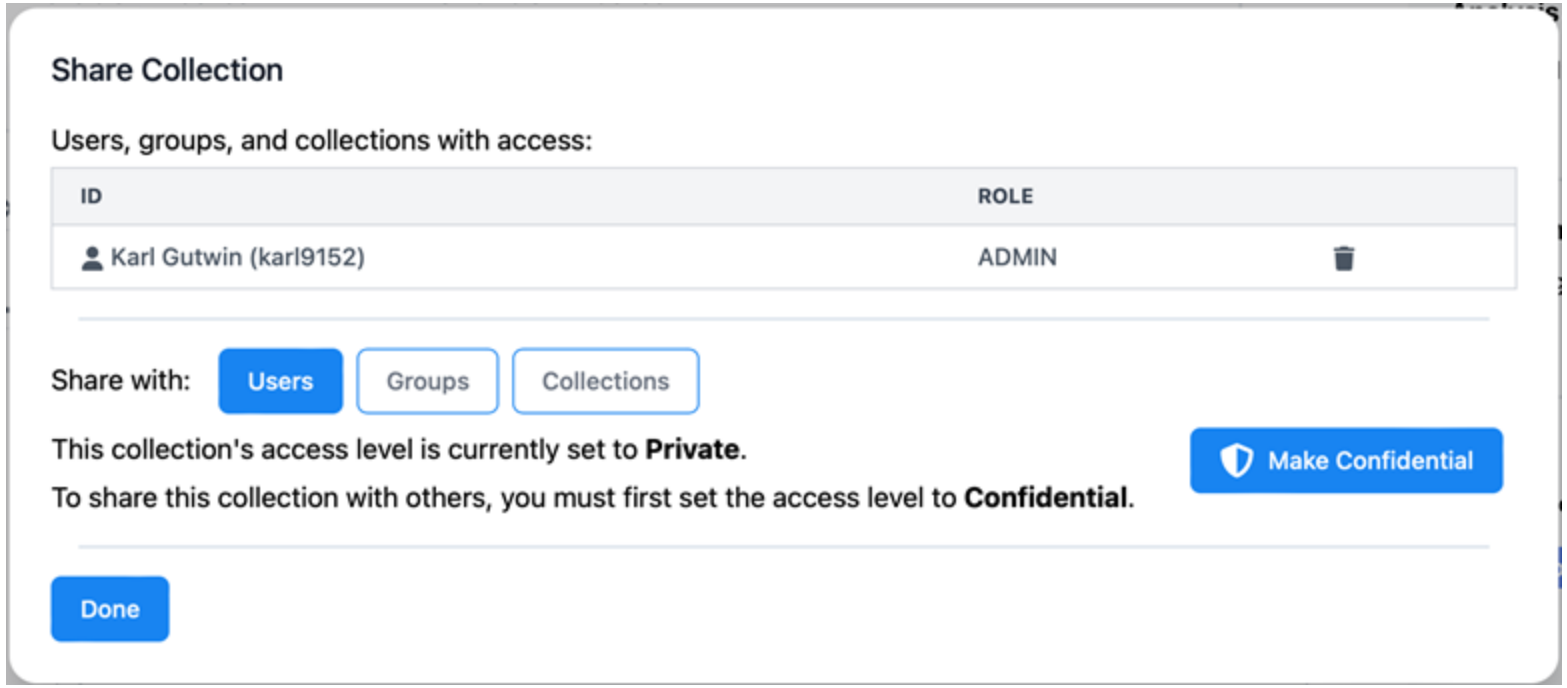


CDE Compliance - Showing how many CDEs have been assigned across the data in the collection

- 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**.



