The SeaDataCloud Virtual Research Environment: the technical perspective

Merret Buurman, German Climate Computing Centre (DKRZ) (Germany), buurman@dkrz.de
Sri Harsha Vathsavayi, CSC IT Center for Science (Finland), sriharsha.vathsavayi@csc.fi
Peter Thijsse, Mariene Informatie Service 'MARIS' BV (The Netherlands), peter@maris.nl
Alexander Barth, GHER, Université de Liège (Belgium), a.barth@uliege.be
Sebastian Mieruch, Alfred Wegener Institut (Germany), sebastian.mieruch@awi.de
Nikolaos Evangelou, Greek Research and Technology Network S.A. (Greece), nikosev@admin.grnet.gr
Marco Rorro, Consorzio Interuniversitario CINECA (Italy), m.rorro@cineca.it
Adrian Coveney, Science and Technology Facilities Council, (United Kingdom), adrian.coveney@stfc.ac.uk

Virtual Research Environments (VREs) allow researchers to execute complex workflows of data-driven experiments in the cloud, thus reducing data transfers and leveraging remote computational facility. The VRE developed by SeaDataNet and EUDAT will allow users to use various processing services on data from a variety of sources. To accommodate this diversity, clear interactions between services are defined. Docker containers make it possible to run very diverse services together in one environment, and to update and migrate services without interfering with other services. Also, this modularity allows to easily scale and extend the services.

With the growth of data size and variety, and processing complexity, Virtual Research Environments (VREs) are getting more and more popular. VREs, also called Science Gateways or Virtual Labs, allow researchers to compose and execute complex workflows of data-driven experiments including finding data, accessing data, processing iteratively data with various tools, visualizing results, sharing relevant findings with colleagues, replicating experiments/workflows and publishing insights online, thus reducing data transfers and leveraging remote computational facility.

SeaDataNet is the primary entry point for many marine scientists to find and access data. After the recent move of SeaDataNet's data to the cloud, offering processing capacities in the cloud too is a logical next step. That is why SeaDataNet is now developing a Virtual Research Environment for their users. It is being developed in close collaboration between the SeaDataNet consortium and the EUDAT research infrastructure, represented by the scientific data and computing centres CSC (Finland), DKRZ (Germany), CINECA (Italy), GRNET (Greece) and STFC (United Kingdom).

This abstract aims to show the more technology-curious public a view behind the curtains of how the VRE is realized and what advantages this setup brings.

The SeaDataNet VRE will allow its users to perform resource-intensive analyses on the cloud. Applications that are commonly used in the marine sciences, such as AWI's OceanDataView (Schlitzer, 2018), Jupyter Notebooks and University of Liège's DIVA (Data-Interpolating Variational Analysis (Barth *et al.*, 2014) to create gridded climatologies from in situ observations), will be available as VRE services. Additional services such as notifying users of outdated datasets,

enabling users to communicate while working on a common task, and chaining services will be added for the users' convenience.

Many of the applications that will be included in the VRE already exist, they are written in different languages, have different user interfaces, different architectures and different dependencies. Thus, the VRE has to be implemented in a way that allows applications that are quite different to be integrated and interoperable, and to add more applications later on. To fulfill these requirements, SeaDataNet has come up with a flexible architecture that is described in the next sections.

To accommodate the very diverse applications that already exist, to easily integrate new ones, and to avoid dependency conflicts, all the applications will run inside Docker containers. This approach will also improve environment portability, scalability, security and scientific reproducibility. Application containers will run on the EUDAT's computing service B2HOST, which provides scalable processing capability, fast access to storage volumes, and scheduling of containers execution. Every application exposes its own service interface, which could be a GUI, a command line interface in a Jupyter Notebook and/or a REST API.

The application containers (i.e., services) interface with the data layer in the backend, with the user interfaces in the frontend, and with the controller.

The frontend will be developed as a responsive JavaScript layer, based on state-of-the-art JavaScript libraries. It will be responsible for authentication and authorization of the users, and for routing them to the interfaces of the services. Services that are common to all processing services can also be accessed from here, such as dataset management component, process chaining (workflows), user communication, and version notification.

For the time being, the main data backend of the VRE will be the EUDAT service B2DROP. B2DROP allows users to upload, store and securely share datasets. It will be accessible to all the

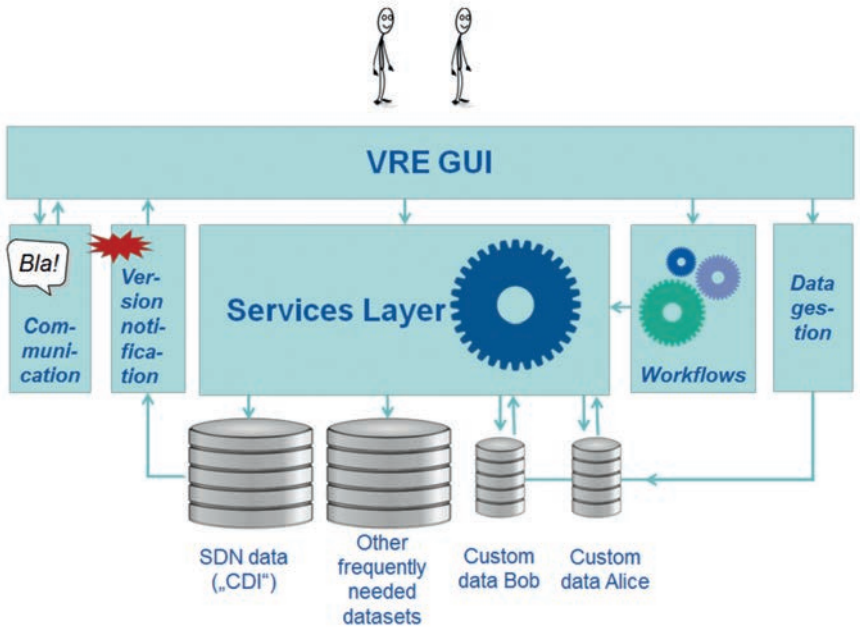

Fig. 1 - Architecture of the Virtual Research Environment.

applications by the WebDAV protocol, allowing the applications to interact with the B2DROP content as a file system.

Behind the frontend, a controller makes up the heart of the VRE and glues all the components together. It sees to the mounting of the user's private data from EUDAT's nextcloud-based B2DROP service, mounting other data and volumes, launches the correct service containers on-demand and it is responsible for all inter-container communications.

Summarising, the SeaDataNet VRE will allow users to use a variety of processing services on data from a variety of sources. To accommodate this diversity, clear interactions between services are defined. Docker containers make it possible to run very diverse processing services together in one environment, and to update and migrate services without interfering with other services. Also, this modularity will allow to scale the services when necessary and easily add new services or extend the existing services.

References

Barth, A., Beckers, J.-M., Troupin, C., Alvera-Azcárate, A., and Vandenbulcke, L., 2014. divand-1.0: n-dimensional variational data analysis for ocean observations, Geosci. Model Dev., 7, 225-241, doi:10.5194/gmd-7-225-2014.

Schlitzer, R., 2018. Ocean Data View, https://odv.awi.de.

https://www.eudat.eu/services/userdoc/b2host.