

EGU2020-17708

<https://doi.org/10.5194/egusphere-egu2020-17708>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



A Web-Based Virtual Research Environment for Marine Data

Merret Buurman¹, Sebastian Mieruch², Alexander Barth³, Charles Troupin³, Peter Thijsse⁴, Themis Zamani⁵, and Naranyan Krishnan⁶

¹German Climate Computing Centre (DKRZ), Data Management, Germany (buurman@dkrz.de)

²Alfred-Wegener-Institut (AWI), Germany

³GHER, Université de Liège, Belgium

⁴MARIS, The Netherlands

⁵Greek Research and Technology Network (GRNET), Greece

⁶Science and Technology Facilities Council (STFC), United Kingdom

Like most areas of research, the marine sciences are undergoing an increased use of observational data from a multitude of sensors. As it is cumbersome to download, combine and process the increasing volume of data on the individual researcher's desktop computer, many areas of research turn to web- and cloud-based platforms. In the scope of the SeaDataCloud project, such a platform is being developed together with the EUDAT consortium.

The SeaDataCloud Virtual Research Environment (VRE) is designed to give researchers access to popular processing and visualization tools and to commonly used marine datasets of the SeaDataNet community. Some key aspects such as user authentication, hosting input and output data, are based on EUDAT services, with the perspective of integration into EOSC at a later stage.

The technical infrastructure is provided by five large EUDAT computing centres across Europe, where operational environments are heterogeneous and spatially far apart. The processing tools (pre-existing as desktop versions) are developed by various institutions of the SeaDataNet community. While some of the services interact with users via command line and can comfortably be exposed as JupyterNotebooks, many of them are very visual (e.g. user interaction with a map) and rely heavily on graphical user interfaces.

In this presentation, we will address some of the issues we encountered while building an integrated service out of the individual applications, and present our approaches to deal with them.

Heterogeneity in operational environments and dependencies is easily overcome by using Docker containers. Leveraging processing resources all across Europe is the most challenging part as yet. Containers are easily deployed anywhere in Europe, but the heavy dependence on (potentially shared) input data, and the possibility that the same data may be used by various services at the same time or in quick succession means that data synchronization across Europe has to take place at some point of the process. Designing a synchronization mechanism that does this without conflicts or inconsistencies, or coming up with a distribution scheme that minimizes the

synchronization problem is not trivial.

Further issues came up during the adaptation of existing applications for server-based operation. This includes topics such as containerization, user authentication and authorization and other security measures, but also the locking of files, permissions on shared file systems and exploitation of increased hardware resources.