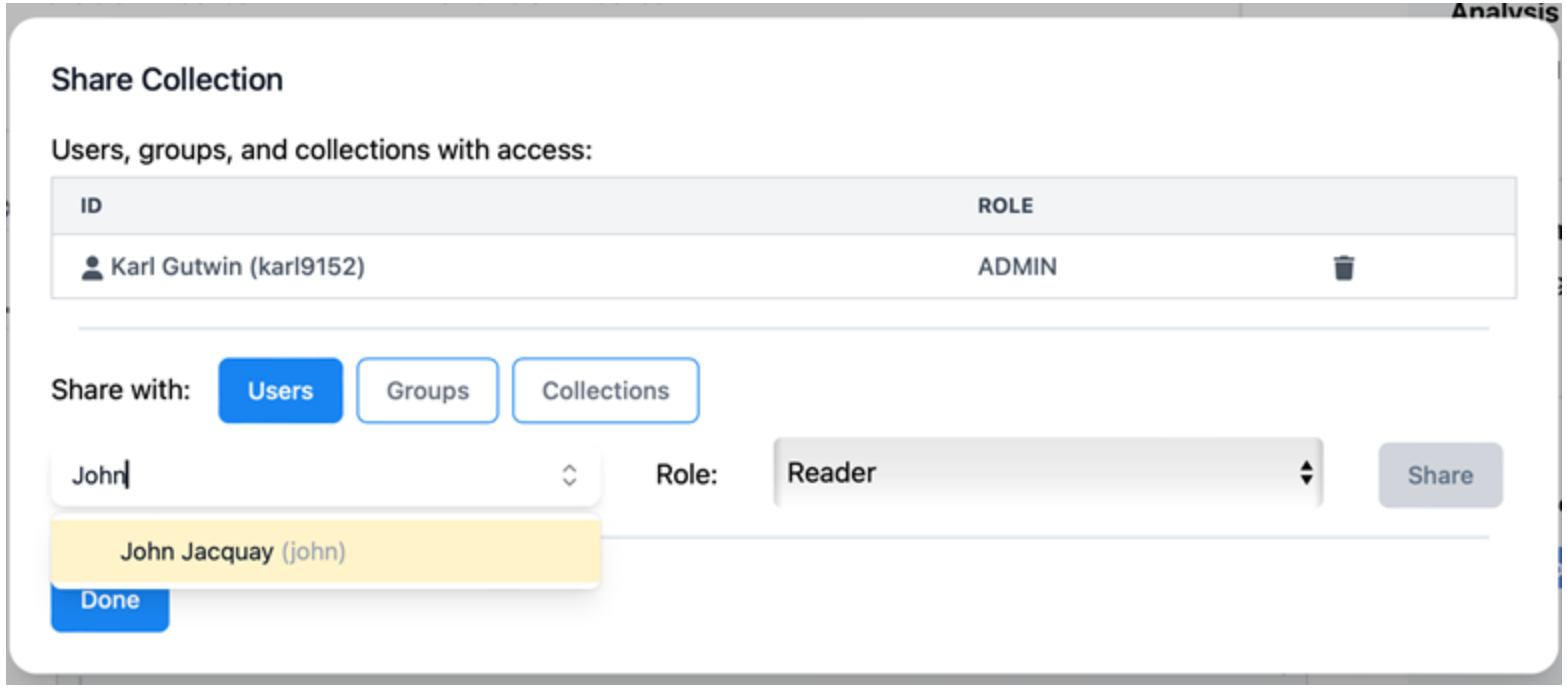
Access Levels

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.



Access Levels

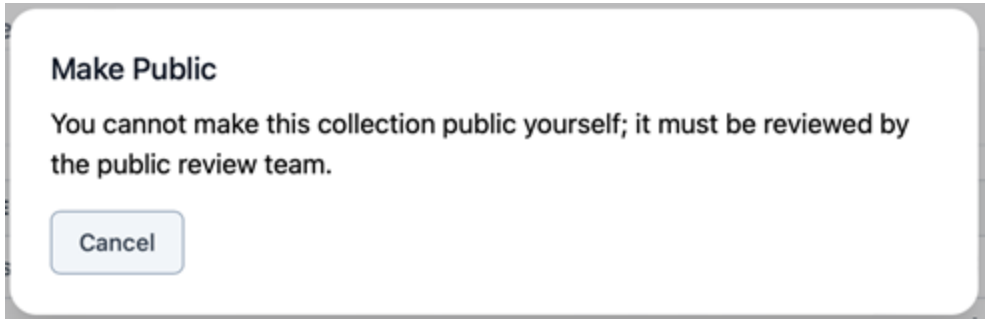
The access level of a collection defines the maximum permissions that can be used to share it with others. The following access levels are supported:

- **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

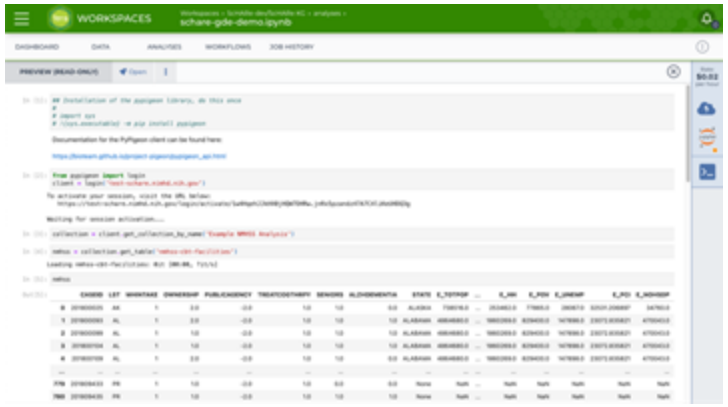
Data Aggregation and Analysis

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

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

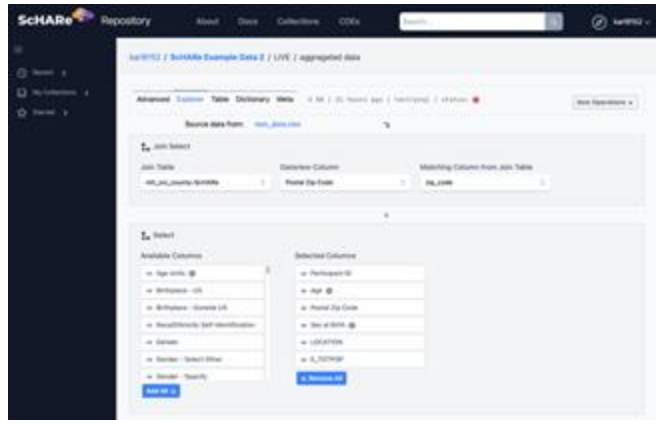
Three quick options:

