

Working with the SeaDataCloud Virtual Research Environment: what can we do for you?

Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be

Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be

Leo Bruvry-Lagadec, IFREMER, Plouzané (France), Leo.Bruvry.Lagadec@ifremer.fr

Merret Buurman, DKRZ, Hamburg (Germany), buurman@dkrz.de

Seppo Kaitala, SYKE (Finland), Seppo.Kaitala@ymparisto.fi

Narayanan Krishnan, UKRI STFC (United Kingdom), narayanan.krishnan@stfc.ac.uk

Sebastian Mieruch, AWI, Bremerhaven (Germany), sebastian.mieruch@awi.de

Giorgio Santinelli, Deltares, Delft (The Netherlands), Giorgio.Santinelli@deltares.nl

Peter Thijsse, MARIS, Nootdorp (The Netherlands), peter@maris.nl

Filip Waumans, VLIZ (Belgium), filip.waumans@vliz.be

Themis Zamani, GRNET, Athens (Greece), themis@admin.grnet.gr

Motivations to build a Virtual Research Environment

For researchers, the conventional work on their desktop computer to analyse data and generate products is getting more and more complex due to the increasing amount of data to process. Three main hurdles often prevent them from performing the required operations:

1. Insufficient CPU power, leading to too long computation times
2. Lack of available memory to process large datasets
3. Insufficient disk storage, making the use of the full datasets impossible.

To overcome these, cloud services are becoming common practice, such as the European Research Data Infrastructure (EUDAT, <https://eudat.eu/>), the European Open Science Cloud (EOC, <https://www.eosc-portal.eu/>). With such services, no more downloading of data is needed: all the processing is performed *close* to the data. In addition, the cloud allows cooperative research on a common platform, with a guarantee that the latest versions of the software tools are installed, and enhances the reproducibility of the scientific work. In order to provide the oceanographic community with seamless access to the SeaDataNet data, standard and software tools, the SeaDataCloud (SDC) Virtual Research Environment (VRE) has been developed.

This paper describes the tools made available to users. The implementation and deployment are detailed in “*SeaDataCloud Virtual Research Environment: Implementation and technical aspects*”.

SeaDataCloud VRE tools

Users login with their MarineID (<https://www.marine-id.org/>) and have access to a dashboard that provides access to a private workspace as well as the different services described hereinafter.

webODV consists of a suite of online services based on Ocean Data View (ODV, <https://odv.awi.de/>), designed to interactively perform analysis, exploration and visualization of ocean data. webODV allows users to aggregate large numbers of SeaDataNet data files and perform quality control. More details about webODV are presented in “webODV – operational and ready for the community” and “Ocean Data View goes Online”.

DIVAnd (Data Interpolating Variational Analysis in n dimensions; <https://github.com/gher-ulg/DIVAnd.jl>) is a cutting-edge software tool designed to efficiently interpolate in-situ observations onto a regular grid, in an arbitrary number of dimensions (for instance longitude, depth and time). A set of *Jupyter notebooks* (<https://github.com/gher-ulg/Diva-Workshops/>) provides a guideline to the user on how to prepare the data, optimise the analysis parameters and perform the interpolation.

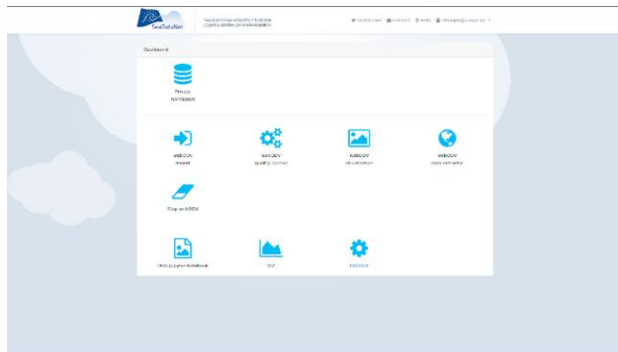
BioQC is a tool to process and to run quality control on biological datasets. BioQC helps researchers to evaluate whether a particular biological occurrence record within the input file is useful for their analysis. It also helps the data providers to identify possible gaps and errors in their datasets.

The tool returns the input file with quality information attached for each occurrence record and a detailed report. This result file will enable the users to filter for suitable records.

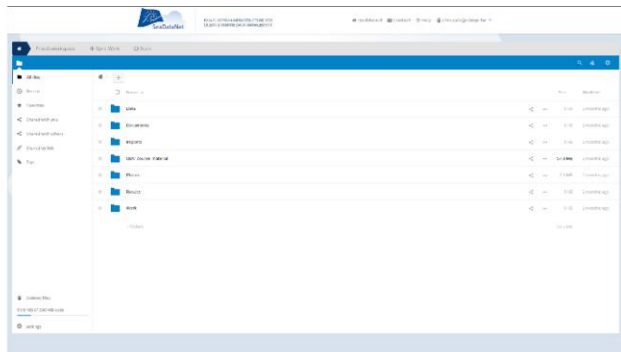
VIZ (<https://github.com/openearth/sdc-visualization>) is a modern and dynamic visualisation service to explore datasets on a map. By clicking on data points, the users see a plot of the full profile prepared with WebODV, and metadata of the input ODV files. A time selector permits to limit the data for the period of interest. Additionally, the visualisation service provides the possibility to explore 4D gridded products prepared with DIVAnd.

The **Subsetting service** strives to make data access easier, by providing services to subset, download and plot data. It returns datasets in various data formats such as CSV, MATLAB, netCDF, ODV and more. An interactive visualization tool allows users to discover and browse through the subset results with modern web technologies.

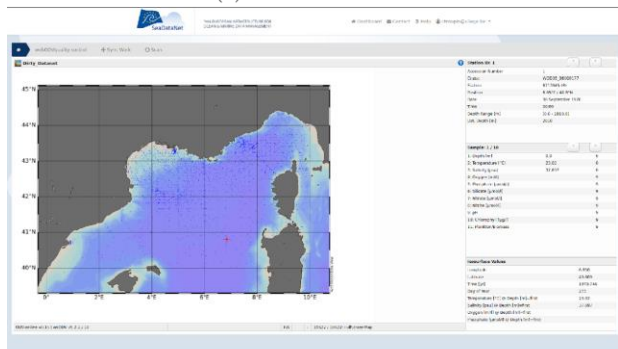
All together, these services make up the SeaDataCloud VRE, which we hope users will find useful and enjoy using.



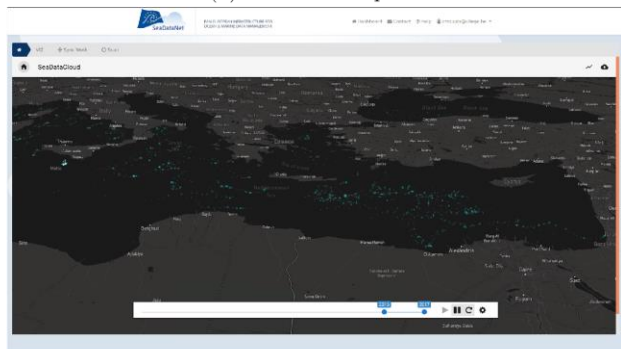
(a) Dashboard home.



(b) Private workspace.



(c) webODV Data quality control.



(d) Data visualisation.

Figure 1: Different components of the VRE.