Data processing levels in the Earth System Sciences

Revision	Authors
Jan-2019	Karsten Peters (DKRZ) Heinke Höck (DKRZ)

The primary focus of climate data preservation in dedicated long-term archives (e.g. DKRZ LTA-WDCC[1] or PANGAEA[2]) is to enable the future reuse of the preserved datasets. Broad reuse of these datasets only becomes feasible when they are accompanied with an extensive set of metadata describing their heritage, postprocessing and usage restrictions. This holds for numerical model output as well as for observational data.

Earth Observation Data in particular require meticulous postprocessing before the geophysical quantity of interest including its geolocation is distilled from the instrument raw data. Data from all stages in this processing chain are suitable for long-term preservation and later reuse. Therefore, an explicit description of the data processing status/level is highly recommended when supplying such data for preservation.

Similar to the common practice of describing data processing levels of e.g. satellite remote sensing datasets, it is recommended that the metadata accompanying preserved observational datasets include an indication of the data processing level.

The data processing level definitions described here are generally formulated following those provided by large Earth Observation Data providers, e.g. NASA[3]:

Level 0:   Level 0 data are unprocessed instrument raw data at the instrument's space and time resolution. All available supplemental information needed for further postprocessing is included in or appended to or supplied with the data files.

Level 1:   Level 1 data are derived from Level 0 data which were processed to values of sensor units, e.g. brightness temperatures, radiances or backscatter ratios. Level 1 data are available at the instrument's full time/space resolution. All supplemental information needed for further processing is included in or appended to or supplied with the data files. Level 1 data may already be geolocated.

---

[1] https://www.dkrz.de/services/long-term-archiving-1/long-term-archiving?set_language=en&cl=en
[2] https://pangaea.de
[3] https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products

Level 2:   Level 2 data are derived from Level 1 data that have been processed to geophysical quantities of interest, e.g. wave height, cloud fraction, aerosol optical depth or radiation at the top of atmosphere. Uncertainty estimates of the retrieved parameters shall be included in or appended to or supplied with the data files.

Level 3:   Level 3 data are derived from Level 2 data that have been spatially and/or temporally resampled, e.g. monthly averages on a global 1°x1° lat-lon grid. Uncertainty estimates of the resampled data shall be included in the data files.

Level 4:   Further derived products, e.g. climate indices, obtained from analysis of multiple lower-level products. Uncertainty estimates, if feasible to compute, may be included in the data files.