

OGC Standard Compliant Platform Following FAIR Principles for Marine and Freshwater Research

Hanna Lahtinen ^{a,*}, Milla Kelhu ^{a,*}, Jaakko Kähkönen ^a, Markus Konkol ^b, Lassi Lehto ^a, Panu Muhli ^a, Henning Sten Hansen ^c, Juha Oksanen ^a

^a Finnish Geospatial Research Institute, hanna.e.lahtinen@nls.fi, milla.kelhu@nls.fi, jaakko.kahkonen@nls.fi, lassi.lehto@nls.fi, panu.muhli@nls.fi, juha.oksanen@nls.fi,

^b 52° North Spatial Information Research GmbH, m.konkol@52north.org

^c Aalborg University, hsh@plan.aau.dk

* Corresponding author

Keywords: European Open Science Cloud (EOSC), research infrastructure, data space, data lake, UN Agenda 2030

Abstract:

Sustainable use of our waters is one of the primary goals of the current EU policy as well as a significant part of the United Nations' Sustainable Development Goals (SDGs) (Tsani et al., 2020). These goals aim to maintain our vibrant and complex environment as viable as possible in the turbulent state of climate and environmental change that we face both in global and local scale. Therefore, the healthiness of marine and inland waters is addressed by multiple projects and missions, such as EU Mission "Restore our Oceans and Waters" as well as SDG 6 of clean water and sanitation which aims to "Ensure availability and sustainable management of water and sanitation for all" and SDG 14 of life below water which targets sustainable water policies (Luchsinger, 2024). One example for tackling these challenges in the context of research infrastructure services is the EU-funded European Open Science Cloud (EOSC) project AquaINFRA, which will create a system that will provide harmonised access to vital spatiotemporal data for freshwater and marine scientists as well as for the use of stakeholders and policymakers (Lehto et al., 2024). In the AquaINFRA context, achieving Technology Readiness Level (TLR) of 6 guarantees the performance of the platform in a proof-of-concept (POC) level.

AquaINFRA data and services follow the EOSC interoperability framework for example by applying FAIR (Findability, Accessibility, Interoperability, and Reusability) principles to its processes. Especially the principles "I" and "R", are highlighted in the AquaINFRA platform which offers open access to extensive library of relevant geospatial, socio-economic statistics and in-situ sensor data which are interoperable with various applications via multiple offered data formats such as COG (Cloud Optimized GeoTIFF) and JSON. In addition to these, end users have access to the demonstrative data of four regional use cases across Europe. AquaINFRA has taken multiple necessary steps towards more interoperable and reusable datasets in the aquatic realm by producing research infrastructure with an interactive data visualization service, the AquaINFRA Interaction Platform (AIP). The AIP is a federated metadata search platform including data sources such as European Environment Agency (EEA) and Copernicus Marine Service (CMEMS). The extent of the available aquatic research data is possible due to the advanced implementation of OGC APIs and custom data provider plugins, with access to 15 data providers. In the beta version of the AIP user can make a query for water related term and find research data from different providers as a result. For example, a query for "secchi" gives the user 39 results from all data providers.

AquaINFRA as an EOSC project will be compliant with other EOSC services as well as other digital twins of oceans (DTOs) such as EDITO (the European Digital Twin of the Ocean) and the Iliad project. Interoperable data advances the development of multiple (aquatic) digital twins. The AquaINFRA Interaction Platform (AIP) is the main gateway for end users to search and access aquatic data through the user interface and analysis tools it provides. The AIP utilizes the AquaINFRA Data Space developed in the project, that consists of Data Discovery and Access Service (DDAS), metadata service CKAN and the AquaINFRA Data Lake with multiple local analysis- and AI-ready data sources of varying harmonization levels (fig 1). The physical location of data is mostly in the in-cloud data repository, the AquaINFRA Data Lake, where data is uploaded with their own data harmonization processes to local repositories such as CSC Allas Object Storage environment. Interoperability is also considered for example with unified file naming conventions in data uploading processes.

