



The AquaINFRA research data infrastructure: Knowledge generation through FAIR open data and reproducible computational workflows

Markus Konkol¹, **Simon Jirka**¹, Henning Sten Hansen², Kaori Otsu³, Sami Domisch⁴, Merret Buurman⁴, Vanessa Bremerich⁴, Astra Labuce⁵, Pekka Latvala⁶, Juha Oksanen⁶, and Björn Grüning⁷

¹52°North Spatial Information Research GmbH, Germany

²Department of Sustainability & Planning, Aalborg University Copenhagen

³Centro de Investigación Ecológica y Aplicaciones Forestales, Universitat Autònoma de Barcelona

⁴Leibniz-Institute of Freshwater Ecology and Inland Fisheries

⁵Latvian Institute of Aquatic Ecology, Agency of Daugavpils University

⁶Finnish Geospatial Research Institute FGI, National Land Survey of Finland

⁷Department of Computer Science, Albert-Ludwigs-University Freiburg

The primary aim of the EU-funded AquaINFRA [1] initiative is the development of a research data infrastructure to help marine and freshwater scientists generate new knowledge for restoring healthy oceans, rivers, and lakes. Several use cases representing pan-Europe, the Baltic Sea, and the North Sea define the scope for implementing the infrastructure and demonstrating its potential. An essential goal of the project is to address high-quality, FAIR, and open multi-disciplinary data. In addition, the infrastructure focuses on making python- and R-based data analyses accessible as reusable tools, services, and workflows. Finally, a specific objective is to develop a platform that is compliant with the European Open Science Cloud (EOSC) Interoperability Framework as an overarching research infrastructure.

In this contribution, we provide an overview of the recent developments in AquaINFRA and show its realization in a use case. The key components of the AquaINFRA research data infrastructure are the Data Discovery and Access Service (DDAS), the AquaINFRA Interaction Platform (AIP), and the Virtual Research Environment (VRE).

The DDAS [2] is the backend of the infrastructure and based on a federated metadata search mechanism sending requests to selected remote metadata providers on the fly. Access to harmonized metadata is accomplished by leveraging community standards, such as the OGC API family including OGC API Records, OGC API Features, and OGC API Coverages. Another component of the DDAS is the Ontology Search that suggests alternative search terms for the keyword entered by the user.

The AIP [3] is the central gateway to the project and provides a search interface to find, access, and reuse aquatic digital resources. It is built on top of the DDAS and allows users to search for data, services, and software. Several features are provided to refine the search query, for instance, using a bounding box and selecting specific data providers.

The purpose of the VRE is to have a set of web-based applications facilitating the reuse of existing tools as well as the contribution of newly developed tools, and connecting them to create readily shareable and reproducible workflows. Hence, the VRE is composed of three modules:

- MyBinder is provided as a virtual lab to let users engage with an existing analysis written in R or Python in a pre-defined computational environment.
- OGC API Processes are provided as a web API service allowing researchers to make remote requests and integrate these into their analysis.
- Access to the Galaxy platform is provided to create reproducible computational workflows. To achieve that, the analysis scripts developed in the case studies are integrated into the Galaxy platform by wrapping the OGC API Processes as Galaxy tools.

Besides the research data infrastructure demonstrating the usefulness of FAIR open data and reproducible computational workflows, the project outcomes are expected to foster collaboration across borders, data infrastructures, and disciplines.

This project has received funding from the European Commission's Horizon Europe Research and Innovation programme. Grant agreement No 101094434.

1) Project website: <https://aquainfra.eu/>

2) DDAS: <https://vm4072.kaj.pouta.csc.fi/ddas>

3) AIP: <https://aquainfra.dev.52north.org/>