ODANext
Oceanographic data acquisition: the next age

Project B2/202/P2/ODANext

SUMMARY

Introduction and Background

The ODANext project, titled "Oceanographic Data Acquisition: The Next Age," was initiated to address the evolving needs in oceanographic research with the commissioning of the new Research Vessel (RV) Belgica. The RV Belgica represents a significant upgrade in Belgium's maritime research capabilities, equipped with advanced sensors and scientific equipment designed to collect vast amounts of data from diverse marine environments.

However, the legacy systems used by its predecessor were inadequate for handling the volume and complexity of data generated by the new vessel. Thus, the project focused on redesigning the data acquisition, management, and processing systems to ensure that the vessel's capabilities could be fully realized, thereby supporting cutting-edge research and compliance with European directives like INSPIRE and Open Data.

Objectives and Methodology

The project's primary objective was to create an automated, efficient data workflow that could handle the data from the RV Belgica, ensuring the data is standardized, accessible, disseminated and of high quality. This included:

1. Establishing a continuous sensor-to-client data flow.

2. Enriching metadata where necessary.

3. Optimizing data storage and ensuring secure backups.

4. Integrating the data into relevant open science repositories.

The methodology adopted was dynamic and flexible, utilizing the MoSCoW prioritization technique and Agile principles to adapt to the challenges posed by the new vessel's technology. The project emphasized adherence to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) to maximize the utility and longevity of the data collected.

Key Developments and Results

1. On-board Data Acquisition and Management

  - The vessel's data acquisition systems were updated, replacing outdated hardware with modern alternatives that could handle the large volume of data generated by the vessel's sensors. We also worked on the optimization of the data acquisition frequencies to store only relevant data.

  - A visual interface has been deployed on-board using the open-source Grafana software. It offers a simple and reliable way to check sensors data in real-time allowing for early sensors malfunction detection.

2. Data Transfer and Storage

  - The project implemented efficient vessel-to-shore data transfer systems using V-SAT technology, which was found to be reliable and cost-effective.

  - A new database architecture was developed using PostgreSQL, optimized with TimescaleDB for managing time-series data, ensuring robust performance as the database grew. The database structure follows an internationally recognized standard developed for sensors data, namely the Open Geospatial Consortium SensorThings standard.

- Data from the previous RV Belgica have been formatted and transferred into the new database to ensure seamless access to both RV Belgica data.

3. Optimization and Quality Control

  - Various optimizations were made across the data acquisition and processing pipeline to ensure that the system could handle the increasing data volume without backlogs. We worked on the acquisition frequencies to limit the growing rate of the database, the hardware on-board and on-shore, the database structure with ad-hoc indexes for fast data recovery, etc.

  - An Automated quality control procedure has been developed in Python and published online on GitHub, allowing for the identification and correction of data anomalies. The procedure performs multiple checks including threshold limits, gradient and spikes flagging, and geo-locations validation. It is a first check that allows for early identification of bad quality data and should be complemented by visual validation. This second validation is not always performed due to the amount of data generated by the research vessel.

4. Dissemination and Valorization

  - The project successfully connected the new data system into existing platforms like the RV Belgica website, GOSUD, SeaDataNet, and INSPIRE, ensuring that data from both the old and new vessels were

accessible to researchers and the public. With this achievement we demonstrated the importance to expose the data online using standardize API (Application programming interface). Clients applications could easily connect to the API end-point and get data as they need, without the need for human interaction.

  - Efforts were made to standardize the data using linked data concepts and controlled vocabularies, improving data interoperability for machine-to-machine data exchange and compliance with international standards.

Impact and Future Work

The ODANext project has significantly improved the data management capabilities associated with the RV Belgica, positioning it as a leading platform for marine research. The systems and workflows developed during the project will facilitate more efficient and widespread dissemination of scientific data, supporting research efforts and policy-making. The project also highlighted the importance of ongoing optimization and the potential for future upgrades to further enhance the vessel's data management systems.

Keywords

FAIR principles, Data acquisition, Data dissemination, Interoperability, Marine Data, API, Research vessel, Water quality