

Data Access Platform for Marine and Freshwater Applications

Lassi Lehto*, Jaakko Kähkönen, Panu Muhli, Juha Oksanen

Finnish Geospatial Research Institute, lassi.lehto@nls.fi, jaakko.kahkonen@nls.fi, panu.muhli@nls.fi, juha.oksanen@nls.fi

* Corresponding author

Keywords: data portal, OGC API, federated search, freshwater

Abstract:

Marine and freshwater conservation processes require easy and wide-reaching access to data. There are several data portals available providing marine datasets for research purposes. These include services like Copernicus Marine, EMODnet and ODIMS. However, there is lack in provision of harmonised access to freshwater related datasets in Europe. The Horizon-funded EOSC-project AquaINFRA (aquainfra.eu) aims at tackling this issue. The four-year project involves 19 European organisations and is coordinated by the Aalborg University, Denmark.

During its first year of operation, the AquaINFRA project has developed a first version of the project's data access portal, Data Discovery and Access Service (DDAS). As the name implies, the portal is supposed to provide support both for data discovery using metadata search functionalities, and for the actual data access mechanisms. These services are being consumed by the AquaINFRA Interaction Platform (AIP), a separate processing layer that provides user interface and analysis support for end users, i.e. researchers working in the domain area. The AquaINFRA project focuses specially on providing access to freshwater-related data sources.

The Data Discovery service on the DDAS platform is based on a federated metadata search approach. Instead of harvesting metadata from remote services into a local metadata storage, the DDAS service carries out on-the-fly metadata queries on a set of selected remote metadata services. At the moment the DDAS Discovery Service is connected to APIs of a dozen remote metadata services, most of them supporting the OGC's Catalogue Service Web (CSW) access interface. On the DDAS platform, the access to metadata and data content is provided via the modern OGC API family of service interface standards: OGC API Records, OGC API Features and OGC API Coverages. All of these services are implemented using the geospatial Python server `pygeoapi` (`pygeoapi.io`)

The federated metadata search is accomplished making use of set of service-specific `pygeoapi` plugin modules that take care of the particularities of each remote metadata service, and carry out the required metadata harmonisation into the metadata model specified in the OGC API Records standard. The background metadata requests to the remote services are performed in parallel for performance reasons. The resulting metadata records are combined and sorted on the DDAS platform and then returned via OGC API Records interface to the calling application. The OGC API Records query currently supports search terms that are matched against the 'title' and 'abstract' fields in the CSW metadata, keywords that are matched to 'subject' field, and spatial extent (BBOX) that is compared with the 'BoundingBox' metadata element. The user can either explicitly select the remote metadata services to be included into the federated search, or use the metadata collection 'AquaINFRA', which will automatically include all available services into the search. Additional support functions being developed for the AquaINFRA Discovery Service include support for ontology-based expansion of the search terms and support for topological constraints in the hydrography network.

In the Data Access service of DDAS, a few important datasets have been uploaded to the platform and are served via the OGC API Features (vector) and OGC API Coverages (raster) services. The vector datasets include for instance hydrography-related layers of the EuroRegionalMap (EuroGeographics), the EU-Hydro dataset (EEA), the WMO Basins and Sub-Basins dataset (Global Runoff Data Centre) and Drainage Basins of the Hydrography90m dataset. The raster datasets currently available include CORINE Land Cover Plus Backbone (Copernicus) and EU-DEM (EEA).

First tests have been carried out on using the same on-the-fly approach, applied in the metadata provision, also for providing the actual data content. A `pygeoapi` data provider plugin module has been developed for accessing a remote data retrieval API and dynamically provide the accessed data via the OGC API Features service of the DDAS platform. This kind of data access approach has been applied for two different data sources. Firstly, a data access plugin has been developed for the OData-compliant interface of the Finnish Environment Institute providing daily discharge and water level observations from all around Finland. Secondly, a plugin has been developed to retrieve State of Environment (SOE) observations data from the DISCODATA service of the EEA. In this plugin the incoming OGC API Features request is

translated to conform to the DISCODATA's SQL API. In both plugins the service-specific JSON-formatted output is transformed to the geospatial GeoJSON format.

In the final state of the AquaINFRA service, a researcher working in the marine and freshwater realm is envisaged to find useful datasets by accessing the DDAS Discovery Service from within the user interface provided by the AIP, immediately access the data source, either directly or via the DDAS Data Access service, by following the download link present in the metadata, and subsequently take the resulting dataset into the AquaINFRA Virtual Research Environment (VRE) for analysis.

Acknowledgements

This research is carried out as part of the AquaINFRA project funded by the European Union through the Horizon Europe programme, grant agreement number 101094434.