

EGU2020-20337

<https://doi.org/10.5194/egusphere-egu2020-20337>

EGU General Assembly 2020

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



Enhancing SeaDataNet e-infrastructure for ocean and marine data, new opportunities and challenges to foster data re-use

Leda Pecci¹, Michele Fichaut², and Dick Schaap³

¹ENEA, SSPT-PROTER-BES, Lerici, Italy (leda.pecci@enea.it)

²IFREMER, Brest, France (michele.fichaut@ifremer.fr)

³Marine Information Service MARIS. B.V, Nootdorp, Netherlands (dick@maris.nl)

Marine and ocean data represent a significant resource that can be used to improve the global knowledge of the seas. A huge amount of data is produced every day by ocean observations all around Europe. The ability to leverage this valuable potential depends on the capacity of the already established European (EU) ocean data infrastructures to support new needs in the field of ocean data management and to adopt the emerging technologies.

The SeaDataNet e-infrastructure (<https://www.seadatanet.org>), built up in early 2000 years, plays an important role for marine scientists and other ocean stakeholders communities, giving access to more than 2.2 million multidisciplinary harmonised marine and ocean data sets coming mainly from the European seas collected by more than 110 data centers, and offering data products and metadata services. Thanks to the 4-year SeaDataCloud Horizon 2020 project, started the 1st of November 2016, the development of a more efficient electronic infrastructure, kept up with the times and offering new services, based on the cloud and High Performance Computing (HPC) technologies, was addressed. It has renewed the original SeaDataNet Information Technology (IT) architecture. The collaboration with the EUDAT consortium, composed of a number of research communities and large European computer and data centres, enabled the migration of the data storage and services into the cloud environment, new instruments, such as High-Frequency Radar (HFR), Flow Cytometer and Glider data, have been standardised in agreement with the respective user communities. Furthermore, a Virtual Research Environment will support research collaboration.

SDN infrastructure is focused on historical digital ocean data and also supports the management of data streams from sensors based on the Sensor Web Enablement (SWE) standards of the Open Geospatial Consortium (OGC).

Harmonisation of ocean data allows more countries to be able to use data for scientific research and for decision-making purpose but data re-use is related also to the trust that the ocean scientific community places in the data. The latter issue involves a well-defined process of data quality checks. In SDN, data producers have to label each individual measurement with a value according to the SDN Quality Check (QC) Flags, and they follow specific procedures presented in the SDN-QC guideline (<https://www.seadatanet.org/Standards/Data-Quality-Control>). Furthermore,

a range of checks are carried out on the data, as part of the process of data products generation to improve the overall quality.

A relevant issue that limits data re-use is that some researchers are reluctant to share their own data, the push to encourage them it is to give them the right acknowledgment for the work done by means of the data citation, for this reason from the SDN portal a Digital Object Identifier (DOI) minting service is freely available for every data producer that shares their data. In addition, data versioning is available on the cloud platform for reproducible analysis.

The SeaDataCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 730960.