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

Towards a Modular Open Data Infrastructure for Heterogeneous Marine Scientific Data Management

Enoc Martínez[1], Ikram Bghiel[1], Daniel M. Toma[1], Matias Carandell[1], Joaquín del Río[1]

[1] *SARTI-MAR Research Group, Universitat Politècnica de Catalunya, Vilanova i la Geltrú, Spain*

***Abstract* – This work introduces the Modular Open Data Infrastructure, an information system that leverages existing open-source tools and services into a unified software infrastructure for marine data management. It includes a comprehensive integration layer that seamlessly manages and integrates all underlying services, providing a centralized service management and a novel approach to address the challenges of handling diverse scientific data in the realm of marine research.**

***Keywords* –** *FAIR principles, data management, metadata, OGC standards*

I. INTRODUCTION

In the marine sciences domain, the need for effective and robust data management solutions has become increasingly important, due to the proliferation of data coming from heterogeneous sources. In response to this challenge, our work focuses on developing a Modular Open Data Infrastructure that adheres to the FAIR principles: Findable, Accessible, Interoperable, and Reusable [1], [2]. These principles provide a framework for ensuring that data are not only well-organized and accessible but also interoperable across different platforms and reusable for diverse scientific endeavors.

The Modular Open Data Infrastructure (MODI), was initially born from the need of a centralized data management tool able to handle heterogeneous from the OBSEA underwater observatory [3]. However, as it evolved into a complete marine data information system, handling data from multiple observation platforms and oceanographic campaigns.

II. MODULAR OPEN DATA INFRASTRUCTURE

Data management is a well-known topic and there are multiple open-source solutions focused as specific tasks. However, there is not a one-size-fits-all solution for data management in the marine domain. The proposed Modular Open Data Infrastructure leverages existing open-source data services into a modular architecture that covers the whole data and metadata lifecycle. The services included in the MODI are:

1. **SensorThings API**: An Open Geospatial Consortium (OGC) standard providing an open and unified framework to interconnect IoT sensing devices, data, and applications over the Web. It is an open standard addressing the syntactic interoperability and semantic interoperability of the Internet of Things
2. **ERDDAP**: a free and open source service that allows the access subsets of scientific datasets in common file formats and make graphs and maps from different data sources.
3. **CKAN**: The Comprehensive Knowledge Archive Network (CKAN) is an open-source open data portal for the storage and distribution of open data. Initially inspired by the package management capabilities of Debian Linux, CKAN has developed into a powerful data catalogue system that is mainly used by public institutions seeking to share their data with the general public.
4. **Grafana**: Grafana is a multi-platform open source analytics and interactive visualization web application. It provides charts and graphs for data visualiation from multiple data sources.
5. **Zabbix**: an open-source monitoring and alarming software tool. It can monitor IT infrastructure such as networks, servers, virtual machines, and cloud services but also physical devices such as oceanographic sensors or observation platforms.
6. **MMAPI**: The Marine Metadata API is a metadata management service inspired by the SensorThings API, but expanding its management capabilities beyond simple sensors to include observation platforms, datasets, people and funding. Additionally, it includes an interoperability layer that seamlessly manages and configures the rest of the MODI components.

The MODI architecture ensures that heterogeneous data can be integrated in real-time by using generic interoperability middleware that relies on sensor abstractions [4], [5]. In addition to its modular architecture and interoperability features, the Open Data Infrastructure excels in the recording and management of comprehensive metadata. This extends beyond generic metadata to encompass critical information related to sensors, platforms, deployments, etc. This ensures that researchers have access to precise information about the instruments generating the data, enabling more accurate analysis and interpretation. Platform metadata, covering vessel or station details, enhances the contextual understanding of the collected data. In addition to operational metadata, this infrastructure also integrates funding source information, specifically tailored to accommodate European projects and national projects using CORDIS (Community Research and Development Information Service) and AEI (Spanish research agency) interfaces. By automatically associating datasets with their respective funding references, researchers benefit from an efficient and streamlined approach to manage and attribute their work to specific funding sources. This ensures that generated datasets carry accurate and up-to-date funding information, facilitating compliance with project requirements and contributing to transparent reporting.

III. CONCLUSIONS

In conclusion, our Modular Open Data Infrastructure goes beyond conventional data management systems by incorporating robust metadata capabilities. The recording of sensor, platform, and data life cycle metadata, coupled with seamless integration of funding sources, establishes a foundation for transparent and accountable marine research. This infrastructure not only facilitates scientific exploration but also ensures that datasets are enriched with context, supporting reproducibility and advancing collaborative efforts in the marine research community.

ACKNOWLEDGEMENTS

This work has been financially supported by the European Commission's HORIZON-INFRA-2021-SERV-01 under the Geo-INQUIRE project (grant agreement 101058518) This work used the EGI infrastructure with the dedicated support of EGI-IFCA-STACK. Enoc Martínez and Matías Carandell acknowledge the financial support from the *Ministerio de Ciencia e Innovación* under the requalification of the Spanish university system program (Margarita Salas grant).

REFERENCES

[1] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, pp. 1–9, 2016, doi: 10.1038/sdata.2016.18.

[2] T. Tanhua *et al.*, "Ocean FAIR Data Services," *Front. Mar. Sci.*, vol. 6, p. 440, Aug. 2019, doi: 10.3389/fmars.2019.00440.

[3] J. Del-Rio *et al.*, "Obsea: A Decadal Balance for a Cabled Observatory Deployment," *IEEE Access*, vol. 8, pp. 33163–33177, 2020, doi: 10.1109/ACCESS.2020.2973771.

[4] E. Martínez, D. M. Toma, S. Jirka, and J. Del Río, "Middleware for plug and play integration of heterogeneous sensor resources into the sensor web," *Sensors*, vol. 17, no. 12, p. 2923, 2017, doi: 10.3390/s17122923.

[5] E. Martinez, A. Garcia-Benadi, D. M. Toma, E. Delory, S. Gomariz, and J. Del Rio, "Metadata-driven Universal Real-time Ocean Sound Measurement Architecture," *IEEE Access*, pp. 1–1, Feb. 2021, doi: 10.1109/access.2021.3058744.