Connect to a Terra Jupyter notebook



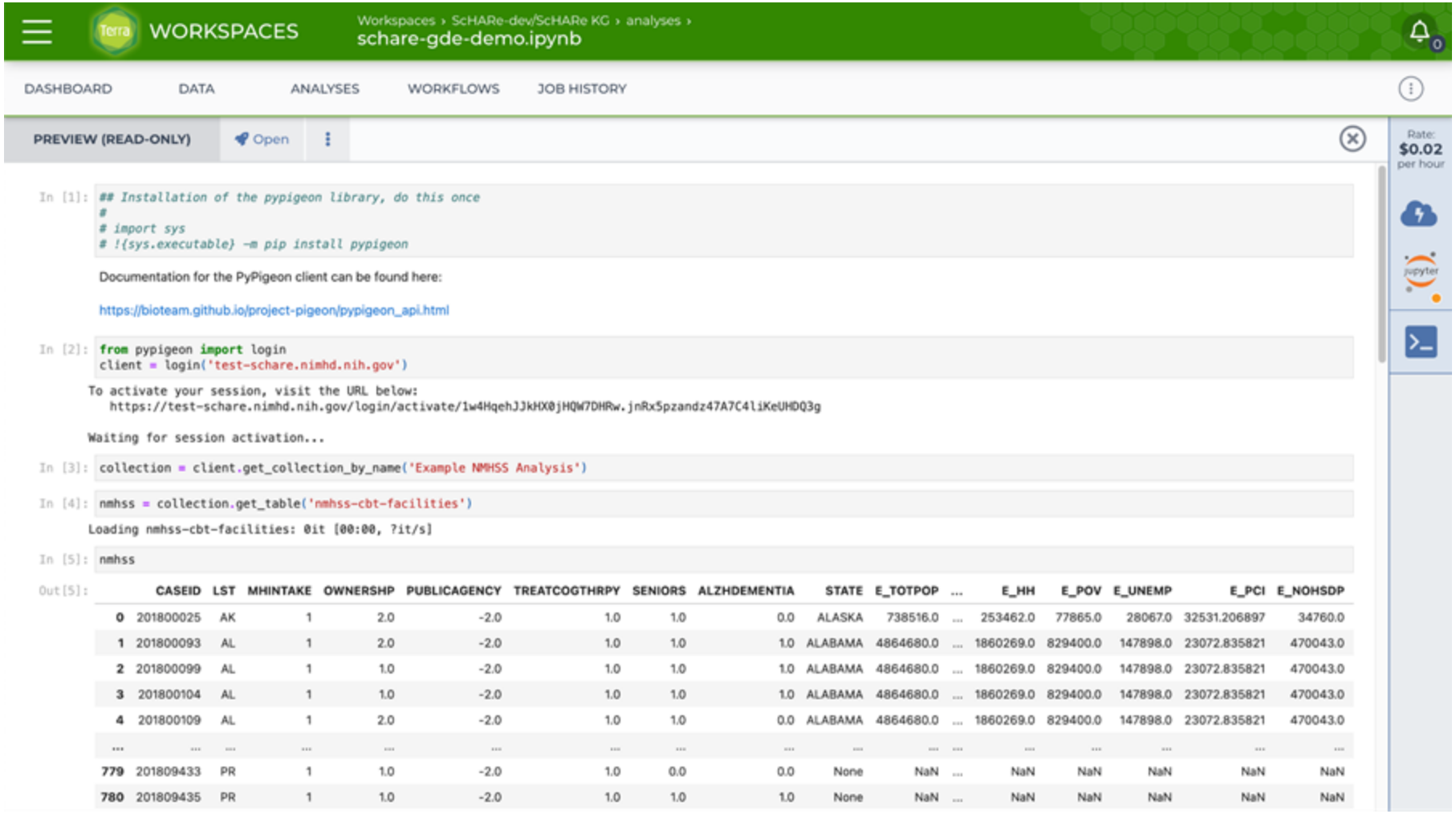
Download to your own computer

Use Dataviews on the Repository



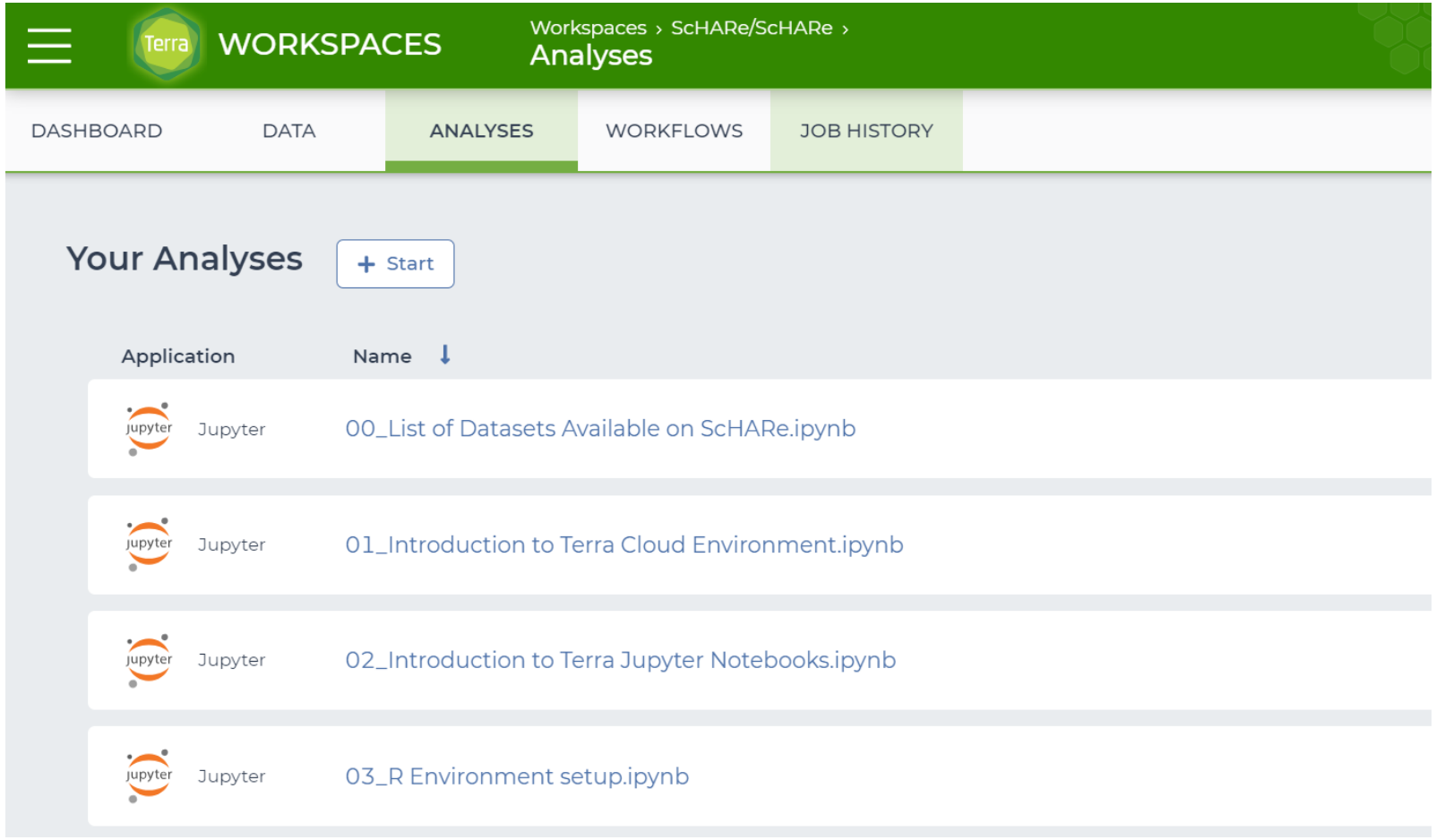
Look for a future Think-a-Thon where we will share more about how to link your data to Terra and how to use the ScHARe Repository to create aggregated data sets!

