

## Why store data in a repository?

- Easy for others to access and associate with a publication, along with metadata
- Get a digital object identifier (DOI) for your data
- Version control (both “major” and “minor” changes can be tracked)

## My own uses of the Harvard Dataverse so far:

- To store replication datasets for published papers *[example 1 on the next slide]*
- To act as a repository (with a DOI) and manage version control for other datasets *[example 2]*

## Other possible uses:

- To store code for published papers (maybe in addition to replication datasets)
- To act as a repository of source files for other projects (e.g., interactive maps, tools, etc.)

## Example 1: replication data associated with a published article

**Replication Data for: Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems**  
May 14, 2023

Liddie, Jahred; Laurel Schaidler; Elsie Sunderland, 2023, "Replication Data for: Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems", <https://doi.org/10.7910/DVN/0C06MR>, Harvard Dataverse, V1

... This repository holds two replication datasets for the study titled, "Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems." Additional information is available in the readme file for download. ...  
Related Publication Citation: Liddie, J.M., Schaidler, L.S., Sunderland, E.M. Sociodemographic Factors are Associated with Abundance of PFAS Sources and Detection in U.S. Community Water Systems. Environmental Science & Technology (2023) DOI: 10.1021/acs.est.2c07255

**Replication Data for: County-Level Associations between Drinking Water PFAS Contamination and COVID-19 Mortality in the United States**  
Oct 5, 2024

Liddie, Jahred; Bind, Marie-Abèle; Karra, Mahesh; Sunderland, Elsie, 2024, "Replication Data for: County-Level Associations between Drinking Water PFAS Contamination and COVID-19 Mortality in the United States", <https://doi.org/10.7910/DVN/PN0RI5>, Harvard Dataverse, V1

... This repository holds two replication datasets for the study titled, "County-Level Associations between Drinking Water PFAS Contamination and COVID-19 Mortality in the United States." Additional information is available in the readme file for download. ...  
Related Publication Citation: Liddie, J.M., Bind, M.A., Karra, M., Sunderland, E.M. County-Level Associations between Drinking Water PFAS Contamination and COVID-19 Mortality in the United States. Journal of Exposure Science & Environmental Epidemiology (2024) DOI: 10.1038/s41370-024-00723-5

### Replication Data for: Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems

Version 1.0

Liddie, Jahred; Laurel Schaidler; Elsie Sunderland, 2023, "Replication Data for: Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems", <https://doi.org/10.7910/DVN/0C06MR>, Harvard Dataverse, V1

Cite Dataset ▾ Learn about [Data Citation Standards](#).

**Access Dataset ▾**  
Contact Owner Share

Dataset Metrics ⓘ  
78 Downloads ⓘ

### Related article

**Description ⓘ**  
This repository holds two replication datasets for the study titled, "Sociodemographic Factors are Associated with the Abundance of PFAS Sources and Detection in U.S. Community Water Systems." Additional information is available in the readme file for download. (2023-04-22)

**Subject ⓘ**  
Earth and Environmental Sciences; Medicine, Health and Life Sciences

**Keyword ⓘ**  
water quality, environmental justice, drinking water, disparities, pollution











**Related Publication ⓘ**  
Liddie, J.M., Schaidler, L.S., Sunderland, E.M. Sociodemographic Factors are Associated with Abundance of PFAS Sources and Detection in U.S. Community Water Systems. Environmental Science & Technology (2023) DOI: 10.1021/acs.est.2c07255doi: 10.1021/acs.est.2c07255

