



Paleoceanography and Paleoclimatology

Supporting Information for

The Paleoenvironmental Standard Terms (PaST) Thesaurus: Standardizing heterogeneous variables in paleoscience

Carrie Morrill^{1,2}, Bridget Thrasher^{1,2,*}, Samuel N. Lockshin^{1,2}, Edward P. Gille^{1,2}, Shelley McNeill^{3,4}, Ethan Shepherd^{3,4}, Wendy S. Gross², and Bruce A. Bauer²

¹ Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder CO 80309.

² NOAA's National Centers for Environmental Information, Boulder, CO 80305.

³ Riverside Technology, Inc., Fort Collins, CO 80525.

⁴ NOAA's National Centers for Environmental Information, Asheville, NC 28801.

* Now at: National Center for Atmospheric Research, Boulder, CO 80307.

Contents of this file

Text S1

Introduction

This supporting information file contains the text of Version 1.0 of the Paleoenvironmental Standard Terms (PaST) Thesaurus governance document.

Governance of the Paleoenvironmental Standard Terms (PaST) Thesaurus

NOAA/World Data Service for Paleoclimatology

Version 1.0

April 8, 2021

I. Description of Thesaurus

The Paleoenvironmental Standard Terms (PaST) Thesaurus contains terms to describe measurements of paleoenvironment and paleoclimate. These terms describe the quantity measured (“What”), the material on which it was measured (“Material”), reported error (“Error”), units (“Units”), analytical or statistical methods (“Method”), transformations made to raw data (“Detail”), reconstructed seasonality (“Seasonality”), proxy or reconstruction category (“Data Type”), and character or numeric format (“Data Format”). The PaST Thesaurus is based upon work supported by the National Science Foundation under Grant Number 1545762¹ and is maintained by the World Data Service for Paleoclimatology (WDS-Paleo) of NOAA's National Centers for Environmental Information. The WDS-Paleo requires data contributors to use these terms, or to propose new terms, when submitting their data. These terms also form the basis of a search of data by measurement type provided by the WDS-Paleo (<https://www.ncdc.noaa.gov/paleo-search>). More information about the thesaurus is available at: <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus>.

II. Scope of this document

This document describes the governance of, and specifically the process of changing, the PaST Thesaurus. This document does not provide complete details of all internal review processes, timelines or schedules, or implementation of the thesaurus within WDS-Paleo data infrastructure.

III. Roles and responsibilities

- a) WDS-Paleo: The staff of the WDS-Paleo have final authority and approval responsibility for changes to the thesaurus or its governance. Staff review data contributions for compliance with the thesaurus and convene advisory panels to review the thesaurus on an ad hoc basis (See Section IV). Staff review changes to the thesaurus proposed by subject matter experts and may delegate to these experts specific aspects of updating the thesaurus or assigning terms to data sets.

¹ Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

- b) Data contributors: All data contributors must use PaST to describe variables in data submissions to the WDS-Paleo. Contributors may propose new terms or corrections to existing terms including their relationships, definitions, and usage.
- c) Advisory panels: On the request of the WDS-Paleo, advisory panels review terminology and relationships in the thesaurus and recommend improvements.
- d) Subject matter experts: Any subject matter expert may propose changes to the thesaurus or to terms assigned to particular data sets.

IV. Process of changing the Thesaurus

Two types of changes may be made to the PaST Thesaurus: additions and modifications. Terms may not be deleted, as further discussed in (d) below.

