The European Multidisciplinary Seafloor and Water Column Observatory to explore the Oceans

Table of Contents

About EMSO ERIC

The **European Multidisciplinary Seafloor and water-column Observatory (EMSO)** is a **European Research Infrastructure Consortium (ERIC)** that includes open-ocean, seafloor observatories down to 4.850 metres depth, and shallow-water test sites from the Northeast Atlantic, across the Mediterranean to the Black Sea (http://emso.eu/).

EMSO ERIC acquires high-quality environmental data and represents a major asset for researchers who have access to multidisciplinary data to respond to pressing scientific and societal challenges. These data cover a multi- and inter-disciplinary range of research areas including biology, geology, chemistry, physics, engineering and computer science, from polar to tropical environments, down to the abyss, and allow researchers to perform investigations of the interactions among geosphere, hydrosphere and biosphere. The data generated in EMSO ERIC allow facing multivariate questions over different space and time scales, overcoming the traditional approach of focusing on single data streams.

The ERIC was established in 2016 with the goal of ensuring long-term, sustained, continuous data streams from the ocean, the majority of the biosphere of our planet.

The Challenge

EMSO ERIC consists of a system of **regional facilities** placed at key sites around Europe, from North East to the Atlantic, through the Mediterranean, to the Black Sea. Observatories are platforms equipped with multiple sensors, placed along the water column and on the seafloor. They constantly measure different biogeochemical and physical parameters that address natural hazards, climate change and marine ecosystems.

In a nutshell, the challenge **EMSO ERIC** is trying to solve is to **provide access to curated data and offer high-quality services** to a large and diverse group of users, from scientists and industries to institutions and policy makers, **for defining environmental policies based on scientific data**. A fundamental technical component of the EMSO ERIC cyber-infrastructure, that allows the integration of data from EMSO regional facilities where the observatories are deployed, is the **EMSO ERIC Data Platform**. This Data Platform has been designed to be scalable, flexible and able to provide high quality data products for a growing set of stakeholders in a wide range of disciplines. From the technical perspective, the platform ingests, consolidates, processes and archives data, integrates the data management architectures of the regionally distributed EMSO nodes and makes data available to the community.

To support this challenge, **EGI contributed to operate different EMSO ERIC services** including the data portal which received thousands of visits from different countries.

The EMSO ERIC Data Management Platform

The overall architecture of the EMSO ERIC Data Management Platform is shown in figure 1:

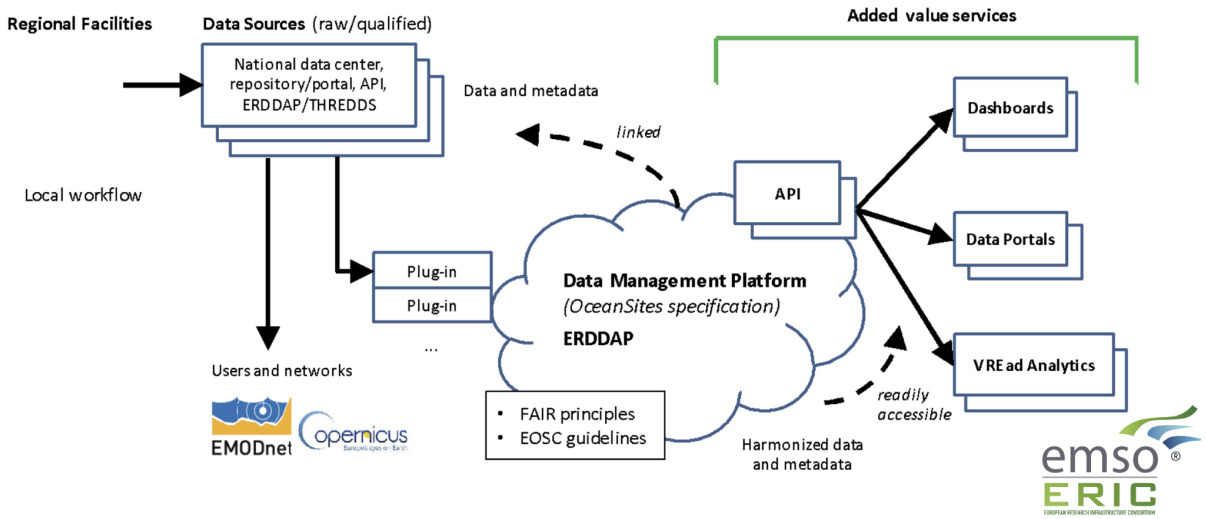

Figure 1. The architecture of the EMSO ERIC Data Management Platform

The data workflow starts at the regional facilities that operate the observatories. The data collected from the sensors is curated and made available through a data source that can be a National data center, repositories, APIs, and tools such as ERDDAP.