Analyzing Data on Terra (Jupyter Notebooks)

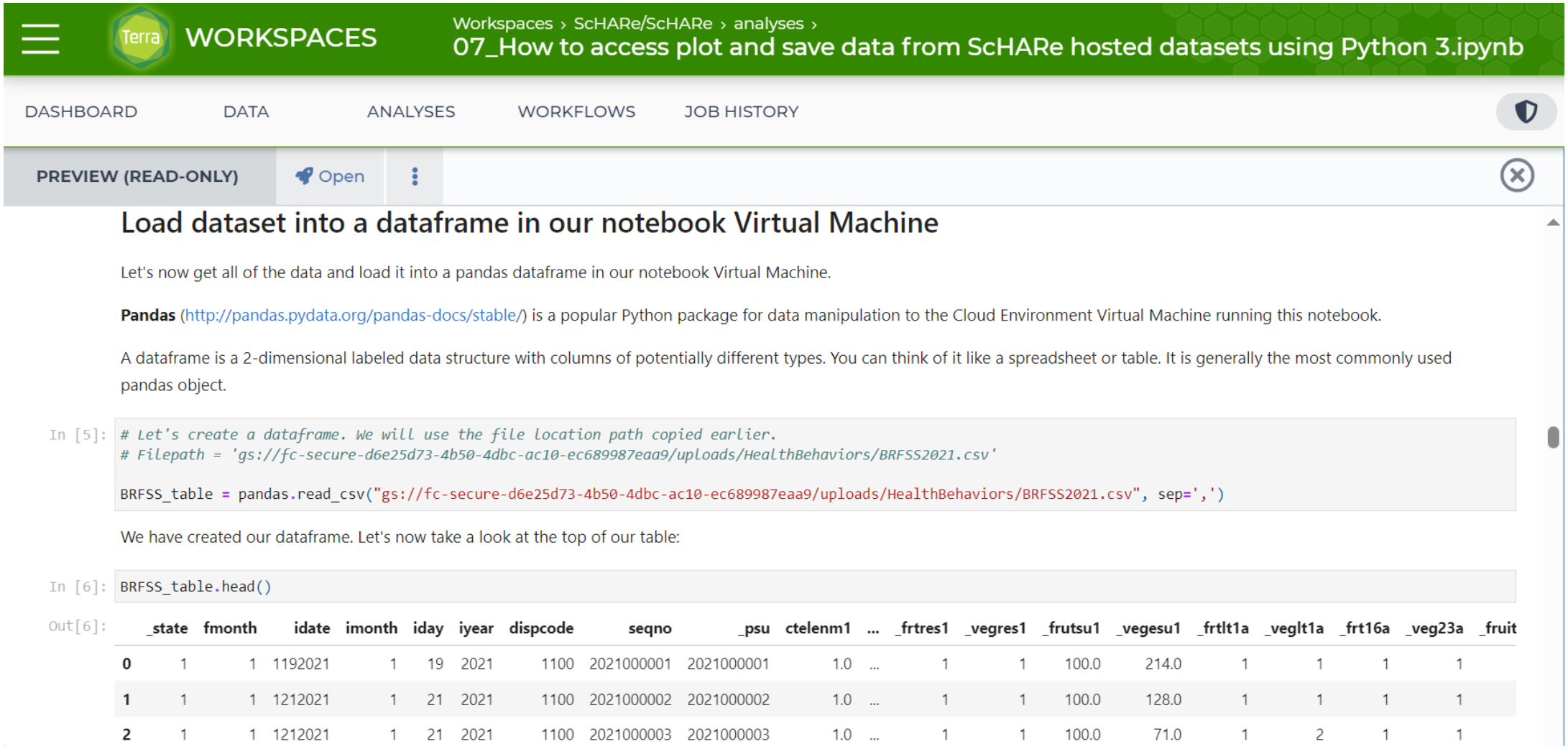


	CASEID	LST	MHINTAKE	OWNERSHP	PUBLICAGENCY	TREATCOGTHRPY	SENIORS	ALZHDEMENTIA	STATE	E_TOTPOP	...	E_HH	E_POV	E_UNEMP	E_PCI	E_NOHSDP
0	201800025	AK	1	2.0	-2.0	1.0	1.0	0.0	ALASKA	738516.0	...	253462.0	77865.0	28067.0	32531.206897	34760.0
1	201800093	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	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)



