doi.org/10.22498/pages.31.1.49

News from paleoscience data organizations

Darrell Kaufman[1] and Lukas Jonkers[2]

A major outcome for many PAGES Working Groups (WGs) (pastglobalchanges.org/science/current-wg) is their data-intensive synthesis product featuring a high level of data stewardship. These follow the long tradition in the paleoscience community of sharing and caring for the community's data resources. PAGES WGs and the broader paleoscience community are fortunate in these efforts for the support of numerous discipline-specific data repositories and organizations. Their services help assure that our data are FAIR – findable, accessible, interoperable, and reusable (go-fair.org/fair-principles). We asked five international paleo data organizations to tell us about their latest activities of interest to the PAGES community.

PANGAEA

PANGAEA (Data Publisher for Earth & Environmental Sciences) (pangaea.de) is an open-access library for archiving, publishing, and disseminating georeferenced data from the Earth, environmental, and biodiversity sciences. It is operated jointly by the German research institutions MARUM in Bremen and AWI in Bremerhaven. Grown from its origins as a paleoclimate archive, PANGAEA now hosts >400000 datasets with more than 24 billion datapoints. Each dataset includes a data citation and a unique Digital Object Identifier (DOI). To add additional value to the archived datasets, PANGAEA now offers data-usage statistics where the number of web-based interactions with each dataset are tracked, and the impact of data sharing can be measured.

LinkedEarth

LinkedEarth (linked.earth) advances paleoclimate data standards and research. Datasets tied to LinkedEarth projects typically employ the Linked PaleoData (LiPD) structure. To improve searching across the breadth of LiPD datasets, the vocabulary for seven key paleoclimate variables has now been standardized on lipdverse.org, while aligning them to NOAA's PaST Thesaurus (ncei.noaa.gov/products/paleoclimatology/paleoenvironmental-standard-terms-thesaurus). The open-source code ecosystem around LiPD is expanding, most notably with geoChronR (McKay et al. 2021), actR (github.com/LinkedEarth/actR) and Pyleoclim (Khider et al. 2022); new LiPD utilities in Python and analytical tools in R are forthcoming. Online workshops are training paleoscientists to use this budding research ecosystem, and an advanced in-person workshop will take place during summer 2023 in California, USA.

EuroClimHist

EuroClimHist (euroclimhist.unibe.ch/en) is a database of historical climate data based on documentary and early instrumental evidence augmented by proxy records from natural archives. It is operated by the Institute of History, Department of Social Economic and Environmental History, and the Oeschger Centre for Climate Change Research at the University of Bern, Switzerland. The volume of data that is publicly accessible has recently increased. Two larger bodies of Swiss weather observations, including large phenological data on a nearly daily basis, have been included (Rudolf Salis-Marschlins, 1781–1800; Jakob Hänni 1839–1870). In collaboration with the World Glacier Monitoring Service (WGMS), a large number of historical paintings, drawings and early photographs of Swiss glaciers have recently been entered into the database. An upcoming project will focus on glaciers from the western Alps and Norway.

World Data Service for Paleoclimatology

The World Data Service for Paleoclimatology (ncei.noaa.gov/products/paleoclimatology), hosted by National Oceanic and Atmospheric Administration (NOAA), has now completed minting new DOIs for the backlog of >10 000 individual datasets that have been contributed over the past three decades. The NOAA Paleoclimatology now routinely assigns DOIs to new contributions. The DOIs are available on each landing page and in metadata files (see this example landing page). Instructions for new submissions, along with requirements and timelines for obtaining a dataset DOI, are being updated and will be announced via the Paleoclimate List when available. Dataset DOIs are important to cite along with publications to give full credit to data generators.

Neotoma

The Neotoma Paleoecology Database (neotomadb.org) features an improved website with access to 31,900 datasets, comprising 10.7 million datapoints from 19,100 sites. Neotoma is working closely with LinkedEarth to build import–export capabilities with the popular LiPD format. Data mobilization campaigns for fossil pollen data are underway, particularly in Africa, the Indo-Pacific, and Latin America. [a] new NSF-funded project is enhancing Neotoma support for both FAIR and CARE (Collective benefit, Authority to control, Responsibility and Ethics) data standards, in partnership with other paleodata resources. A new project is developing controlled vocabulary for lipid biomarker variables, with support from Belmont Forum, and another is partnering with a coalition of labs to support the curation of sedimentary (ancient) DNA datasets. Online and in-person educational workshops are planned, including at INQUA2023.

ACKNOWLEDGEMENTS

We thank Julien Emile-Geay, Frank Oliver Gloeckner, Carrie Morrill, Christian Rohr, and John (Jack) Williams for their responses.

AFFILIATIONS

[1]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA
[2]MARUM Center for Marine Environmental Sciences, University of Bremen, Germany

CONTACT

Darrell Kaufman: Darrell.Kaufman@nau.edu

REFERENCES

Khider D et al. (2022) Paleoceanogr Paleoclimatol 37(10): e2022PA004509

McKay NP et al. (2021) Geochron 3(1): 149-169

Figure 1: The paleoscience community is fortunate to be supported by discipline-specific data repositories and organizations which help assure that our data are FAIR – findable, accessible, interoperable, and reusable.