In order to deliver added-value services, the EMSO ERIC Data Management Platform harmonizes the data sets following Oceansites specifications, FAIR principles, and EOSC guidelines.

The resulting harmonized data will be offered via REST APIs for building services such as data portals, dashboards, or analysis tools.

EGI services used by EMSO ERIC

The EMSO ERIC Data Management Platform is using the following EGI services:

- The EGI Cloud Compute and the cloud-based EGI Online Storage to distribute the computations.
- The EGI Check-In to enable users' registration and authentication mechanisms.

Support from EGI

The collaboration with EGI started back in 2017 when 4 cloud providers of the EGI Federation agreed to support the EMSODEV project during the design and implementation of the EMSODEV Data Portal. In total, the **4** providers offered **9TB of storage capacity and about 340 vCPU cores**. This collaboration continued in the EOSC-hub project where a dedicated Competence Centre across the NGIs, involving user experts in the scientific domain, resource centres, and technology providers of the EGI Collaboration was set-up with the goal to further support the community needs.

From a technical perspective, through the Competence Centre, EGI delivered expertise and support on a variety of technical areas including platform development, data migration, and training on the EGI advanced services.
The EOSC-hub project has been instrumental for having continuity to this effort and to transition the Data Management Platform to pre-production. The cloud-based resources provided from two geo-distributed datacenters in Italy (INFN-CLOUD-BARI) and Spain (CESGA) belonging to the EGI Federation, guaranteed by an SLA, have been used to support three environments:

- **Development/test site** that provides an environment for software evolution and testing, including configuration management, continuous integration and functional testing,
- **Core site** that supports the EMSO ERIC Data Portal software stack, including back-end processes and data services exposed to users; and
- **Backup core site** that represents a mirror of the core site for system resiliency and business continuity, including data and services mirroring and fail-over capabilities.

In the context of EGI-ACE the agreement with the two providers has been recently renewed until **June 2023** in order to continue the operation of the platform as one of the EGI-ACE Data Space providers.

Resource providers supporting EMSO ERIC

A Service Level Agreement with the following providers was agreed with EMSO ERIC:

- CESGA, Spain member of ES-NGI, and
- INFN-CLOUD-BARI, Italy, member of INFN

More specifically, INFN-CLOUD-BARI is used as a primary data centre to deploy the Data Management Platform (production, development and testing environments). INFN-CLOUD-BARI provided **300 vCPU cores, 1.2TB of RAM and 10TB of block storage**.

In CESGA are replicated services that are used in fail-over for disaster recovery and business continuity. The total resources offered by CESGA amount to: **192 vCPU cores, 512GB of RAM and 600GB HDD of block storage**. For further details, about the collaboration agreement between EGI and EMSO ERIC, please check the EGI documents repository.

In 2020, the EMSO-ERIC data services deployed in the EGI Infrastructure were in operational status. During that time, the quality of service has been very close to 100% without any significant incident.

EMSO ERIC Data Portal in numbers

During the EOSC-hub project, the EMSO ERIC services operated using EGI resources such as the data portal received visits from more than one thousand distinct users from **85** countries. The countries with the largest number of visits include China, Italy, Spain, France, Greece, Portugal, United Kingdom, United States, Japan, and Germany.

Usage of the EGI computing resources

Over the last 2 years (Dec. 2019 - May 2021) EMSO-ERIC:

- Consumed more than **2,879,779** of (Cloud) CPU hours and,
- Instantiated **242** VMs in the EGI Federated Cloud infrastructure.

For more updated usage, please check the metrics reported in the EGI Accounting Portal.

More information

- **Website:** http://emso.eu/
- **EMSO ERIC Data Portal (in EOSC Marketplace):** https://marketplace.eosc-portal.eu/services/emso-eric-data-portal
- **EMSO ERIC SLA OLAs:** https://documents.egi.eu/document/3539
- **EGI news:** https://www.egi.eu/about/newsletters/the-emso-data-management-platform-from-prototype-to-full-production/