

BEACON - High performance data access supporting marine data lakes

Wednesday, 2 October 2024 11:15 (15 minutes)

In many of the societal and scientific challenges, such as Digital Twins of the Oceans and virtual research environments, fast access to a large number of multidisciplinary data resources is key. However, achieving performance is a major challenge as original data is in many cases organised in millions of observation files which makes it hard to achieve fast responses. Next to this, data from different domains are stored in a large variety of data infrastructures, each with their own data-access mechanisms, which causes researchers to spend much time on trying to access relevant data. In a perfect world, users should be able to retrieve data in a uniform way from different data infrastructures following their selection criteria, including for example spatial or temporal boundaries, parameter types, depth ranges and other filters. Therefore, as part of the EOSC Future and Blue-Cloud 2026 projects, MARIS developed a software system called 'BEACON' with a unique indexing system that can, on the fly with incredible performance, extract specific data based on the user's request from millions of observational data files containing multiple parameters in diverse units.

The BEACON system has a core written in RUST (low-level coding language) and its indexed data can be accessed via a REST API that is exposed by BEACON itself meaning clients can query data via a simple JSON request. The system is built in a way that it returns one single harmonised file as output, regardless of whether the input contains many different data types or dimensions. It also allows for converting the units of the original data if parameters are measured in different types of units (for this it e.g. makes use of the NERC Vocabulary Server (NVS) and I-Adopt framework).

EOSC-FUTURE Marine Data Viewer

Showcasing the performance and usability of BEACON, the BEACON system is applied to the SeaDataNet CDI database, Euro-ARGO and the ERA5 dataset from the Climate Data Store. These are also connected to a Marine Data Viewer that was developed as part of the EOSC-FUTURE project to co-locate Copernicus Marine satellite derived data products for Temperature and Salinity with observed in-situ data, made available through BEACON instances for the Euro-Argo and SeaDataNet marine data services.

The user interface of the Marine Data Viewer (<https://eosc-future.maris.nl/>) is designed to allow (citizen) scientists to interact with the data collections and retrieve parameter values from observation data. Enabled by the performance of BEACON, the user can filter the data on-the-fly using sliders for date, time and depth. At present, the ocean variables concern temperature, oxygen, nutrients and pH measurements, from Euro-Argo and SeaDataNet. The in-situ values are overlayed at the same time and space with product layers from Copernicus Marine, based upon modelling and satellite data.

Presentation

During the presentation more details will be given about the BEACON software and its performances. Moreover, latest developments will be presented, which includes deploying BEACON instances for several leading marine and ocean data repositories as part of Blue-Cloud 2026 to provide data lakes to the VRE user community and DTO.

Topic

Data innovations: Data Management/Integration/Exchange

Primary author: THIJSSSE, Peter (MARIS)

Co-authors: Mr KOOYMAN, Robin (MARIS); KRIJGER, Tjerk (MARIS)

Presenter: KRIJGER, Tjerk (MARIS)

Session Classification: Maximizing Data Efficiency and Availability: Innovations in Data Management Platforms