Contribution ID: **109** Type: **TALK**

Harmonizing Marine Data for Real-Time Ingestion into Digital Twins of the Ocean (DTOs): An Open Data Infrastructure Approach

Monday 4 November 2024 12:05 (20 minutes)

The study of climate change and its impact on marine environments requires large-scale, multidisciplinary data that are often collected by various national and marine institutes, fishery associations, as well as by research groups. With the proliferation of underwater observatories, profilers, and autonomous underwater vehicles (AUVs), significant progress has been made in collecting continuous, high-resolution data for in-situ ecological monitoring. However, much of this data remains static and stored in formats such as NetCDF or CSV, making it difficult to integrate into dynamic DTO systems. Furthermore, distribution shifts—variations in the data due to differing collection methods or environmental conditions—pose significant challenges for AI-based systems, which rely on consistent and harmonized data for training and prediction.
Archived and ecological monitoring network data from in-situ robotic and other scientific and societal sources, while essential, are highly heterogeneous and encoded in different formats, posing significant challenges for harmonization and integration. In this context, the Digi4Eco Project (https://digi4eco.eu/the-project/) focuses on addressing the lack of tools to effectively harmonize this vast amount of marine data, ensuring a suitable format for ingestion into DTOs. To address these challenges, our work focuses on developing a comprehensive Open Data Infrastructure (ODI) that adheres to FAIR principles: Findable, Accessible, Interoperable, and Reusable. The ODI will harmonize data typologies, procedures, and instrument specifications, making it easier to process and feed into DTO systems. This pipeline will include automated data validation and quality control mechanisms, following best practices. Special attention will be given to ensuring the data is suitable for AI applications, particularly in solving distribution shift problems.
The proposed ODI integrates existing open-source data services into a modular architecture that covers the entire data and metadata lifecycle. Key components include the SensorThings API for data/metadata storage, ERDDAP for data delivery, Zabbix for system monitoring and alerting, and Grafana for visualizations. Additionally, to ensure long-term impact and community adoption, all procedures and tools developed within this framework will be made open-source and publicly accessible, fostering standardized practices across the marine research community.

Please specify ”other”

In addition, please add 3 to 5 keywords.

DTOs, marine data harmonization, data interoperability, quality control

Please specify ”other”

For whom will your contribution be of most interest?

Data professionals who provide and maintain data infrastructure

Please assign yourself (presenting author) to one of the following groups.

Researchers

Primary authors: Dr MARTÍNEZ, Enoc (SARTI-UPC); Dr MIHAI TOMA, Daniel (SARTI-UPC); Dr CARANDELL, Matías (SARTI-UPC); Prof. DEL RÍO, Joaquín (SARTI-UPC); AGUZZI, Jacopo (ICM-CSIC)

Presenter: Dr MARTÍNEZ, Enoc (SARTI-UPC)

Session Classification: Session B1

Track Classification: Connecting research data: 6. Interoperable semantics at domain and application level