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

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

Title of dataset	siData: Gas flux and driver data used for analysis
URL of dataset	10.6084/m9.figshare.5220001 (for manuscript doi:10.1002/lol2.10073)
Abstract	We collected published data consisting of CO2, CH4, and N2O flux from lakes and impoundments as well as the associated nutrient (TP and TN), chlorophyll a, and surface area data if available. We derived empirical models of flux for each gas using these variables. For the upscaling, the global chlorophyll a distribution in lakes from Sayers et al. 2015 was propagated through 3 different global lake size distributions (Downing et al. 2006, Verpoorter et al. 2014, Messager et al. 2016) and these tables can be found in the SI (Tables S2, S3, S4) as well as in the Figshare repository. Finally, we used the empirical size-chlorophyll a (or TP) models (found in Table S5 of the SI) to estimate flux in each size-chlorophyll bin (Tables S2-S4) and then summed these fluxes for global estimates. The data presented in file siDATA.csv in the repository contains the collected data used to build the empirical models (i.e., gas fluxes, nutrients, chlorophyll a, lake size).
Keywords	Methane, carbon dioxide, nitrous oxide, greenhouse gas, flux, nutrients, lakes, size, productivity, ebullition, diffusion, phosphorus, nitrogen, chlorophyll
Dataset lead author	Tonya DelSontro
Position of data author	Postdoctoral researcher
Address of data author	Dept. des sciences biologiques Université du Québec à Montréal Case postale 8888 Succursale Centre-ville Montréal H3C 3P8
Email address of data author	tdelsontro@gmail.com
Primary contact person for dataset	Tonya DelSontro Jake Beaulieu John Downing
Position of primary contact person	Postdoctoral researcher, University of Quebec at Montreal Government scientist, US Environmental Protection Agency Director, Minnesota Sea Grant
Address of primary contact person	Same as above
Email address of primary contact person	tdelsontro@gmail.com beaulieu.jake@epa.gov downing@d.umn.edu
Organization associated with the data	
Usage Rights	Publicly available and free to use, but request that the user cites this manuscript (DelSontro et al. 2017 in Limnology and Oceanography Letters) as the source of this collated data.
Geographic region	The data for our study was collected from previously published work. In general, data are from at least 54 countries spanning all continents, except Antarctica.
Geographic coverage	Detailed information given in individual publications (see Table S1 of SI for references).

Temporal coverage - Begin date	As data was synthesized from previous work, temporal coverage spans many years. Specific information found in individual references (see Table S1).
Temporal coverage - End date	As data was synthesized from previous work, temporal coverage spans many years. Specific information found in individual references (see Table S1).
General study design	As data was synthesized from previous work, precise study design would be found in the individual studies (see Table S1). Majority of studies sampled gas concentration and/or flux, as well as nutrients and chlorophyll, in one location, usually the center of the water body.
Methods description	As data was synthesized from previous work, precise methods would be found in the individual studies (see Table S1).
Laboratory, field, or other analytical methods	As data was synthesized from previous work, precise methods would be found in the individual studies (see Table S1). In general, CH4 and N2O concentrations were measured directly and fluxes were either measured directly using chambers or estimated using concentration and gas exchange models. Most CO2 concentrations were estimated using the carbonate equilibria approach and fluxes were estimated using gas exchange models. A handful of studies measured CO2 fluxes directly using chambers.
Quality control	As the data were collected from previously published works, quality control should have been dealt with during those studies and details can be found in individual manuscripts.
Additional information	siData.csv is of pruned dataset (see methods for details). All data in the siData.csv file is from previous work. References for collected data are located in Table S1 of the SI.

Table 2. Description of the variables (i.e., columns) in EACH dataset in sufficient detail for another user to understand and use the data. If there are 10 variables (i.e., columns) in the dataset, then there should be 10 rows in this column that describe each column.

Dataset filename: siData.csv

Column name	**Definition**	**Units**
The name of the variable in the dataset	*A detailed definition of the variable*	*Units the variable is measured in*
CH4.Total.mg.CH4.C.m.2.d.1	Total CH4 flux (Diffusion + Ebullition) from system if reported as total in literature	mg C-CH_4 m^{-2} d^{-1}
CH4.Diffusive.mg.CH4.C.m.2.d.1	Diffusive CH4 flux from system	mg C-CH_4 m^{-2} d^{-1}
CH4.Ebul.mg.CH4.C.m.2.d.1	Ebullitive CH4 flux from system	mg C-CH_4 m^{-2} d^{-1}
CO2.Diffusive.mg.CO2.C.m.2.d.1	Diffusive CO2 flux from system	mg C-CO_2 m^{-2} d^{-1}
N2O.Diffusive.mg.N2O.N.m.2.d.1	Diffusive N2O flux from system	mg N-N_2O m^{-2} d^{-1}
chla.ug.l	Chlorophyll a concentration	µg l^{-1}
Surface.Area.km2	Lake surface area	km^2
TP.ug.l	Total phosphorus concentration	µg l^{-1}
TN.mg.L.	Total nitrogen concentration	µg l^{-1}