Guide to naming variables using the PaST Thesaurus

NOAA/World Data Service for Paleoclimatology (WDS-Paleo)
Last modified April 8, 2021

Table of Contents

1. Introduction

The WDS-Paleo uses the Paleoenvironmental Standard Terms (PaST) thesaurus (https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus) to describe measured and inferred variables in datasets. This standardized terminology offers a common language for describing paleoclimate information and is the basis of the WDS-Paleo search by variable (https://www.ncdc.noaa.gov/paleo-search). New contributions to the WDS-Paleo are required to use the PaST thesaurus to document all variables. This guide provides instructions to aid data contributors in creating variable names.

2. Structure of WDS-Paleo variable names

WDS-Paleo variable names have multiple components to capture information critical for understanding and reusing data (**Figure 1**). Nine of these components require terms from the PaST thesaurus: “What”, “Material”, “Error”, “Units”, “Seasonality”, “Data Type”, “Details”, “Method”, and “Data Format.”

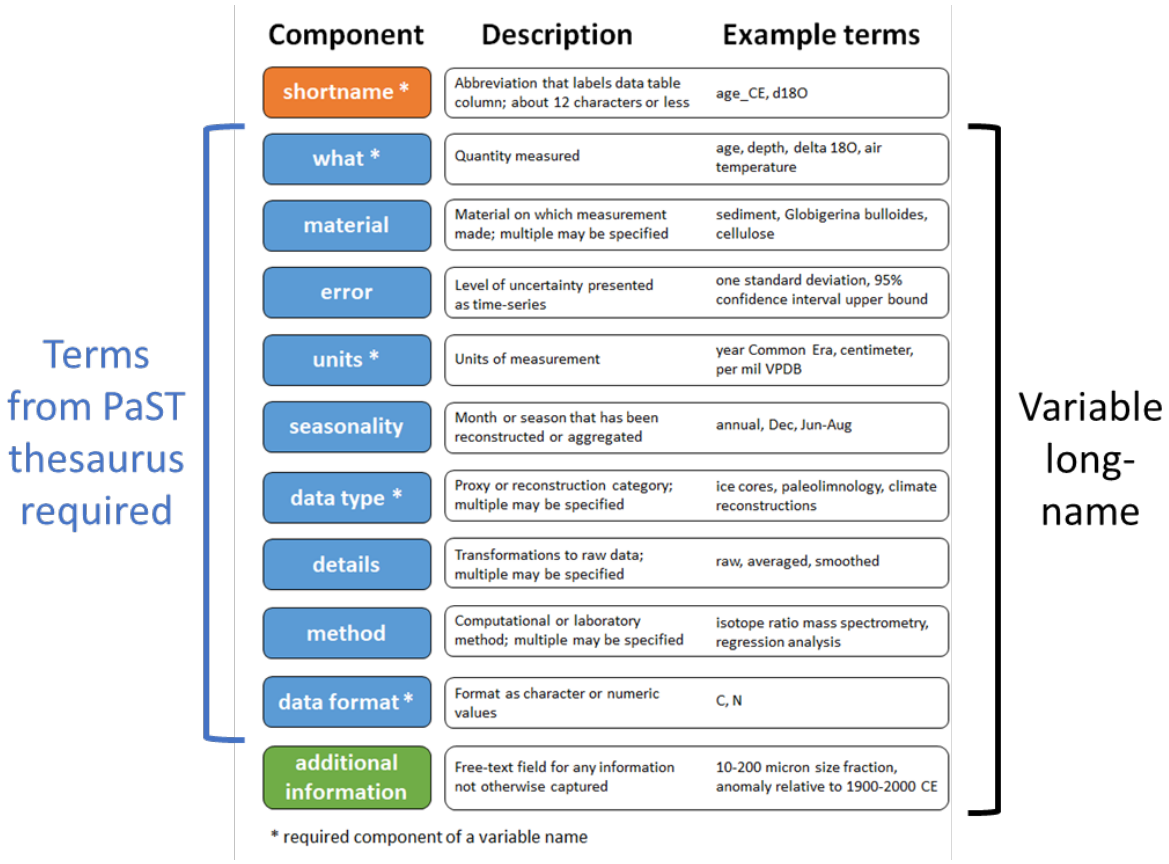
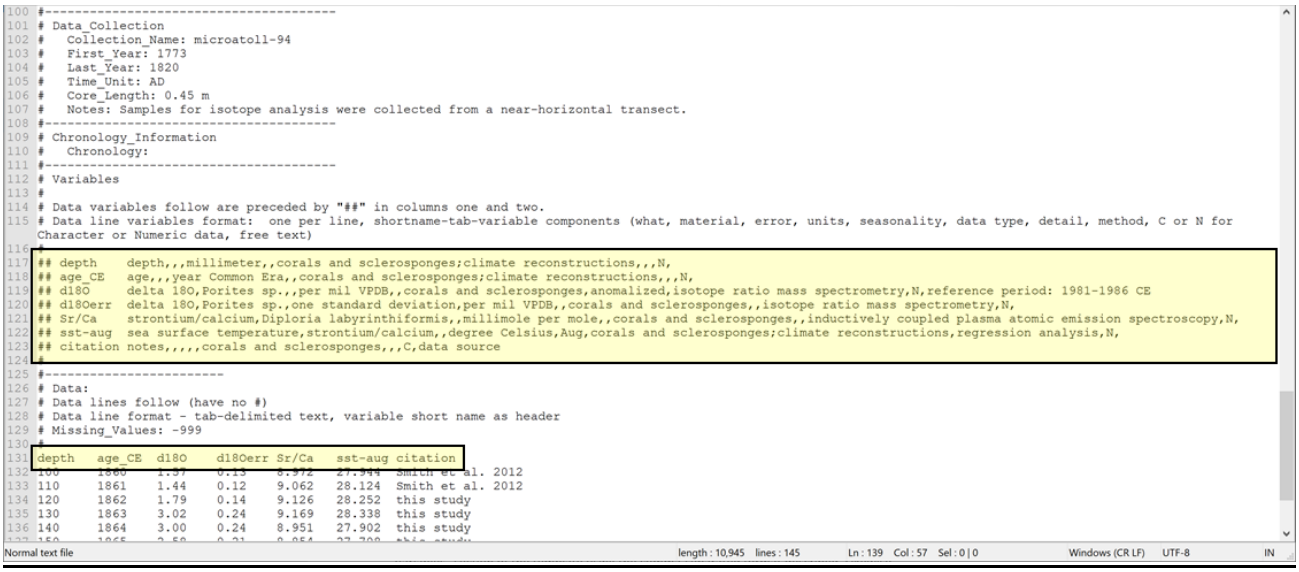


