

Short guide to naming variables using the PaST Thesaurus

NOAA/World Data Service for Paleoclimatology

Last modified October 2019

The WDS-Paleo uses a nine-part variable name, along with a variable abbreviation and an optional free-text field for additional variable information, to describe each column of data in a data table. Terms used in the nine-part variable name must exist in the WDS-Paleo Paleoenvironmental Standard Terms (PaST) Thesaurus. These variable names offer a common language for describing paleoclimate information and are the basis of the WDS-Paleo search by variable at www.ncdc.noaa.gov/paleo-search. This guide provides instructions to aid data contributors in creating variable names.

General instructions

- Nine fields plus a free-text entry collectively form the variable “long name,” and the variable abbreviation is called the “short name.” The short name is used when a shortened form of the variable name is needed, for example as headers at the top of the data table.
- Each column of data in a data table must have a long name and short name, as defined in the bullet above, 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.
- Nine fields of the long-name must be populated from the pre-defined set of terms in the PaST Thesaurus (www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus). These fields, which are described in more detail in the specific instructions below, are: “What,” “Material,” “Error,” “Units,” “Seasonality,” “Data Type,” “Detail,” “Method,” and “Data Format.”
- Controlled terms, as well as their definitions, for the nine fields of the long name are available at: www.ncdc.noaa.gov/paleo-search/cvterms.
- There are 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. Ensure there is a comma separating each category, even where no value exists. 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.
- Avoid abbreviations and acronyms within the long name.

Specific instructions

“Short name”: Abbreviation of variable

- Spaces 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 of a variable name in such a case:

What	Material
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 of the variable names for these three data series would be the following:

What	Material	Error	Units
sea surface temperature			degrees Celsius
sea surface temperature		one standard deviation lower bound	degrees Celsius
sea surface temperature		one standard deviation upper bound	degrees 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)

- 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.
- 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 partial variable names for a climate reconstruction and the raw paleoclimate data on which it was based:

What	Material	Units	Data Type
delta 18O	Porites sp.	per mil VPDB	corals and sclerosponges
sea surface temperature	delta 18O	degrees Celsius	climate reconstructions; corals and sclerosponges

“Detail”

- Provide information about alterations or transformations made to the raw data.
- 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 nine 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 year other than 1950 CE).
- This field is also suitable for storing project-specific information (e.g., PAGES 2K climate interpretations in JSON format).