

EGU2020-18477

<https://doi.org/10.5194/egusphere-egu2020-18477>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Using and acquiring time-series data with the EMSO ERIC DataLab

Raul Bardaji¹, Jaume Piera², Juanjo Dañobeitia^{1,3}, and Ivan Rodero³

¹Marine Technology Unit - CSIC, Barcelona, Spain

²Marine Science Institut - CSIC, Barcelona, Spain

³EMSO ERIC, Rome, Italy

In marine sciences, the way in which many research groups work is changing as scientists use published data to complement their field campaign data online, thanks to the large increase in the number of open access observations. Many institutions are making great efforts to provide the data following FAIR principles (findability, accessibility, interoperability, and reusability) and are bringing together interdisciplinary teams of data scientists and data engineers.

There are different platforms for downloading marine and oceanographic data and many libraries to analyze data. However, the reality is that scientists continue to have difficulty finding the data they need. On many occasions, data platforms provide information about the metadata, but they do not show any underlying graph of the data that can be downloaded. Sometimes, scientists cannot download only the data parameters of interest and have to download huge amounts of data with other not useful parameters for their studies. On other occasions, the platform allows to download the data parameters of interest but offers the time-series data as many files, and it is the scientist who has to join the pieces of data into a single dataset to be analyzed correctly. EMSO ERIC is developing a data service that helps reduce the burden of scientists to search and acquire data as much as possible.

We present the EMSO ERIC DataLab web application, which provides users with capabilities to preview harmonized data from the EMSO ERIC observatories, perform some basic data analyses, create or modify datasets, and download them. Use case scenarios of the DataLab include the creation of a NetCDF file with time-series information across EMSO ERIC observatories.

The DataLab has been developed using engineering best practices and trend technologies for big data management, including specialized Python libraries for web environments and oceanographic data analysis, such as Plotly, Dash, Flask, and the Module for Ocean Observatory Data Analysis (MOODA).