Figure 1. *Components of the WDS-Paleo naming scheme for measured or inferred variables. Components requiring terminology standardized by the PaST thesaurus are shown in blue. Variable short names label columns of a data table, while the comma-separated components of the variable long name provide more detailed information about the variable (see* ***Figure 2****).*

Contributors submitting data in WDS-Paleo text template format should specify variable names in the "Variables" section of the template, which is available at: https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/contributing. In the text template snippet in **Figure 2**, the names of measured and inferred variables are highlighted in yellow. The upper yellow box has seven variable names, one per row. Each variable row begins with the variable short name (e.g., "depth" or "age_CE"), followed by a tab and the variable long name. This long name consists of the nine components that must be specified using terms in the PaST thesaurus plus one component ("Additional Information") that is a free-text field. The lower yellow box highlights how the variable short names label columns of the data table.

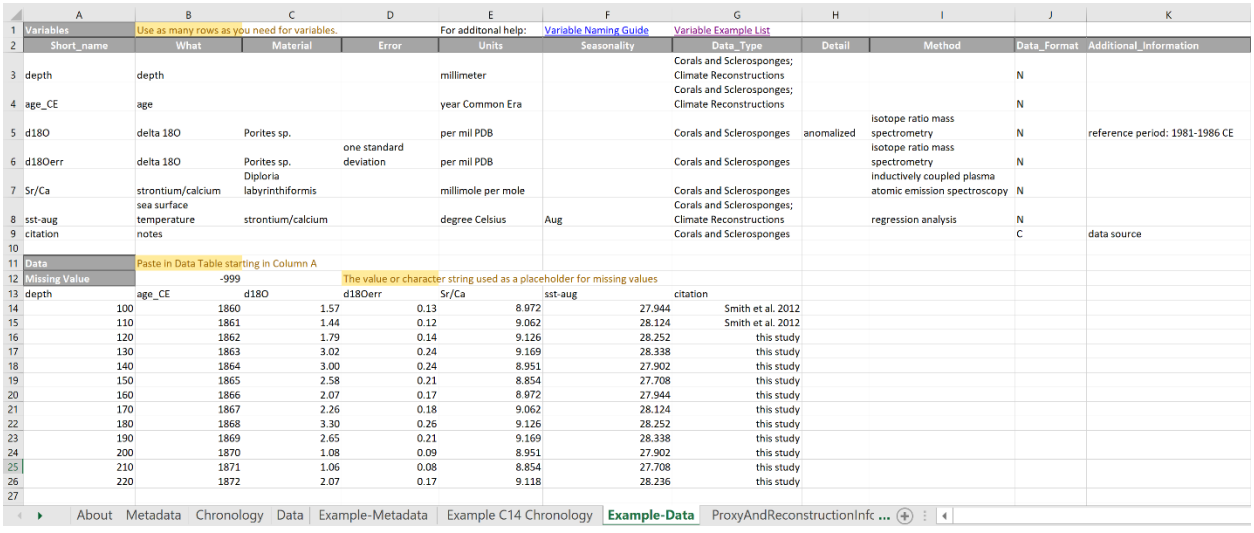
Contributors submitting data in WDS-Paleo Microsoft Excel template format should enter variable names in the "Data" tab of the template, which is also available at: https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/contributing. **Figure 3** shows how the variable names used in the text template example are entered in the Excel template.



```
#----------------------------------------
# Data_Collection
#   Collection_Name: microatoll-94
#   First_Year: 1773
#   Last_Year: 1820
#   Time_Unit: AD
#   Core_Length: 0.45 m
#   Notes: Samples for isotope analysis were collected from a near-horizontal transect.
#----------------------------------------
# Chronology_Information
#   Chronology:
#----------------------------------------
# Variables
#
# Data variables follow are preceded by "##" in columns one and two.
# Data line variables format:  one per line, shortname-tab-variable components (what, material, error, units, seasonality, data type, detail, method, C or N for
Character or Numeric data, free text)
#
## depth    depth,,,millimeter,,corals and sclerosponges;climate reconstructions,,,N,
## age_CE   age,,,year Common Era,,corals and sclerosponges;climate reconstructions,,,N,
## d18O     delta 18O,Porites sp.,,per mil VPDB,,corals and sclerosponges,anomalized,isotope ratio mass spectrometry,N,reference period: 1981-1986 CE
## d18Oerr  delta 18O,Porites sp.,one standard deviation,per mil VPDB,,corals and sclerosponges,,isotope ratio mass spectrometry,N,
## Sr/Ca    strontium/calcium,Diploria labyrinthiformis,,millimole per mole,,corals and sclerosponges,,inductively coupled plasma atomic emission spectroscopy,N,
## sst-aug  sea surface temperature,strontium/calcium,,degree Celsius,Aug,corals and sclerosponges;climate reconstructions,regression analysis,N,
## citation notes,,,,,corals and sclerosponges,,,C,data source
#
#------------------------
# Data:
# Data lines follow (have no #)
# Data line format - tab-delimited text, variable short name as header
# Missing_Values: -999
#
depth   age_CE  d18O    d18Oerr Sr/Ca   sst-aug citation
100     1860    1.57    0.13    8.972   27.944  Smith et al. 2012
110     1861    1.44    0.12    9.062   28.124  Smith et al. 2012
120     1862    1.79    0.14    9.126   28.252  this study
130     1863    3.02    0.24    9.169   28.338  this study
140     1864    3.00    0.24    8.951   27.902  this study
```

