Martech 2024.
Marine Technology Workshop
06, 07th June. Palma, Spain

Autonomous Ocean Observations: The challenges for FAIR and reliable data for Digital Twins Development

Nikolaos D. Zarokanellos[1], Lara Diaz[1], Miguel Charcos Llorens[1], Emma Reyes[1], Melanie Juza[1], Albert Miralles[1], Manu Rubio[1], Patricia Rivera Rodríguez[1], Benjamín Casas[1], and Joaquín Tintoré[1,2]

[1] Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB), Palma, 07122, Spain,
[2]IMEDEA (CSIC-UIB), Esporles,07190, Spain

Abstract
Ocean observations and their effective integration in the Digital Twins are critical for understanding and managing marine ecosystems. The Balearic Islands Coastal Observing and Forecasting System (SOCIB) is a marine research infrastructure that monitors the ocean state and the variability of the Western Mediterranean Sea using multiple observing platforms, including autonomous underwater vehicles such as the Ocean Gliders. They represent a transformative way to gather observations efficiently and cost-effectively regarding the marine environment, also contributing to the global target of advancing towards net zero in carbon in 2040 for a global observing network. They have become an integral asset for oceanographic research by offering their unique ability to collect high-resolution data over extended periods, covering vast oceanic regions that may be challenging for traditional sampling methods. They can provide real-time (RT) or near-real-time (NRT) observations in all weather conditions via satellite communication networks, allowing researchers to facilitate adaptive sampling strategies. Autonomous platforms can monitor physical (temperature, conductivity, and depth) and biochemical (oxygen, chlorophyll fluorescence, CDOM, and bbp700nm) variables, making them suitable for various research applications. SOCIB ensures the data quality of Ocean Gliders through best practices, continuous monitoring of the sensor's calibration, automated validation of the observations, and comprehensive documentation. In addition, SOCIB has developed data management plans for the gliders that integrate a systematic approach to data collection, curation, preservation, and the description of the data and workflows, encompassing precise metadata documentation to ensure the traceability and reproducibility of oceanographic observations. These SOCIB initiatives aim to improve data quality and reliability, ultimately contributing to the development of Ocean Best Practices and fostering data interoperability. Furthermore, the SOCIB data repository adheres to CoreTrustSeal standards for digital repositories, thus contributing to creating TRUST (Transparency, Responsibility, User focus, Sustainability, and Technology) in the research data infrastructure to ensure SOCIB data are accessible in the future. These measures, encompassing thorough data validation and verification processes, contribute to maintaining the reliability, accuracy, and high standards for storing oceanographic observations. Upcoming integrations of Science Knowledge Graphs, FAIR evaluation tools, and machine-actionable DMP (maDMP) interoperability in the OSTrails project context will further enhance SOCIB's capabilities for Digital Twins development.

Keywords *– Autonomous underwater vehicles, Ocean Gliders, ocean observations, EOVs, Western Mediterranean, Quality Control, ocean best practices, data interoperability, FAIR principles, OSTrails, CoreTrustSeal, Digital Twins.*

I. The significance of ocean gliders data quality in enhancing digital twin capabilities

Autonomous platform observations help us better understand the physical and biogeochemical characteristics of the ocean since they accurately monitor the ocean's temporal and spatial variability. In addition, they help us to detect patterns and trends in measured Essential Ocean Variables (EOVs) and contribute to international efforts to observe ocean health and climate. Furthermore, the assimilation of autonomous observations into ocean models improves marine model accuracy and model-data fit and gives insights into storms and marine heat waves. The accuracy, dependability, and predictiveness of the Digital Twins depend on high-resolution data since they provide a more complete and realistic system representation of the mesoscale and submesoscale processes of the real ocean twin.

Over the last decade, particularly in response to the growing urgency to comprehend climate change trends, there has been a substantial rise in the demand for precise and reliable data. Even though manufacturers have improved laboratory calibrations and instrument stability, the ability to ensure in-field delayed mode (DM) correction to a world-class standard limits gliders' use as instrument platforms to record legacy data and/or as part of multiplatform observing systems [1]. To generate sufficiently high scientific quality data, the autonomous platform observations have to be cross-validated with data from CTD stations and water samples collected during seasonal cruises. Furthermore, SOCIB has also implemented best practices in the lab that ensure and evaluate the conditions of the oxygen glider sensors.

