Session: C21 Poster: T246A

Enabling data-intensive science at NOAA's World Data Center for Paleoclimatology

Wendy Gross[†]; David Anderson; Bruce Bauer; Rodney Buckner; Gille Edward; Michael Hartman; Carrie Morrill; Anju Shah; Eugene Wahl
[†] National Climatic Data Center, USA
Leading author: Wendy.Gross@NOAA.gov

Data-intensive science is a global trend in data reuse and management. It uses large volumes of data representing work by thousands of researchers world-wide. NOAA's National Climatic Data Center's World Data Center for Paleoclimatology collects data from scientists working at institutes around the world, and makes the data available for climate research on topics ranging from global trends to regional decadal variability to abrupt climate change. The majority of the data are peer-reviewed and published, and represent the results of research by over 2,500 scientists. Compared to other sources of climate data, paleoclimate data are unusually diverse, yielding thousands of variables originating from greater than a dozen different proxy types. Spatial scales range from local to global in scope, and the time spans the present day to 60 million years before present. To provide and harvest this diverse collection, we employ the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The GCMD NASA Directory Interchange Format (DIF) metadata standard provides a lowest-common-denominator content standard among Paleoclimate metadata providers and harvesters, and cross-walks among NOAA and other federal metadata requirements. Site-level granularity of the metadata enables mapping via GIS interfaces including ArcGIS, Google Maps/Earth, IDL, and Java animations. The metadata includes information needed to create thematic presentations for topics including the last millennium reconstructions, abrupt climate change, and hydrology, allowing researchers to explore and access data at the site level. Aggregation, rather than sub-setting, is the prevalent user-need for these small data sets that often consist of a single time series. All of the data are shared via FTP and World Wide Web protocols. To promote interoperability, we are migrating some data towards NetCDF with Climate and Forecast (CF) conventions where possible. Gridded proxy reconstructions, observation networks, and climate model output, are all easily distributed in NetCDF.