Normal text file — length : 10,945 lines : 145 — Ln : 139 Col : 57 Sel : 0 | 0 — Windows (CR LF) UTF-8 IN

Figure 2. *Snippet of WDS-Paleo machine-readable data template, showing metadata for measured and inferred variables in yellow highlight. Terms used in this example follow specifications of the PaST thesaurus. The text template is the standard archive version of WDS-Paleo data files. While nearly all data sets at the WDS-Paleo are searchable using PaST, only text templates created in and after March 2020 display PaST-compliant variable names.*



	A	B	C	D	E	F	G	H	I	J	K
1	Variables	Use as many rows as you need for variables.			For additonal help:	Variable Naming Guide	Variable Example List				
2	Short_name	What	Material	Error	Units	Seasonality	Data_Type	Detail	Method	Data_Format	Additional_Information
3	depth	depth			millimeter		Corals and Sclerosponges; Climate Reconstructions			N	
4	age_CE	age			year Common Era		Corals and Sclerosponges; Climate Reconstructions			N	
5	d18O	delta 18O	Porites sp.		per mil PDB		Corals and Sclerosponges	anomalized	isotope ratio mass spectrometry	N	reference period: 1981-1986 CE
6	d18Oerr	delta 18O	Porites sp.	one standard deviation	per mil PDB		Corals and Sclerosponges		isotope ratio mass spectrometry	N	
7	Sr/Ca	strontium/calcium	Diploria labyrinthiformis		millimole per mole		Corals and Sclerosponges		inductively coupled plasma atomic emission spectroscopy	N	
8	sst-aug	sea surface temperature	strontium/calcium		degree Celsius	Aug	Corals and Sclerosponges; Climate Reconstructions		regression analysis	N	
9	citation	notes					Corals and Sclerosponges			C	data source
10											
11	Data	Paste in Data Table starting in Column A									
12	Missing Value	-999		The value or character string used as a placeholder for missing values							
13	depth	age_CE	d18O	d18Oerr	Sr/Ca	sst-aug	citation				
14	100	1860	1.57	0.13	8.972	27.944	Smith et al. 2012				
15	110	1861	1.44	0.12	9.062	28.124	Smith et al. 2012				
16	120	1862	1.79	0.14	9.126	28.252	this study				
17	130	1863	3.02	0.24	9.169	28.338	this study				
18	140	1864	3.00	0.24	8.951	27.902	this study				
19	150	1865	2.58	0.21	8.854	27.708	this study				
20	160	1866	2.07	0.17	8.972	27.944	this study				
21	170	1867	2.26	0.18	9.062	28.124	this study				
22	180	1868	3.30	0.26	9.126	28.252	this study				
23	190	1869	2.65	0.21	9.169	28.338	this study				
24	200	1870	1.08	0.09	8.951	27.902	this study				
25	210	1871	1.06	0.08	8.854	27.708	this study				
26	220	1872	2.07	0.17	9.118	28.236	this study				
27											

About | Metadata | Chronology | Data | Example-Metadata | Example C14 Chronology | **Example-Data** | ProxyAndReconstructionInfc ...

Figure 3. *Snippet of WDS-Paleo Microsoft Excel data template, showing how variable terminology is input. The WDS-Paleo auto-generates text template files from contributed Excel templates. The Variables section in the text template file generated from this Excel example would appear as in* ***Figure 2****.*

3. Accessing the PaST thesaurus

The PaST Navigator (https://www.ncdc.noaa.gov/paleo-search/cvterms) provides a way to search for terms within PaST and to access term definitions and relationships. The Navigator interface (**Figure 4**) allows users to find terms via a text-based search and also by clicking through the hierarchical structure of the thesaurus using broader and more specific terms.