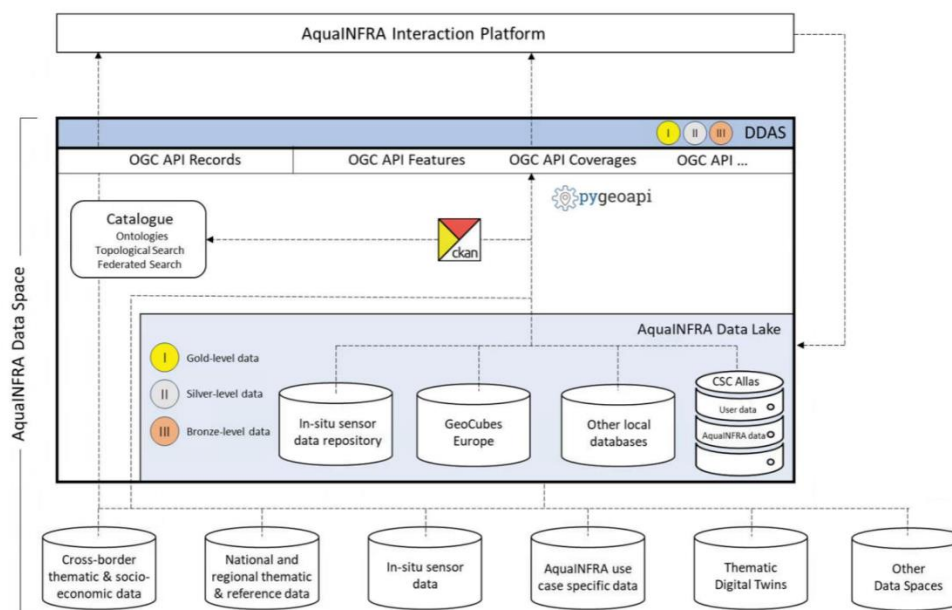


Figure 1. Architecture of the AIP and the AquaINFRA Data Space.

To follow the latest developments in geospatial standards, DDAS provides access to metadata and data via OGC API family access interfaces (Open Geospatial Consortium, 2025). For example, the data discovery service provides metadata via OGC API Records, and the vector data service utilizes OGC API Features. All these interfaces are provided in DDAS through pygeoapi. The Data Discovery Service of DDAS allows for finding, searching and filtering datasets via the metadata catalogue on DDAS platform as on-the-fly searches to external services –with the only exception being the local data stored in the AquaINFRA Data Lake and related metadata served in a CKAN instance. However, we also encountered some challenges in developing the research data infrastructure, as not all data providers offer machine-readable access to their holdings. For example, some providers require users to manually accept a disclaimer for data use before receiving the download link. Links to relevant resources may also not be clearly marked in the metadata and as such, not intuitively found, or they may be missing completely. A supposed download link may also be non-functional or lead to a more general data discovery portal. Some metadata services were also found lacking in search functionalities, such as geospatial search. Such mechanisms make seamless technical integration difficult and limit the reusability of the data. The Data Discovery Service generally supports geospatial search with bounding box, temporal search or type- and text-based search targeting the title, description and keywords of a metadata record. There is also an additional service, the AquaINFRA Ontology Search, for enriching the search functionalities. The Ontology Search has been implemented as a Python-based web service that provides a list of related term suggestions to the metadata search query term by using GEMET (General Multilingual Environmental Thesaurus) data to retrieve the related terms.

Acknowledgements

This research is carried out as part of the AquaINFRA (aquainfra.eu) project funded by the European Union through the Horizon Europe programme, grant agreement number 101094434, and coordinated by the Danish Aalborg University. We made use of the data and geocomputing services provided by the Open Geospatial Information Infrastructure for Research (Geoportti, urn:nbn:fi:research-infras-2016072513) funded by the Research Council of Finland, CSC – IT Center for Science, and other Geoportti consortium members. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

References

- Lehto, L., Kähkönen, J., Muhli, P., and Oksanen, J., 2024. Data Access Platform for Marine and Freshwater Applications. In: *Abstracts of the International Cartographic Association*, September 9–11, 2024, Vienna, Austria, Vol. 7.
- Luchsinger, G., 2024. The Sustainable Development Goals report 2024. Retrieved from <https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf>
- Open Geospatial Consortium, (2025). OGC API Standards. Retrieved from <https://ogcapi.ogc.org/#standards>
- Tsani, S., Koundouri, P., and Akinsete, E., 2020. Resource management and sustainable development: A review of the European water policies in accordance with the United Nations' Sustainable Development Goals. In: *Environmental Science and Policy*, Vol.114, pp. 570–579.