II. The Importance of FAIRness, Trustworthy digital repositories, and QC for the Digital Twins.

In order to effectively distribute data and information to the scientific community and society, it is important to standardize and harmonize our observations according to FAIR principles. SOCIB is committed to improving the fairness of digital objects across the data cycle by aligning its data management methods [2] with internationally recognized best practices [3]. However, the development of the SOCIB Digital Twins presents a unique challenge: it necessitates an additional effort to significantly enhance the quality and usability of data collected from autonomous platforms.

SOCIB maintains its digital twin's effectiveness and supports transparency, accountability, and glider observations' long-term sustainability in oceanographic research by following FAIR principles. Projects like CoreTrustSeal contribute to the FAIRness of data within the Digital Twin, promoting wider adoption and usability. The CoreTrustSeal principles guide SOCIB in maintaining high-quality data stewardship practices, including metadata standards, versioning, and long-term preservation, fostering reliability in Digital Twin datasets. In addition, by improving the interoperability, we can facilitate the seamless integration of diverse oceanographic datasets into the Digital Twin. The latter enhances the Digital Twin's capability to incorporate various observational data sources.

In the OSTrails project, SOCIB will integrate the measure of FAIRness along the entire data cycle by connecting the catalog of the digital objects involved in the data flow to data management and FAIRness measurement processes. This integration will guarantee consistency and compatibility among various components of the Digital Twin.

Processing and analyzing data and metadata are needed to provide high accuracy, and the QC data sets [4] are ingested in Digital Twin models. Thus, reliable data repositories that fulfill international standards are essential. SOCIB employs quality control tests to validate the sensor performance. These tests ensure that observational data integrated into the Digital Twin is accurate and reliable, enhancing the quality and trustworthiness of the Digital Twin outputs.

III. CONCLUSIONS

Finally, SOCIB improved ocean data quality of autonomous observations aiming to better understand physical and biogeochemical processes in the marine ecosystem. Leveraging best practices, improving interoperability, having trustworthy digital data repositories that meet international standards, and integrating autonomous observations into European portals can improve marine data integration and help develop ocean digital twins. Harmonizing all observations and metadata before feeding them into digital twin models will maximize the Digital Twin's potential for data-driven decision-making and resource management, promoting sustainable practices to protect the marine environment.

REFERENCES

[1] Allen JT, Munoz C, Gardiner J, Reeve KA, Alou-Font E and Zarokanellos N (2020) Near-Automatic Routine Field Calibration/Correction of Glider Salinity Data Using Whitespace Maximization Image Analysis of Theta/S Data. Front. Mar. Sci. 7:398. DOI: https://doi.org/10.3389/fmars.2020.00398

[2] Marasco, M., Zarokanellos, N.D., Charcos, M., Fernández, J.G., Miralles, A., Rotllán, P., and Tintoré, J. (2021) SOCIB Glider - Canales Endurance Line Data Management Plan (Version 1.0). Balearic Islands Coastal Observing and Forecasting System, SOCIB, 13pp. DOI: https://doi.org/10.25704/vj13-eh49

[3] Voanov, Y., Corgnati, L., Zarokanellos, N., Charcos, M., Anastasopoulou, G. and Perivoliotis, G. (2022) JERICO-S3 D6.3 - Data Management Best practices report for physical and BGC platforms. Version 2.0. Brest, France. IFREMER for JERICO S3, 68pp. (ERICO-S3-WP6-D6.3-300522-V1.0). DOI: https://doi.org/10.25607/OBP-1954

[4] Mantovani, C., Pearlman, J. and Simpson, P. (eds) (2023) JERICO-S3 Deliverable 5.2. Electronic Handbook for Mature Platforms: Mooring - HF Radar - FerryBox – Glider. Version 1.1. IFREMER for JERICO-S3, 195pp. (JERICO-S3-WP5-D5.2.-310123-V1.1). DOI: https://doi.org/10.25607/OBP-1945