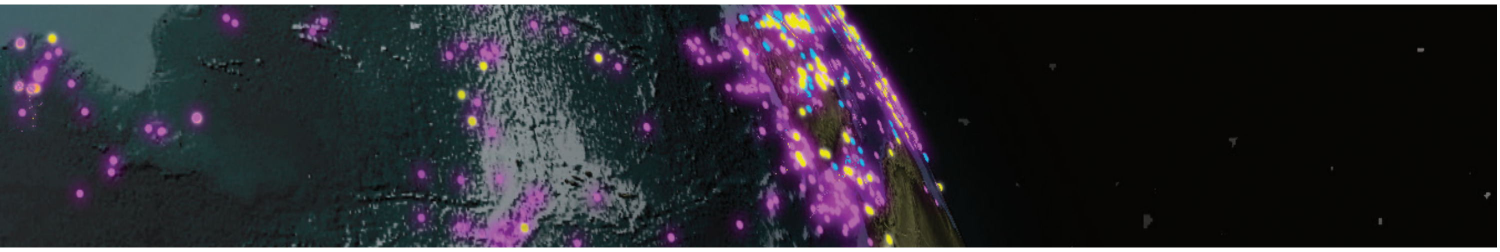
doi.org/10.22498/pages.26.2.49

Building and harnessing open paleodata

John W. Williams[1,2], D.S. Kaufman[3], A. Newton[4,5] and L. von Gunten[6]



Open data in the paleogeosciences have a long and fruitful history. Many of the primary open-data resources in the paleoenvironmental sciences are now at least two decades old, including the NOAA World Data Center for Paleoclimatology (Gross et al., p. 58), PANGAEA (Diepenbroek, p. 59), Paleoclimate Modelling Intercomparison Project (PMIP, Peterschmitt et al., p. 60), and the Paleobiology Database (Uhen et al., p. 78), all founded in the 1990s, with others, such as the Neotoma Paleoecology Database (Grimm, p.64), tracing their roots to constituent databases from this era and to influences spanning the last century. Indeed, this special issue can be viewed as a 20th-anniversary celebration of the 1998 "*Paleodata*" issue of PAGES news (the former name of *Past Global Changes Magazine*) that established many of the advances reviewed here (PAGES IPO 1998).

The history of open data in the paleogeosciences is long because the scientific motivation is so clear and unambiguous. In the large, complex, and ever-changing Earth system, scientific insight requires the open availability and close integration of multiple observational systems with Earth system models, to better understand the past and present, and better forecast the future (Crucifix 2012; Dietze et al. 2018). And, as the Great Acceleration continues (Steffen et al. 2015), such efforts have increased urgency; the past offers a uniquely important set of model systems for the strange new world of the coming decades.

Over these last two decades of open data, much has changed. The dividing line between “data generator” and “data user”, so apparently bright in the 1990s (PAGES Scientific Steering Committee 1998), has blurred as a new generation has arisen, with cross-over expertise in data generation, synthesis, and modeling. The information revolution races on, with the data sciences emerging both as a distinct academic discipline (Blei and Smyth 2017) and as a key employment opportunity for many scientists. Access to open-data resources is now essential to career advancement for early-career scientists, while lack of access to training is a key barrier (Koch et al., p. 54).

Contributing one's data to open-data resources, once largely voluntary, is now required by most journals, funders, and professional societies (Newton, p. 52; Belmont Forum, p. 56). The bar has been raised for open-data resources, to ensure that they meet the FAIR standards of Findable, Accessible, Interoperable, and Reusable (PAGES Scientific Steering Committee, p. 48; Gross, p. 58). New funding initiatives are being launched to increase the power and interoperability of existing data resources (e.g. NSF's EarthCube; Belmont Forum, p. 56), leading to new and flexible data standards and software that leverage and link open-data resources (Uhen et al., p. 78; McKay and Emile-Geay, p. 71). New geovisualization approaches such as Flyover Country, using open data and mobile technologies, are bringing paleodata to new audiences (Myrbo et al., p. 74). And, our understanding of data is changing as well, as we recognize that open data require ongoing curation and improvement, supported by community-curated data resources and linked networks of data stewards (Williams et al., p. 50).

These advances in open-data systems are opening up new scientific frontiers. Data-model assimilation, in which paleoenvironmental inferences from data and models are closely integrated, weighted by uncertainty, are active fields in paleoclimatology (Hakim et al., p. 73) and paleoecology (McLachlan and the PalEON Project, p. 76). Computer scientists are experimenting with artificial-intelligence approaches to age-model development (Bradley et al., p. 72) and extracting geological knowledge from the peer-reviewed literature (Marsicek et al., p. 70). Open paleodata have reached new audiences, as biogeographers and macroecologists combine the fossil record with big-data genetic repositories to study the processes governing the distribution and diversity of life (Fordham and Nogues-Bravo, p. 77), and as archaeologists bring big data to bear on the interplay between humans and the environment (Kohler et al., p. 68).

More needs to be done. Many key data remain "dark", requiring inordinate effort to gather and synthesize (Stenni and Thomas, p. 66). The paleoscience communities need to commit to conventions for reporting data and essential metadata, with shared adoption by scientists, data resources, publishers, and funding agencies. Established open-data resources need commitments of sustained support from funding agencies, with opportunities to build new data resources or extend existing data models to serve new kinds of data and science. The recent advances in assigning digital object identifiers (DOIs) to datasets needs to be more fully leveraged so that data generators are appropriately credited for data use. Scientific data services are needed that better streamline the passing of data from individual labs to community data resources. And, most of all, we need better integrated training programs in paleoscience and data science, to train the next generation of cross-over scientists.

In short, these are exciting and changing times. This special issue is more progress report than final authority. Nevertheless, we hope that the articles enclosed will provide useful information about the latest updates from some of the major open-data resources in the paleogeosciences, the efforts to build new resources and interlink existing resources, the emergence of new software and science powered by open data, and the ever-evolving interplay among cultural norms, technological advances, and scientific discovery.

AFFILIATIONS

[1]Department of Geography, University of Wisconsin-Madison, USA
[2]Neotoma Paleoecology Database
[3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA
[4]Nature Geosciences Editorial Office, London, UK
[5]Geological Society of London, UK
[6]PAGES International Project Office, Bern, Switzerland

CONTACT

John (Jack) W. Williams: jww@geography.wisc.edu

REFERENCES

Blei DM, Smyth P (2017) PNAS 114: 8689-8692

Crucifix M (2012) Quat Sci Rev 57: 1-16

Dietze MC et al. (2018) PNAS 115: 1424-1432

PAGES IPO (1998) PAGES news 6(2)

PAGES Scientific Steering Committee (1998) PAGES news 6(2): 1-2

Steffen W et al. (2015) The Anthropo Rev 2: 81-98