Data near processing support for climate data analysis

Stephan Kindermann (1), Carsten Ehbrecht (1), and Nils Hempelmann (2)

(1) German Climate Computing Centre - DKRZ, Hamburg, Germany (kindermann@dkrz.de), (2) Institut Pierre Simon Laplace, Paris, France

Climate data repositories grow in size exponentially. Scalable data near processing capabilities are required to meet future data analysis requirements and to replace current "data download and process at home"
workflows and approaches.
On one hand side, these processing capabilities should be accessible via standardized interfaces (e.g. OGC WPS), on the other side a large variety of processing tools, toolboxes and deployment alternatives
have to be supported and maintained at the data/processing center.

We present a community approach of a modular and flexible system supporting the development, deployment and maintenace of OGC-WPS based web processing services. This approach is organized in an open source github project (called "bird-house") supporting individual processing services ("birds", e.g. climate index calculations, model data ensemble calculations), which rely on basic common infrastructural components (e.g. installation and deployment recipes, analysis code dependencies management).
To support easy deployment at data centers as well as home institutes (e.g. for testing and development) the system supports the management of the often very complex package dependency chain of climate data analysis packages as well as docker based packaging and installation.

We present a concrete deployment scenario at the German Climate Computing Center (DKRZ). The DKRZ one hand side hosts a multi-petabyte climate archive which is integrated e.g. into the european ENES and worldwide ESGF data infrastructure, and on the other hand hosts an HPC center supporting (model) data production and data analysis. The deployment scenario also includes openstack based data cloud services to support data import and data distribution for
bird-house based WPS web processing services. Current challenges for inter-institutionnal deployments of web processing services supporting
the european and international climate modeling community as well as the climate impact community are highlighted. Also aspects supporting future WPS based cross community usage scenarios supporting data reuse and data provenance aspects are reflected.