A SEMANTIC DATABASE OF TEMPERATURE PROXIES COVERING THE COMMON ERA

Julien Emile-Geay¹, Nicholas P. McKay², Jianghao Wang¹, Darrell Kaufman² & PAGES2k consortium

Abstract—Reconstructions of surface temperature over the past 2,000 years extend our knowledge of climate system behavior beyond the instrumental era, helping to distinguish between exogenous and endogenous sources of climate variability, a fundamental frontier of climate science. In this study, we describe the latest incarnation of the PAGES 2k global multi-proxy database, a multi-proxy, community-curated pool of paleoclimate records. The database is structured as Linked Open Data using a JSON-LD container, allowing for semantic relations to be discovered between its objects and other Linked Data. We describe elementary statistical analyses possible with this new data resource, present a reconstruction of global surface temperature via Markov random fields, and encourage experimentation via other forms of machine learning and artificial intelligence.

I. MOTIVATION

Low-frequency climate variability is crucial to adaptation and planning, and can only be adequately constrained by paleoclimate observations. Today, however, the vast majority of paleoclimate datasets are in disparate formats, incommensurate with each other and with climate model output. This fundamentally limits our ability to use them for the validation of Earth system models and, hence, for effective decision-making. It is therefore essential to bring all relevant observations into a consistent format. Two ingredients have made this possible:

1) PAGES2k¹, a community-driven effort to synthesize all publicly-archived, temperature-sensitive proxy records of the past 2,000 years [1].
2) LiPD, a new container designed to make paleoclimate data intelligible to machines [2]

The Linked Paleo Data (LiPD) format uses a JSON-LD framework to flexibly annotate metadata, explicating domain-specific definitions via a context file. The data themselves are stored in .csv files.

Corresponding author: J. Emile-Geay, Department of Earth Sciences, University of Southern California, Los Angeles, CA²
School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ ²
¹http://www.pages-igbp.org/ini/wg/2k-network/data

II. DATABASE SYNOPSIS

The database presently contains 724 records from 667 locations around the globe, spanning all or part of the past 2,000 years, with resolution going from monthly to centennial (Fig 1). There are 11 proxy types and dozens of measurement types. Records were selected by many volunteers worldwide, based on the following criteria:

Relation to temperature
The dataset includes proxy records for which statistical or mechanistic evidence of temperature exists

Duration
≥ 500y for non-annually resolved archives, 300y for terrestrial archives, 50y for annual marine archives.

Chronological accuracy
For non-annually resolved records, primary chronological information was archived to enable age modeling.

Resolution
≥ 1 data point every 50 years on average (except marine sediments, for which 200 years is the minimum average sample interval).

Fig. 1. Spatiotemporal data availability in the PAGES2k database
models, establishing relations to other climate fields (e.g. drought indices) or to (pre-)historical events. Once published [8] the data will be available as Linked Open Data, allowing for web-based discovery and linkages to other datasets.

ACKNOWLEDGMENTS

Funding for the authors was provided by NSF grants AGS-1003818, EAR-1347221 and ICER-1541029.

REFERENCES