Geophysical Research Abstracts Vol. 21, EGU2019-14982, 2019 EGU General Assembly 2019 © Author(s) 2019. CC Attribution 4.0 license.



SuperMegaHyperCubes: earth system data discovery, aggregation and analysis with Iris and Pangeo

Peter Killick

Met Office Informatics Lab, United Kingdom (peter.killick@informaticslab.co.uk)

As data volumes grow, we more and more need to take processing to the data than rely on local copies: not only is it often infeasible to process the data locally, but often the data are too large to maintain a local copy. This situation holds true from single files on single machines to whole datasets on large-scale high performance compute clusters.

To meet this need, we present an implementation of the Pangeo ecosystem which leverages Iris, a mature, open-source Python library for analysis and visualisation of Big Earth Data. We will particularly focus on recent work that has integrated dask into the very core of Iris to provide scalable, out-of-core processing of Data Cubes from the smallest to the largest of sizes. We show how we can use this platform to generate and perform analysis on single contiguous orthogonal hypercubes of earth system data constructed from multiple overlapping input datasets – data objects we refer to as SuperMegaHyperCubes.