- a) New terms may be added to the thesaurus on an as-needed basis to describe measurements being contributed to the WDS-Paleo. If none of the terms in the thesaurus are sufficient to describe a new data contribution, the dataset contributor may suggest new terminology at the time of data submission. WDS-Paleo staff will then review the suggestion and, if necessary, confer with the contributor to verify that the proposed term is distinct from any existing terms and meets term recommendations as described in section IV below. Once verified, WDS-Paleo staff will add the new term, its definition, any non-preferred terms, and all relevant related terms (including its broader term and, in the case of “What” or “Material” terms, its related data type or types) to the thesaurus in consultation with the contributor of the new dataset.
- b) Proposed modifications to existing terms, their definitions, their non-preferred terms, and/or their related terms that are small (i.e., having little impact on other users) may be made on an as-needed basis and without external review. Such modifications may be made in response to a user comment or request, or following identification of a necessary update or correction by WDS-Paleo staff.
- c) More significant modification (i.e., having greater impacts to other users) of the thesaurus, by adding a new set of terms or by modifying existing terms, definitions, non-preferred terms, and/or related terms, will be reviewed externally. Like small modifications, these changes may be in response to user comments or requests, or in response to needs identified by WDS-Paleo staff. When a significant subset of terms needs to be re-examined or added, the WDS-Paleo will convene a virtual advisory panel of approximately 2 to 4 subject matter experts to ensure that the proposed changes align with community standards.

- d) In order to maintain backward compatibility, no term will be deleted from the thesaurus. In the case that an existing term is to be renamed, the existing term will be preserved as a non-preferred term.
- e) Changes to the thesaurus, either as additions or modifications, will be documented internally by the WDS-Paleo in a variable history field. This history field tracks the date of the change, the WDS-Paleo staff member making the change, and the reason for the change.
- f) Incremental changes occur to the thesaurus through the processes described above. The frequency of the incremental updates depends on the frequency of new contributions and of user feedback. Minor version numbers (e.g., version 2.1) represent a snapshot in time of the thesaurus and will be given following a number of incremental changes. Minor versions will be released periodically on an as-needed basis, but generally not more than once per year.
- g) Major version numbers (e.g., version 2.0), which also represent a snapshot in time of the thesaurus, will be given following significant changes to multiple subcomponents of the thesaurus. WDS-Paleo staff will convene a virtual advisory panel before each new major version release. This panel does not need to review the thesaurus as a whole, but will review the components of the thesaurus that have undergone significant change since the last major version release.
- h) Users who wish to be notified of the release of any new major or minor versions of the thesaurus may send a request to paleo@noaa.gov to be added to a mailing list.

V. Term recommendations

The following are strongly encouraged recommendations for generating new terms or for modifying existing terms.

- a) Abbreviations and acronyms should be avoided, but may be added as non-preferred terms.
- b) All characters must exist and be encoded in UTF-8.
- c) Multiple terms that represent essentially the same concept should be avoided (e.g., mineral matter and lithic grain). The more commonly-used term should be the preferred term and any other terms should be maintained as non-preferred terms.
- d) Terms should generally be singular rather than plural.
- e) Terms should be in lowercase, with the exception of terms that are always or nearly always capitalized (e.g., *Globigerina bulloides*, North Atlantic Oscillation).

- f) Very specific terms that are applicable to a limited number of datasets should be avoided. In this case, a more general term should be applied and the more specific information included as “Additional Information.”
- g) Terms should reflect known, rather than inferred, quantities or values. For example, a measurement thought to reflect past temperature should not be assigned a temperature term unless temperature has been explicitly reconstructed.
- h) Upon recommendation of subject matter experts, the thesaurus includes genus and species names for corals, coralline algae, and selected foraminifera only.
- i) When a term in the thesaurus exists in another domain-specific vocabulary, this connection should be referenced by including the name of the other vocabulary and the relevant term identification number from the other vocabulary in the PaST Thesaurus definition field.

VI. Accessing the Thesaurus

All numbered versions of the thesaurus in SKOS format are available at: <https://www.ncdc.noaa.gov/paleo/study/31892>. The current thesaurus including changes made since the last released version can be downloaded in SKOS format, searched, browsed, and visualized at: <https://www.ncdc.noaa.gov/paleo-search/cvterms>.

VII. Versioning the Governance Document

Changes to this governance document will receive new version numbers. All current and past versions of the governance document with change logs are available at: <https://www.ncdc.noaa.gov/paleo/study/31892>.