Analyzing Data on Terra (Jupyter Notebooks)



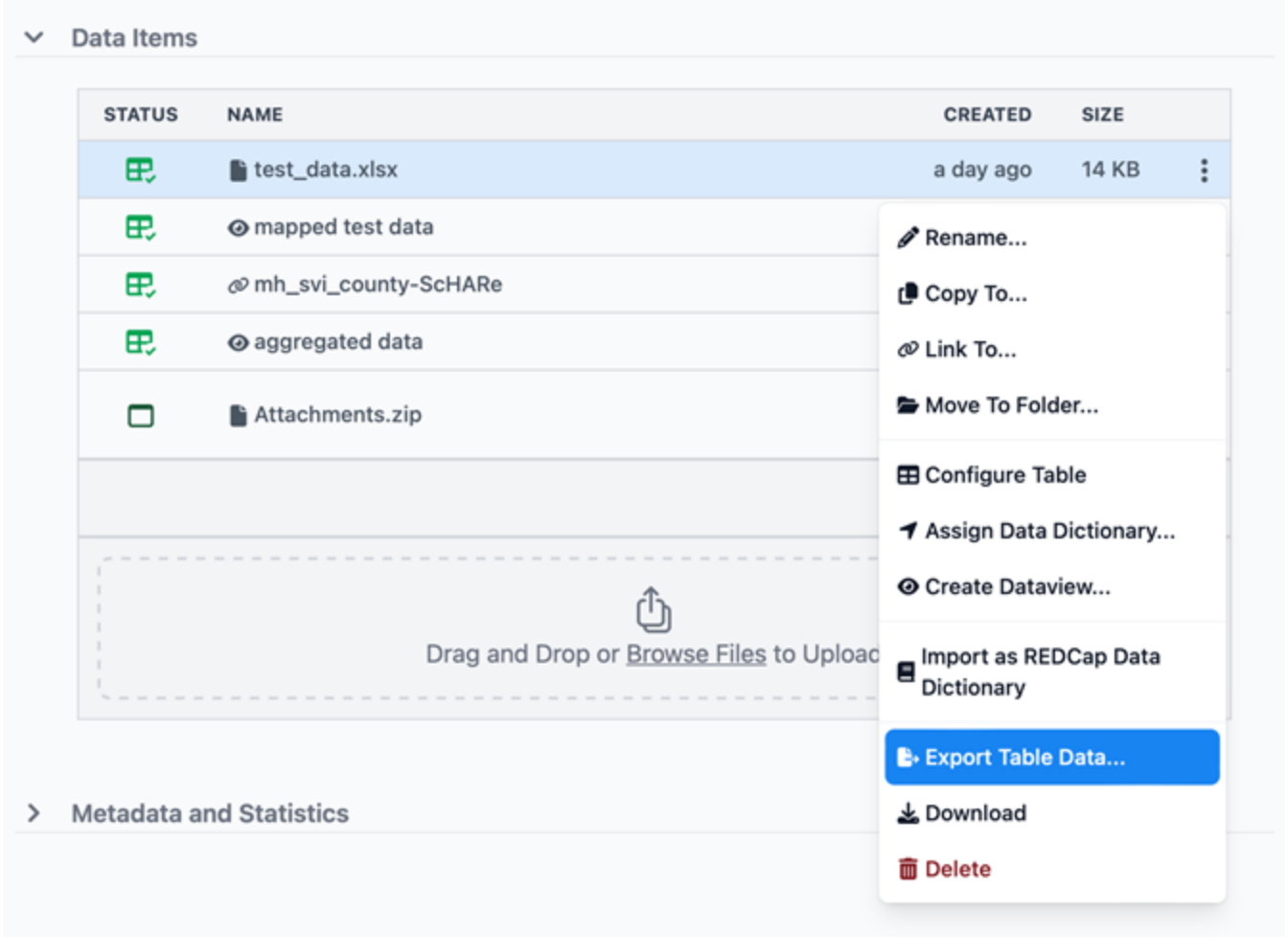
Downloading data to your own computer

Use the item's menu to download the original file

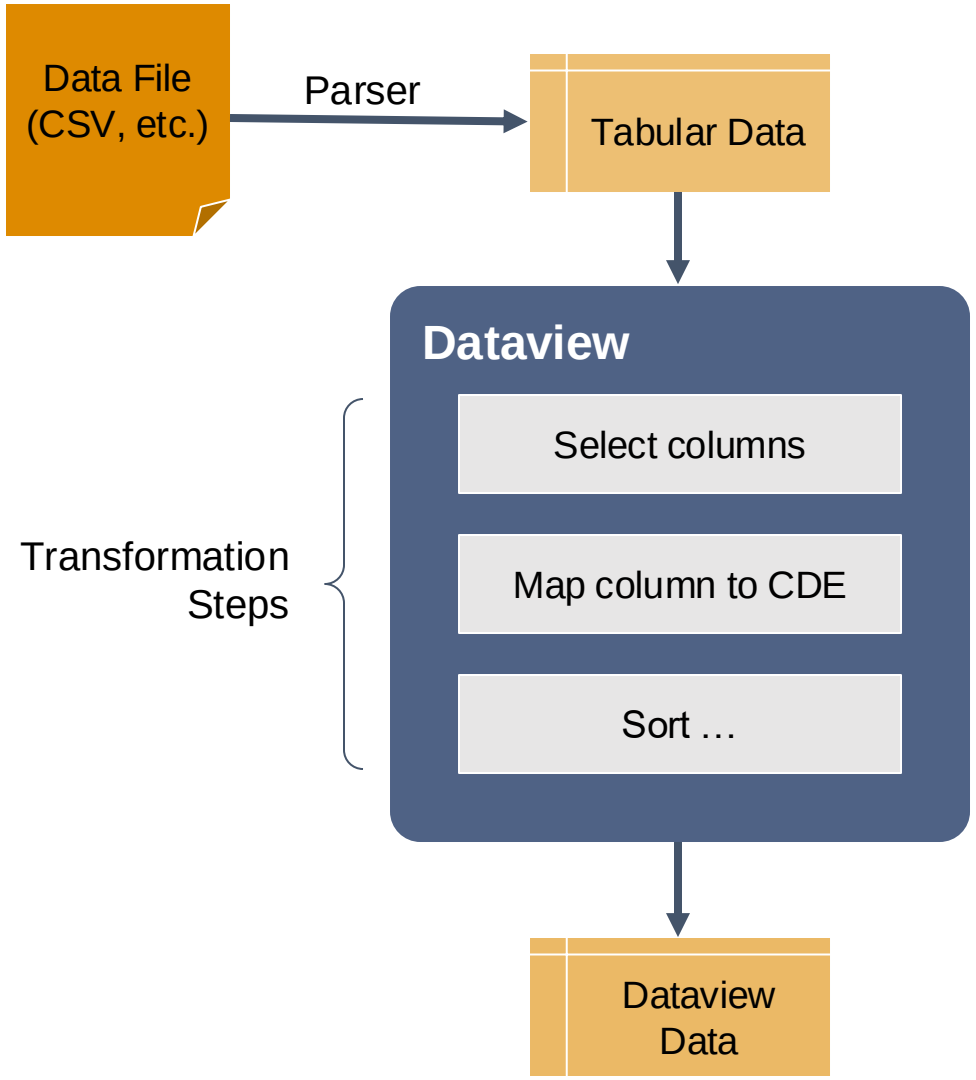
- Best for working directly with raw data