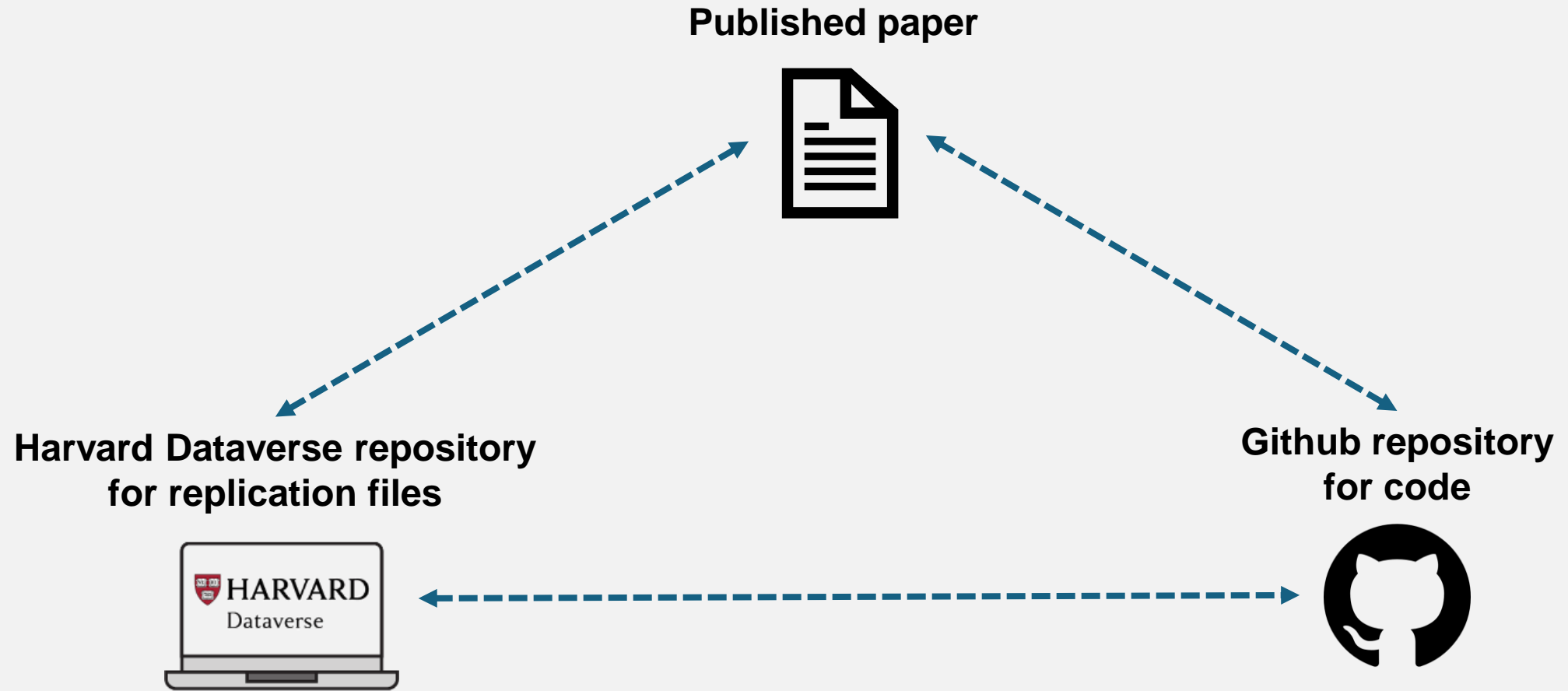
**Replication data (i.e., processed datasets that can be used to replicate the results)**

**A "codebook" describing the variables in the replication datasets**

**A "readme" file with the appropriate citation, description of each file, and contact information**

1 to 4 of 4 Files Download

	<a href="#">allCWS_replication.csv</a> Comma Separated Values - 2.6 MB Published May 14, 2023 22 Downloads MD5: b93...998	 
	<a href="#">HUC_replication.geojson</a> GeoJSON - 21.1 MB Published May 14, 2023 17 Downloads MD5: 3ee...51a	 
	<a href="#">replication_data_codebook-1.xlsx</a> MS Excel Spreadsheet - 14.2 KB Published May 14, 2023 21 Downloads MD5: c61...141	
	<a href="#">replication_readme.tex</a> LaTeX - 1.2 KB Published May 14, 2023 18 Downloads MD5: a8f...9ad	



