



Scientific Platform as a Service - Tools and solutions for efficient access to and analysis of oceanographic data

Aleksander Vines, Morten W. Hansen, and Anton Korosov

Nansen Environmental and Remote Sensing Center, Bergen, Norway (aleksander.vines@nersc.no)

Existing infrastructure international and Norwegian projects, e.g., NorDataNet, NMDC and NORMAP, provide open data access through the OPeNDAP protocol following the conventions for CF (Climate and Forecast) metadata, designed to promote the processing and sharing of files created with the NetCDF application programming interface (API). This approach is now also being implemented in the Norwegian Sentinel Data Hub (satellittdata.no) to provide satellite EO data to the user community. Simultaneously with providing simplified and unified data access, these projects also seek to use and establish common standards for use and discovery metadata. This then allows development of standardized tools for data search and (subset) streaming over the internet to perform actual scientific analysis. A combination of software tools, which we call a Scientific Platform as a Service (SPaaS), will take advantage of these opportunities to harmonize and streamline the search, retrieval and analysis of integrated satellite and auxiliary observations of the oceans in a seamless system.

The SPaaS is a cloud solution for integration of analysis tools with scientific datasets via an API. The core part of the SPaaS is a distributed metadata catalog to store granular metadata describing the structure, location and content of available satellite, model, and in situ datasets. The analysis tools include software for visualization (also online), interactive in-depth analysis, and server-based processing chains. The API conveys search requests between system nodes (i.e. interactive and server tools) and provides easy access to the metadata catalog, data repositories, and the tools. The SPaaS components are integrated in virtual machines, of which provisioning and deployment are automatized using existing state-of-the-art open-source tools (e.g., Vagrant, Ansible, Docker). The open-source code for scientific tools and virtual machine configurations is under version control at <https://github.com/nansencenter/>, and is coupled to an online continuous integration system (e.g., Travis CI).