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



Using Dataviews to Aggregate Data

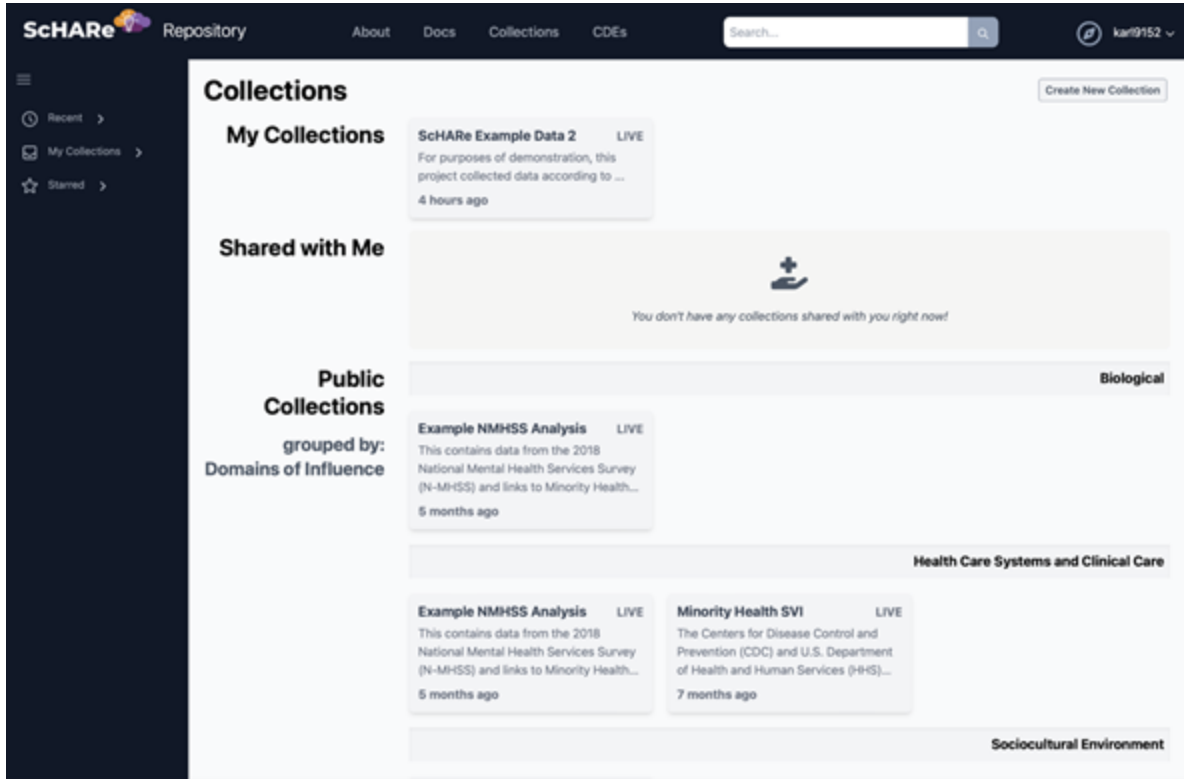


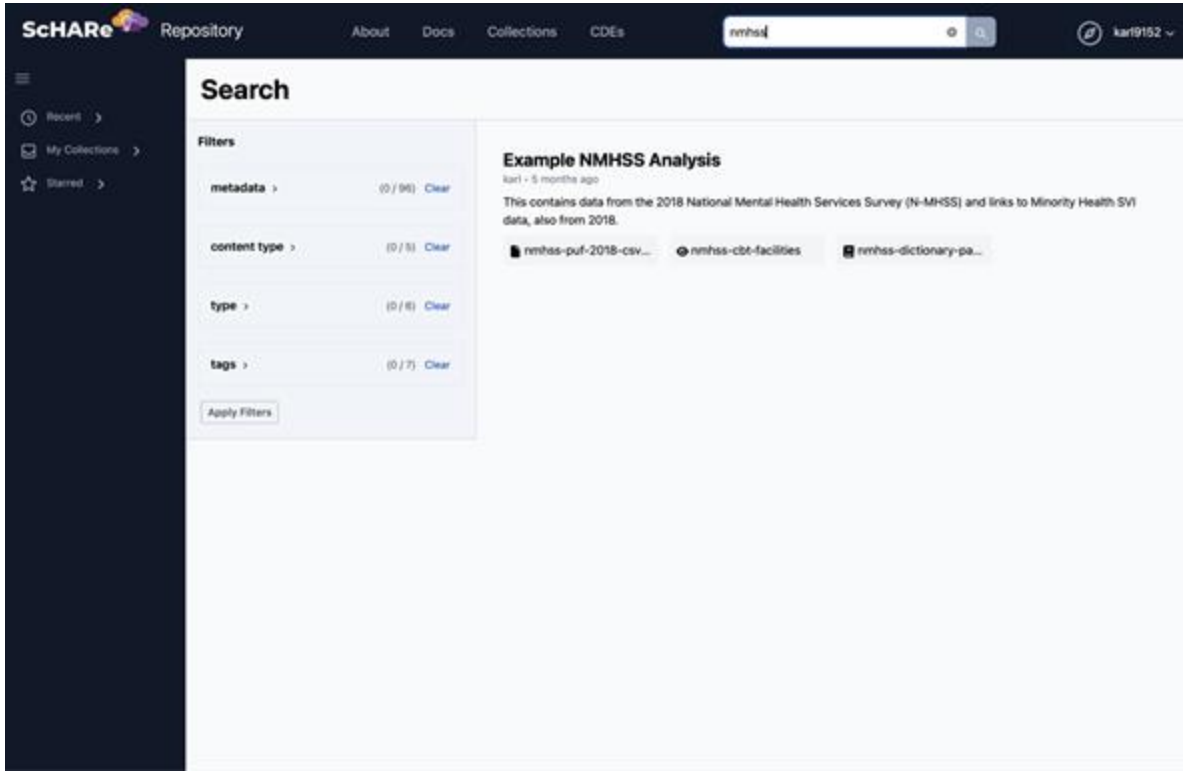
Dataviews take data from one or more sources, apply a series of transformation steps to that data *(filtering, sorting, mapping, etc.)* resulting in a new table of data as output.