## Example 2: repository for an (ongoing) data compilation effort

**PFAS Statewide Sampling Dataset** Published Curator

Mar 11, 2025 - Harvard Dataverse

Liddie, Jahred, 2022, "PFAS Statewide Sampling Dataset", <https://doi.org/10.7910/DVN/8LPLCF>, Harvard Dataverse, V4

This repository contains a sample of U.S. community water systems (CWS) included in the PFAS statewide sampling dataset. In the first versions of this dataset, 18 states were included as part of Liddie et al., 2023. The statewide sampling dataset now...

### PFAS Statewide Sampling Dataset

**Version 4.1**

Liddie, Jahred, 2022, "PFAS Statewide Sampling Dataset", <https://doi.org/10.7910/DVN/8LPLCF>, Harvard Dataverse, V4

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

**Version control along with citation**

**Description**

This repository contains a sample of U.S. community water systems (CWS) included in the PFAS statewide sampling dataset. In the first versions of this dataset, 18 states were included as part of Liddie et al., 2023. The statewide sampling dataset now includes compiled data from 28 U.S. statewide sampling campaigns of CWS for per- and polyfluoroalkyl substances (PFAS) in drinking water. The full dataset of samples spans the 2016-2024 period. All data were either publicly available or provided by state agencies upon request. The full dataset contains approximately 98,000 samples from 11,138 CWS in Alabama, Arizona, California, Colorado, Georgia, Illinois, Indiana, Iowa, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Utah, Vermont, Virginia, Washington, and Wisconsin. Concentrations

[Access Dataset](#)

[Edit Dataset](#)

[Link Dataset](#)

[Contact Owner](#) [Share](#)






**Dataset Metrics**

294 Downloads

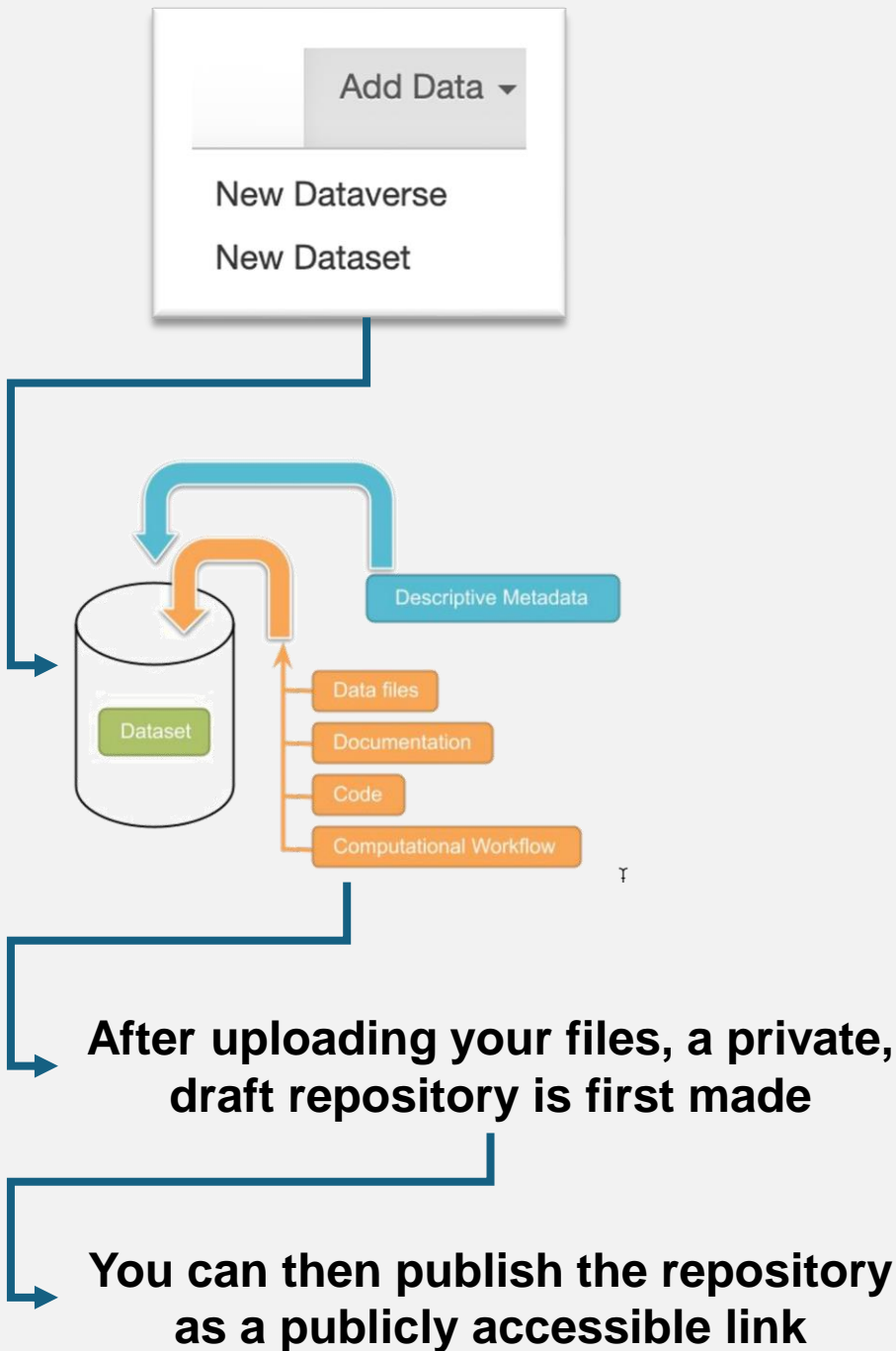
Readme file and a bibliography of data sources

