

EGU2020-22155

<https://doi.org/10.5194/egusphere-egu2020-22155>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Climate Index Metadata and its Implementation

Klaus Zimmermann and Lars Barring

SMHI, Rosaby Centre, Norrköping, Sweden (klaus.zimmermann@smhi.se)

Climate indices play an important role in the practical use of climate and weather data. Their application spans a wide range of topics, from impact assessment in agriculture and urban planning, over indispensable advice in the energy sector, to important evaluation in the climate science community. Several widely used standard sets of indices exist through long-standing efforts of WMO and WCRP Expert Teams (ETCCDI and ET-SCI), as well as European initiatives (ECA&D) and more recently Copernicus C3S activities. They, however, focus on the data themselves, leaving much of the metadata to the individual user. Moreover, these core sets of indices lack a coherent metadata framework that would allow for the consistent inclusion of new indices that continue to be considered every day.

In the meantime, the treatment of metadata in the wider community has received much attention. Within the climate community efforts such as the CF convention and the much-expanded scope and detail of metadata in CMIP6 have improved the clarity and long-term usability of many aspects of climate data a great deal.

We present a novel approach to metadata for climate indices. Our format describes the existing climate indices consistent with the established standards, adding metadata along the lines of existing metadata specifications. The formulation of these additions in a coherent framework encompassing most of the existing climate index standards allows for its easy extension and inclusion of new climate indices as they are developed.

We also present Climix, a new Python software for the calculation of indices based on this description. It can be seen as an example implementation of the proposed standard and features high-performance calculations based on state-of-the-art infrastructure, such as Iris and Dask. This way, it offers shared memory and distributed parallel and out-of-core computations, enabling the efficient treatment of large data volumes as incurred by the high resolution, long time-series of current and future datasets.