**Instructions for sidb Data Entry**

Caitlin Hicks Pries

2019-03-06

An entry into the soil incubation database (sidb) consists of three files: metadata.yaml, timeSeries.csv, and initConditions.csv. The file metadata.yaml contains the metadata, which includes information about the location the soils were sampled from, field treatments, and laboratory treatments. The timeSeries.csv file contains all gas flux data taken over time in a wide format with time in days as the first column. The flux data can either be reported as rates or as cumulative flux. The initConditions.csv file contains information about soil properties at the start of the incubation such as carbon and nitrogen contents, bulk density, texture, and pH.

1. Download metadatatemplate.yaml and initConditions.csv files.
2. Follow instructions here, in yaml, and look at data files for examples. There are many entries and likely an entry already exists that demonstrates an experimental design similar to yours.
3. Conventions:
   1. We use camelCase to name fields in the yaml file. This consists of starting a phase with a lowercase letter and capitalizing all subsequent letters with no spaces between the words.
   2. If you type NA into a yaml file, R reads it as a factor not an NA. So leave any NA’s blank. For example, in an array, type – then <space> then <space>
   3. Please use these recommended experimental and laboratory treatment names for common treatment/soil designations:

temperature

moisture

midDepth

1. Unit recommendations:

|  |  |
| --- | --- |
| **Parameter** | **Unit** |
| latitude and longitude | Decimal degrees |
| MAT | Celsius |
| MAP | mm |
| depth | cm, m |
| temperature | Celsius |
| moisture | percentGWC, percentFieldCapacity, percentWaterFilledPoreSpace |
| carbon | percent, mg/gSoil, g/gSoil, microg/gSoil |
| nitrogen | percent, mg/gSoil, g/gSoil, microg/gSoil |
| bulkDensity | g/cm3 |
| redox | mV |
| fluxes | gC-CO2/gSoil, gC/gC, microgC/gC/h, molC/gSoil/d, or combinations of the above |

1. How to treat hierarchy
   1. If a hierarchal field in the yaml file, do not fill in the line on the top of the hierarchy, instead fill in value and units that are below that field. Example below.

moisture: LEAVE THIS BLANK

value: enter number here

units: enter units here

1. The importance of tabs or spaces. Yaml’s use tabs or spaces to distinguish levels of hierarchies. We used tabs to create the metadata file and suggest you do the same to be consistent. We also suggest using a text editor (such as ATOM or sublime text) that provides vertical lines to help see the hierarchies created by indentations. A common reason that your yaml cannot be read (see below) is when there is a tab in the wrong place or a tab missing where there should be one (as in delineating hierarchies).
2. If you have multiple levels for a site, depth, treatment, etc., you will create an array below the field indented with a tab followed by a dash, a <space>, and then the value. If you have a single level, you will enter it after the field’s colon on the same line. Example below.

temperature:

* 5
* 15

temperature: 5

1. Site is an important variable. If you have multiple sites, you will need to repeat the length of the site array for country, latitude, longitude, MAT, MAP, elevation, land cover, veg note, and soil taxonomy. These fields should have arrays the same length as the number of sites you have.
2. When finished, check that your files are readable by reading it into a computer program (R, Ruby, Python).
   1. R example:

#to test incubation data entry

devtools::install\_github('SoilBGC-Datashare/sidb/Rpkg', force=TRUE)

library(sidb)

readEntry(path=" ", entryName=" ")

#if error in yaml, you get a helpful error message of which line and column to #look at. Look for errant spaces and tabs or characters like colons or commas out #of place or not in quote.

1. If your files are readable in R and do not give any error or warning messages, you are now ready to submit your data. To do so, contact ??.