Example representative datasets

Codebook

1 to 5 of 5 Files	
	<a href="#">readme.rtf</a> Rich Text Format - 5.0 KB Published Mar 11, 2025 0 Downloads MD5: 2fe...44f readme file
	<a href="#">statewide_sampling_data_sources_03112025.rtf</a> Rich Text Format - 16.0 KB Published Mar 11, 2025 0 Downloads MD5: 198...972 Data source links
	<a href="#">statewide_dataset_subset_samples_V4.csv</a> Comma Separated Values - 1.3 MB Published Mar 11, 2025 0 Downloads MD5: abe...ae5
	<a href="#">statewide_dataset_subset_V4.csv</a> Comma Separated Values - 165.5 KB Published Mar 11, 2025 0 Downloads MD5: 434...3b3
	<a href="#">statewide_sampling_dataset_codebook_03112025.xlsx</a> MS Excel Spreadsheet - 19.7 KB Published Mar 11, 2025 0 Downloads MD5: 143...0fa Codebook

Have not used them, but there are also ways to restrict access / require data requests on HD



Much more can be found in the Harvard Dataverse [user guide](#) and [prior presentations](#)

### GATHER AND VERIFY

Determine **what will be shared** and **ensure the metadata describes what is being shared**: raw data, output data, documentation, readme, codebook, etc...

Can you **share the data openly**? **Copyright?** **Deidentification?** (human subjects, vulnerable/endangered species); **Access restrictions** needed (restrict. embargo)

What **metadata standards** are required to describe your data properly?

**\*Sensitive data support is not yet available in Harvard Dataverse Repository. All data must be deidentified!**

**Parting thought: you can manage your repositories after changing affiliations!**

**Convert your Dataverse installation account away from your Institutional Log In**

If you are leaving your institution and need to convert your Dataverse installation account to the Dataverse Username/Email log in option, you will need to contact support for the Dataverse installation you are using. On your account page, there is a link that will open a popup form to contact support for assistance.

**Much more can be found in the Harvard  
Dataverse [user guide](#) and [prior presentations](#)**