

New advances at NOAA's World Data Service for Paleoclimatology - Promoting the FAIR principles

Wendy Gross¹, C. Morrill^{1,2} and E. Wahl¹

Guided by FAIR principles as best data management practices (Wilkinson et al. 2016), the World Data Service for Paleoclimatology (WDS-Paleo) at NOAA's National Centers for Environmental Information (NCEI) has recently deployed new capabilities and data-format standards. These and planned future developments facilitate the standardization and aggregation of WDS-Paleo's small, long-tail, and heterogeneous datasets into larger standardized collections. These capacities enhance the value of the data, analogous to how large volumes of well-managed big data can be transformed into valuable information (Lehnert and Hsu 2015).

WDS-Paleo archives and provides paleoclimatology data products derived from a variety of sources, such as tree rings, ice cores, corals, and ocean and lake sediments, along with web-based services to access these products. To attain the goal of long-term professional preservation and dissemination of its data, WDS-Paleo partners with its user communities and maintains long-standing relationships with PAGES, PANGAEA and Neotoma. WDS-Paleo works with these partners to offer aggregated search capabilities. NCEI data stewardship operations meet the responsibilities of an Open Archival Information System (oais.info), and new and existing capabilities follow FAIR best practices as follows.

Findable

WDS-Paleo makes its data findable via geographic map-based searches, along with a web service featuring an application programming interface (API) for programmatic use and a graphical user interface (GUI) that also acts as an API-builder tool (ncdc.noaa.gov/paleo-search). Recently a new controlled vocabulary, *Paleoenvironmental Standard Terms* (PaST; ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus) has been developed for documenting variables (i.e. paleoclimate measurements, units and methods). With the input of 25 subject-matter experts, terms have been assigned to over 100,000 paleoclimatic time series, powering new search capabilities that complement other web-service features. Paleoclimate data are extremely heterogeneous, and with PaST terminology WDS-Paleo's search capabilities now capture this heterogeneity. In the future, interactive visualizations of PaST will allow users to obtain more-detailed information about terms and will enhance data discovery. The governance structure for PaST is described at the above link.

Accessible

A recently released feature of the WDS-Paleo web service provides users with capacity to bundle and download search results, thus easing the process of procuring sets of data appropriate for specific use.

A bundle includes data files and manifest information, maintaining provenance of both data and metadata.

Interoperable

Upcoming and long-standing data and metadata formats promote interoperability via machine readability and common tools. Going ahead, a NOAA Standard for the Linked Paleo Data Format (LiPD, lipd.net) (now in development), is designed to facilitate interoperability between LiPD and the NOAA WDS-Paleo Template data format, including use of PaST terms. Standardized metadata formats, including DIF, ISO, and JSON, facilitate data discovery and federated search capabilities. Community-specific data formats and software tools, including those developed and used by the International Tree Ring Databank (ITRDB) and International Multiproxy Paleofire Database (IMPD), provide key resources for scientific discovery.

Reusable

PaST-enhanced web-service search capabilities aggregate WDS-Paleo's small, long-tailed datasets into larger, standardized collections. This can facilitate large-scale data syntheses, which is a key thrust in paleoclimatology (e.g. PAGES 2k Consortium 2017), and also promotes reuse of paleo data beyond paleoclimate specialists. WDS-Paleo is implementing persistent identifiers and locators for datasets via the provision of DOIs.

Going forward, PaST and a future project of standardizing the reporting of age determination will greatly enhance the interoperability of WDS-Paleo data formats, allowing for easier and more robust aggregation of datasets.

The WDS-Paleo website, including search, access, data contribution, and PaST information is at: ncdc.noaa.gov/paleo

AFFILIATIONS

¹NOAA's National Centers for Environmental Information, Boulder, USA

²Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, USA

CONTACT

WDS-Paleo: paleo@noaa.gov

REFERENCES

Lehnert K, Hsu L (2015) 11: 368-369

PAGES 2k Consortium (2017) *Sci Data* 4: 170088

Wilkinson MD et al. (2016) *Sci Data* 3: 160018

Paleoenvironmental Standard Terms

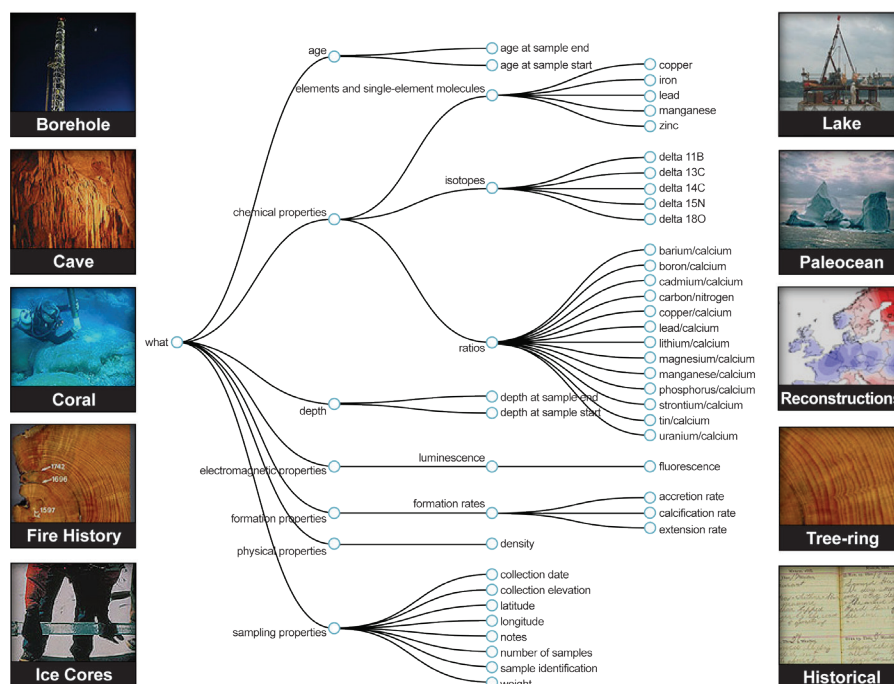


Figure 1: Example of Paleoenvironmental Standardized Terms (PaST) controlled vocabulary.