Metadata template[1] for datasets of *L&O-Letters* articles

Table 1. Description of the fields needed to describe the creation of your dataset.

Title of dataset	*daymetr_meteorological_sites.csv*
URL of dataset	*https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/daymetr_meteorological_sites.csv*
Abstract	*Heatwaves are increasing in frequency, duration, and intensity in ocean, coastal, and lake ecosystems. While positive water temperature trends have been documented in many rivers, heatwaves have not been analyzed. This study examined heatwaves in rivers throughout the United States between 1996-2021. Riverine heatwaves increased in frequency over the study period, with significant increases in summer, in mid to high order streams, and at free-flowing sites and sites above a reservoir. The increase in heatwave frequency was accompanied by an increase in moderate strength heatwaves as well as an increasing trend in the average number of heatwave days each year. Riverine heatwaves were most likely to occur during periods of normal or below-normal discharge conditions and at sites with a mean annual discharge ≤ 250 m3 s-1. These results provide the first assessment of heatwaves in rivers for a large geographic area in the United States.* *The dataset described here provides station specific details regarding site number and location that is used with the 'daymetr' R package to access the meteorological Daymet web service. The resulting meteorological data is used to develop multiple linear regressions for each site to fill in gaps in the USGS derived water temperature time series so long as the regression R^2 ≥ 0.80*
Keywords	*Daymet, stream, river, meteorological, heatwaves*
Lead author for the dataset	*Spencer J. Tassone*
Title and position of lead author	*Graduate Student*
Organization and address of lead author	*Department of Environmental Sciences* *University of Virginia* *Clark Hall* *291 McCormick Rd* *P.O. Box 400123* *Charlottesville, VA 22904*
Email address of lead author	*sjt7jc@virginia.edu*
Additional authors or contributors to the dataset	*NA*
Organization associated with the data	*Oak Ridge National Laboratory (ORNL), National Aeronautics and Space Administration (NASA)*
Funding	*NA*
License	***CCO***

[1] *This document liberally borrows from a similar document provided by the Environmental Data Initiative*

Geographic location – verbal description	*United States of America*
Geographic coverage bounding coordinates	*Alabama, Alaska, Arizona, Arkansas, California, Colorado, Georgia, Michigan, Montana, Nevada, New Jersey, New York, Oregon, Pennsylvania, South Carolina, Texas, Washington*
Time frame - Begin date	*01-01-1996*
Time frame - End date	*12-31-2021*
General study design	*This study analyzed publicly, and freely available USGS derived water temperature data to examine riverine heatwaves throughout the U.S. Meteorological data, also publicly and freely available using the 'daymetr' R package were used to build regressions models to fill in gaps in the USGS water temperature data so long as the models were well fit ($R^2 \geq 0.80$).*
Methods description	*The R package 'daymetr' provides a programmatic interface to the Daymet web services. Daymet is a NASA supported product that is produced by the Oak Ridge National Laboratory and provides long-term, continuous, gridded estimates of daily weather and climatology variables at a 1 kilometer x 1 kilometer resolution. For batch downloads from daymetr, a comma separated file (i.e., csv) of site, latitude, and longitude is required. Using the 'download_daymet_batch' function, the comma separated file location must be specified in order to download the long-term, continuous daily weather observations.*
Laboratory, field, or other analytical methods	*The USGS conducts high-frequency surface water monitoring throughout the U.S. as part of its national water information system. All sites with daily mean water temperature records available for the 26-year period of 1996-2021 were identified using the R package 'dataRetrieval' version 2.7. Tidally influenced and lake sites were removed, as were all flagged data other than those 'Approved', 'Approved Revised', 'Approved Edited', or 'Provisional'. Sites with < 90% of their daily records were also excluded. Linear interpolation was applied to water temperature gaps ≤ 2 days. For larger gaps, multiple linear regression models were developed using 1 km^2 resolution, daily climate data using the R package 'daymetr' version 1.6. Only those sites with regressions where $R^2 \geq 0.80$ were used in this analysis (mean ± SD $R^2 = 0.91 \pm 0.04$). Seventy long-term water temperature sites were identified, resulting in a total of 1,820 station years of water temperature data available for analysis.* *The 'daymetr' R package requires a comma separated file (csv) with site names, latitude, and longitude for batch downloads. This csv provides those USGS site numbers and locations for which multiple linear regression models were developed using meteorological data.*
Taxonomic species or groups	*NA*
Quality control	*Once downloaded, the meteorological data were inspected for completeness as well as for extremes in each meteorological variable.*
Additional information	*Many USGS site numbers begin with the number zero. Noted in table 2, all site numbers in this analysis < 10301500 (< line 52) begin with a zero.*

Table 2. Data dictionary: description of the variables (i.e., columns) in EACH dataset.

Dataset filename: *daymetr_meteorological_sites.csv*
Dataset description: *This dataset provides the USGS site numbers and locations that are used to download long-term, daily weather observations using the R package 'daymetr'.*

Column name	Description	Units	Code explanation	Data format	Missing data code
site	*Site number issued by the USGS. All site numbers < 10301500 (< line 52) start with a zero which gets dropped in csv file format*	*NA*	*NA*	*Numeric*	*No missing data*
lat	*Site latitude provided by the USGS*	*Decimal degrees*	*NA*	*Numeric*	*No missing data*
lon	*Site longitude provided by the USGS*	*Decimal degrees*	*NA*	*Numeric*	*No missing data*

Table 3. Data provenance

Dataset title	Dataset DOI or URL	Creator (name & email)	Contact (name & email)
daymetr: Interface to the 'Daymet' Web Services	*https://cran.r-project.org/web/packages/daymetr/index.html*	*Koen Hufkens, koen.hufkens@gmail.com*	*Koen Hufkens, koen.hufkens@gmail.com*
Daymet: Daily surface weather and climatological summaries	*https://daymet.ornl.gov/*	*Oak Ridge National Laboratory, uso@daac.ornl.gov*	*ORNL DAAC User Services Center, uso@daac.ornl.gov*

Table 4. Scripts/code (software)

File name	Description	Scripting language
DataPull.R	*This file pulls the daily mean water temperature and discharge (Q) data from USGS using the 'dataRetrieval' R package. Similarly, meteorological data is accessed using the 'daymetr' R package.*	*R*

Notes and Comments:

- The Riverine Heatwave GitHub page where all the code to download and analyze the data for the riverine heatwave project is here: https://github.com/spencer-tassone/RiverineHeatwaves

- The specific code and GitHub page where the 'daymetr' package was used is here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/DataPull.R

- The GitHub link to the csv file used for the daymetr batch download is here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/daymetr_meteorological_sites.csv

- This metadata for the csv file above is also linked to the Riverine Heatwave GitHub here: https://github.com/spencer-tassone/RiverineHeatwaves/blob/main/ Metadata-LO-Letters-RiverineHeatwaves_Meteorological.pdf