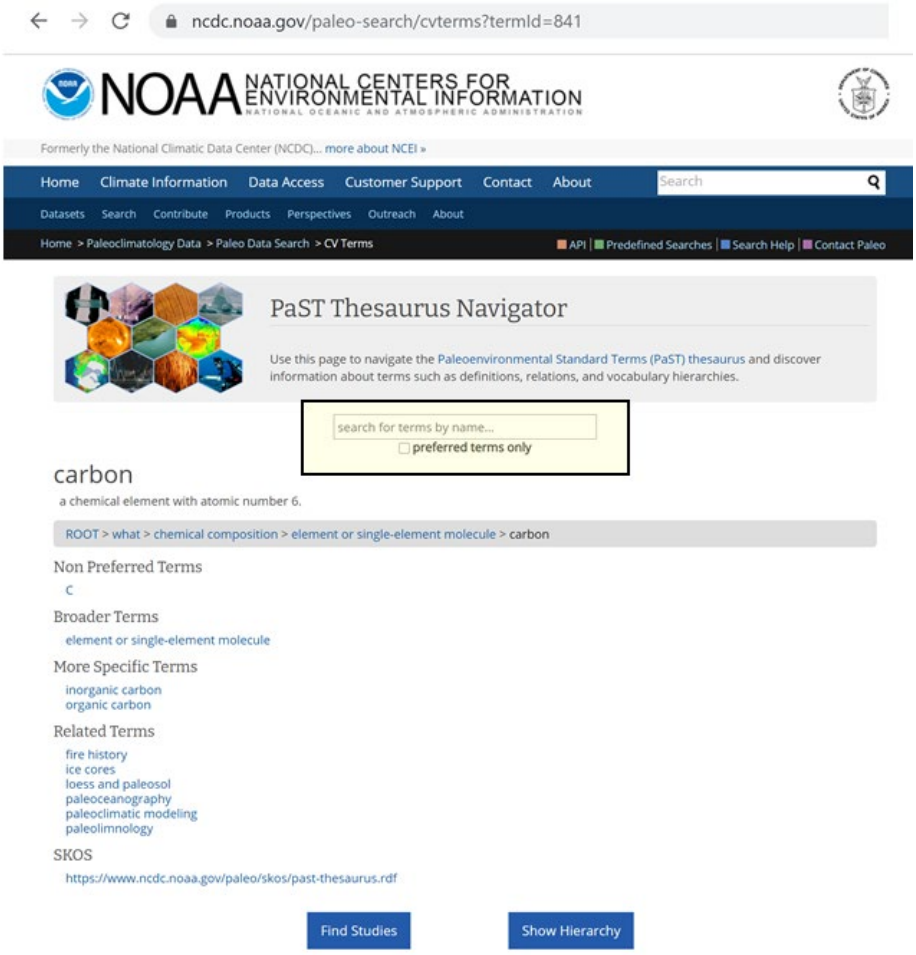


Figure 4. *Interface of PaST Thesaurus Navigator, showing the entry for "carbon" as an example. The text search, highlighted in yellow, matches strings to either preferred terms only or to both preferred and non-preferred terms. All terms displayed are clickable links to separate entries in the thesaurus. The "Find Studies" button searches for all studies containing the displayed term, and the "Show Hierarchy" button produces a tree diagram for the displayed term.*

The WDS-Paleo Data Search (https://www.ncdc.noaa.gov/paleo-search) is also useful for finding examples of variable names that are relevant for your dataset. Both the general (free-text) and the advanced variables searches locate previously-archived studies that employ particular terms. While the advanced search is limited to the "What", "Material", and "Seasonality" components, the general search works for all components of the variable long name.

4. General instructions for naming variables

There are several general points to keep in mind when naming variables:

- Each column of data in a data table must have a long name and short name that are unique within the data file.
- Not all of the long-name fields must be used for each variable. In fact, some variables, particularly age and depth variables, will employ only a few of the fields.
- There are only four required fields in the long name: "What," "Units," "Data Type," and "Data Format." The more complete the variable description is, however, the more useful it will be to others.
- The fields comprising the long name are separated by commas. If a comma is needed *within* one of the fields, enclose it in quotation marks. No tabs or newline characters (carriage returns) can be used. In text templates, ensure there is a comma separating each category, even where no value exists.
- Data contributors may propose new terminology when making their submission if existing terms in the PaST thesaurus are not sufficient to describe their variables. Enter the proposed term directly into the data template. WDS-Paleo staff will verify that the new term does not duplicate existing terms in PaST.