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

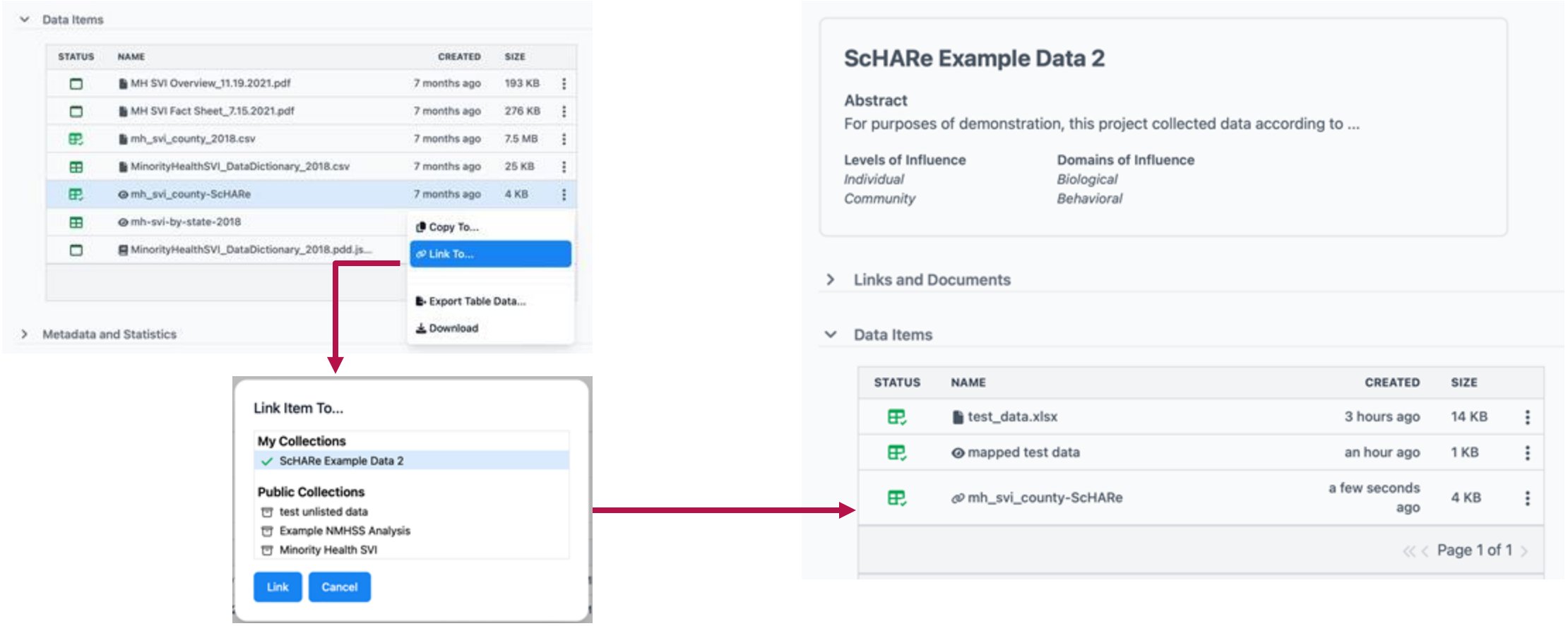




Look for improvements to the data organization and search experience in 2025!

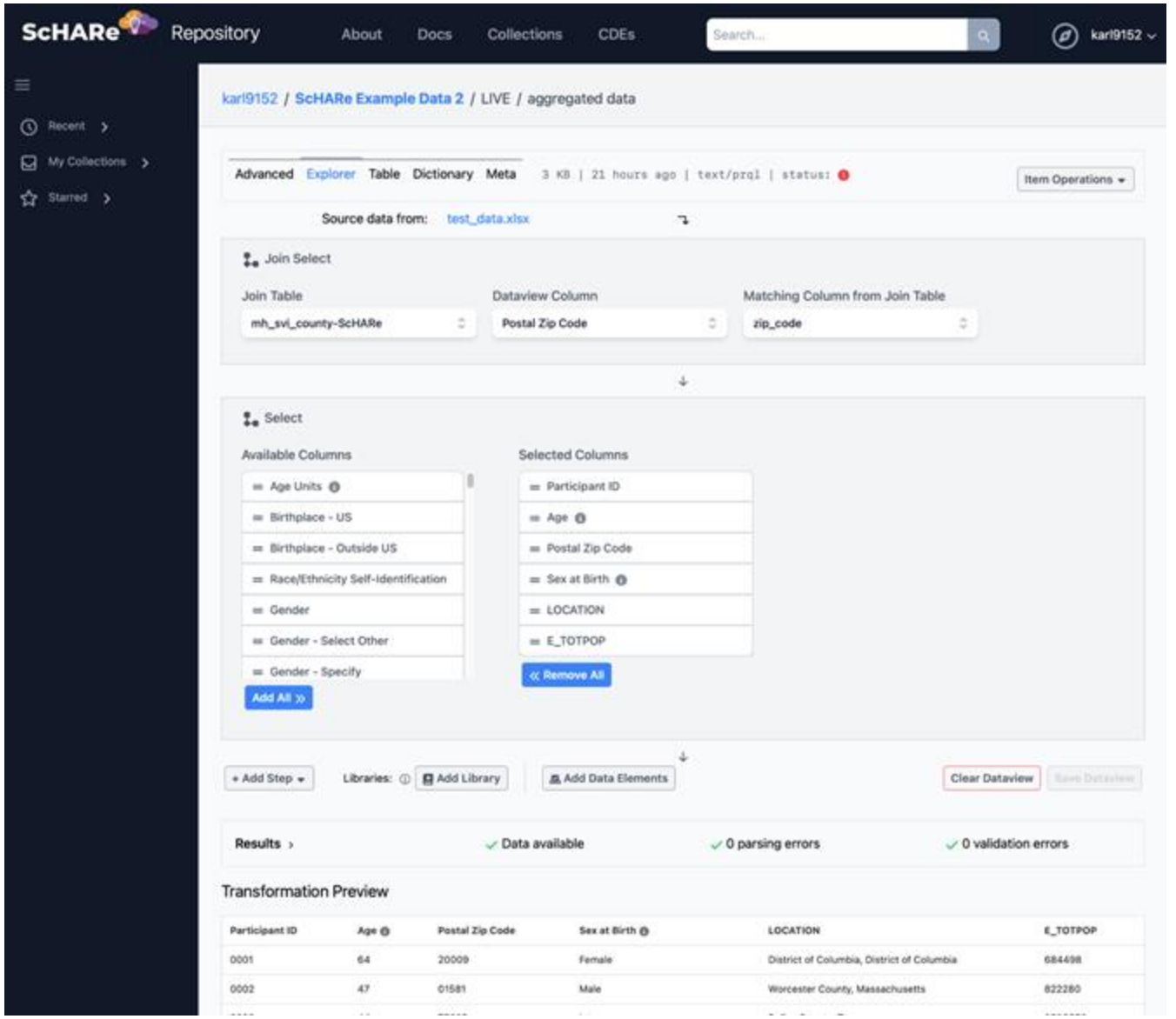
Linking and Aggregating Data

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.



Linking and Aggregating Data

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



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
Response time: within 24 hours

For Technical Support inquiries, contact:
✉ 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

Register for ScHARe:

bit.ly/join-schare

schare@mail.nih.gov