



Guidelines for Citing Datasets

Data is a legitimate, citable product of research. Treating data as a scholarly product ensures that the hard work and effort of creating, curating, and publishing data is recognized and credit is duly given. Funders and publishers are increasingly expecting funded research data to be Findable, Accessible, Interoperable, and Reusable (FAIR). This is a good thing - research consumes large amounts of resources and it is only “fair” to make results accessible to review and build upon. Proper dataset citation can help.

Benefits of Data Citations

- Data producers: Provides proper attribution and credit; helps demonstrate impact of work; connects publications and supporting data; and creates a timeline of publication and usage.
- Data users: Easier to find datasets; supports dataset persistence; encourages reuse of data.
- Everyone: Increased transparency and reproducibility.

Dataset Citation Best Practices

- Include
 - Author
 - Date of publication
 - Title of publication
 - Publisher and/or distributor
 - Electronic location or persistent identifier (e.g., a DOI)
 - Version (when appropriate)
 - Date accessed (when appropriate)
- Add the completed data citation to your list of references so that it is registered by indexing programs.
- Some publishers may specify a format or ask for additional citation information. You should follow these requirements when applicable.
- If you are publishing a paper and your data underlies the research in the paper, cite your own data in your reference list.

Example Citations

- APA:
 - Yan, H., Sun, N., Eldardiry, H., Thurber, T., Reed, P. M., Malek, K., Gupta, R. S., Kennedy, D., Swenson, S., Vernon, C. R., Burleyson, C. D., & Rice, J. S. (2022). CLM5 CAMELS Basins Ensemble (v0.0.2) [Data set]. MSD-LIVE Data Repository. <https://doi.org/10.57931/1884563>
- Chicago:
 - Smith, Tom W., Peter V. Marsden, and Michael Hout. 2011. General Social Survey, 1972-2010 Cumulative File. ICPSR31521-v1. Chicago, IL: National Opinion Research Center. Distributed by Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR31521.v1
- Figshare:
 - Rodriguez, Tommy (2013): 17,170 Base Pair Alignment of Thirteen Time-Extended Lineages [data: (complete) mtDNA; format: ClustalW]. figshare. <https://dx.doi.org/10.6084/m9.figshare.815894> Retrieved: 16 26, Jan 04, 2016 (GMT)