5. Specific instructions for naming variables

Considerations specific to individual components of a variable name include:

"Short name" (REQUIRED): Abbreviation of variable

- White-space characters are not allowed. Use underscore, period, or dash instead.
- Keep these short – about 12 characters or less.

"What" (REQUIRED)

- Describe what was measured.

"Material"

- Describe the material on which measurements were made.
- Multiple materials may be specified; separate each with a semicolon. The following example shows the first two components ("What" and "Material") of a long name with two materials:

delta 15N, Porites sp.; alanine, …..

"Error"

- Use ONLY for data series that are measurements of uncertainty or error. A variable that makes use of this field is always used in conjunction with another variable to which the error measurements applies.
- For example, a reconstruction of sea surface temperature is reported with additional columns recording the one standard deviation lower bound and upper bound errors for each reconstructed data point. The first four components ("What", "Material", "Error", and "Units") of the long names for these three data series are as follows; note that they differ only in the error term:

sea surface temperature, , , degree Celsius, …..
sea surface temperature, , one standard deviation lower bound, degree Celsius, …..
sea surface temperature, , one standard deviation upper bound, degree Celsius, …..

- If the error is reported as a constant, this information should be placed in the "Description and Notes" portion of the template rather than in the data table.

"Units" (REQUIRED FOR NUMERIC DATA, BUT NOT FOR CHARACTER-BASED DATA)

- Provide the units of measurement.

"Seasonality"

- Describe either (1) the specific part of the annual cycle that has been explicitly reconstructed from raw data or (2) the time window (e.g., Annual, Mar, Dec-Feb) over which aggregation of raw sub-annually resolved data has occurred. Examples of these cases, respectively, are:

precipitation, ring width, , millimeter, Jun-Aug, tree ring; climate reconstructions, , , N,
delta 18O, Porites lutea, , per mil VPDB, Dec-Feb, corals and sclerosponges, averaged, isotope ratio mass spectrometry, N, monthly samples averaged to three month mean

- Do not use this field for interpretations or inferences of seasonality.

"Data Type" (REQUIRED)

- Valid entries are: Borehole, Climate Forcing, Climate Reconstructions, Corals and Sclerosponges, Fauna, Fire History, Historical, Ice Cores, Insect, Instrumental, Lake Levels, Loess, Other Collections, Paleoceanography, Paleoclimatic Modeling, Paleolimnology, Plant Macrofossils, Pollen, Speleothems, and Tree Ring.
- Multiple data types may be specified; separate each with a semicolon. This is particularly common in the case of a climate reconstruction. The following example shows long names for raw delta 18O data and a climate reconstruction of sea surface

temperature derived from these raw data:

delta 18O, Porites sp., , per mil VPDB, , corals and sclerosponges, , , N,
sea surface temperature, delta 18O, , degree Celsius, , corals and sclerosponges; climate reconstructions, , , N,

"Detail"

- Provide information about alterations or transformations made to the raw data.
- Specific details may be provided in the "Additional Information" field (e.g., "anomalized to the 1981-1986 CE base period" or "smoothed using a 5-year running mean").
- Multiple details may be provided; separate each with a semicolon.

"Method"

- Provide information about the instrument, analytical technique, or reconstruction method used.
- Multiple methods may be provided; separate each with a semicolon.

"Data Format" (REQUIRED)

- Define whether data are numeric ("N") or character-based ("C").

"Additional Information"

- Provide any additional information as free-text (UTF8) that is not captured by the fields described above. Examples include further details on data transformations (e.g., how a dataset was corrected or smoothed) or units (e.g., "before present" anchored to a year other than 1950 CE).
- This field is also suitable for storing project-specific information (e.g., PAGES 2K climate interpretations in JSON format).
- Newline characters are not allowed.

6. Contact information

Send questions about contributing data or about using the PaST Thesaurus to describe variables to: paleo@noaa.gov