

Facilitating the Publication of Real-Time Marine Observation Data: The SeaDataCloud SWE Ingestion Service

Simon Jirka, 52°North GmbH (Germany), jirka@52north.org
Christian Autermann, 52°North GmbH (Germany), c.autermann@52north.org

Near real-time in-situ sensor measurements are of high importance for answering a broad range of scientific questions in the marine domain. At the same time, collected measurements are not only relevant for one specific question but they may be of high value in other research contexts, as well. Thus, there is a significant movement towards the development of research data infrastructures which shall facilitate the sharing of such scientific data sets.

Besides the availability of research data infrastructures, it is also important to consider interoperability aspects. This comprises especially the application of common standards for data access interfaces, data models, and encodings to ensure that data sets from different sources can be easily integrated. An important framework of standards addressing exactly this aspect is the Sensor Web Enablement (SWE) family of standards of the Open Geospatial Consortium (OGC).

In order to increase the availability of near-real time in-situ observation data for researchers, the SeaDataCloud project aims, among other aspects, at facilitating the data publication process. For this purpose, the SeaDataCloud SWE Ingestion Service is currently being developed as a tool to support the publication of near real-time in-situ observations in interoperable research data infrastructures. This component is intended to support sensor operators, researchers and data owners during the publication of collected marine observation data by offering a configurable and re-usable publication workflow.

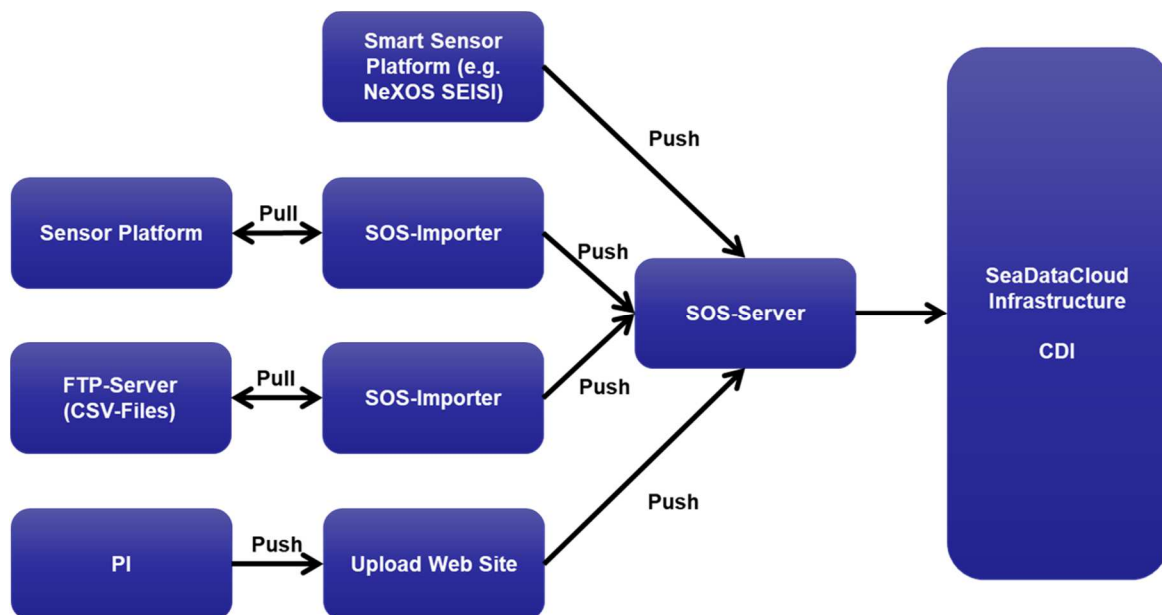


Figure 1: Overview of the Proposed Data Ingestion Workflow

Figure 1 illustrates the proposed ingestion workflow: Central component is an OGC Sensor Observation Service (SOS) instance which is capable of storing observation data and metadata from different sources. Through dedicated adapters and importer components different sources can be linked to the central SOS server:

- Smart sensors (for example sensors supporting the Smart Electronic Interface for Sensor Interoperability (SEISI) developed by the NeXOS project): Such sensors are able to push their metadata and data autonomously to an SOS server.
- Sensor platforms with a well-described data access interface: Such sensor platforms can be regularly queried by an importer for new available data. In this case a formal description of the sensor platform interface is necessary (based on the OGC Sensor Model Language (SensorML) standard) so that an importer can be automatically configured to the commands and data formats of the sensor platform.
- Data archives such as files or servers containing CSV files: Automatic harvesting of data by an importer. In this case a description of the data structure is necessary so that the importer can be configured to the specific format in which the data is provided.
- Uploading (e.g. by researchers) so that the data is automatically processed. Also in this case a description of the data structure is necessary so that the importer can be configured to the specific format in which the data is provided.
- Internet of Things data sources (e.g. MQTT brokers)

The harvested metadata and data can subsequently be coupled with the SeaDataCloud infrastructure. An important element of this data flow is the provision of the collected information through the SeaDataCloud Common Data Index (CDI) so that researchers are able to discover and access the available data sets. In this case, the SOS server will act as a feeder for the CDI through interfaces offered by the CDI. Furthermore, the publication of the data via SOS servers allows the visualisation of the previously ingested data in tools such as the Helgoland Sensor Web Viewer (Figure 2).

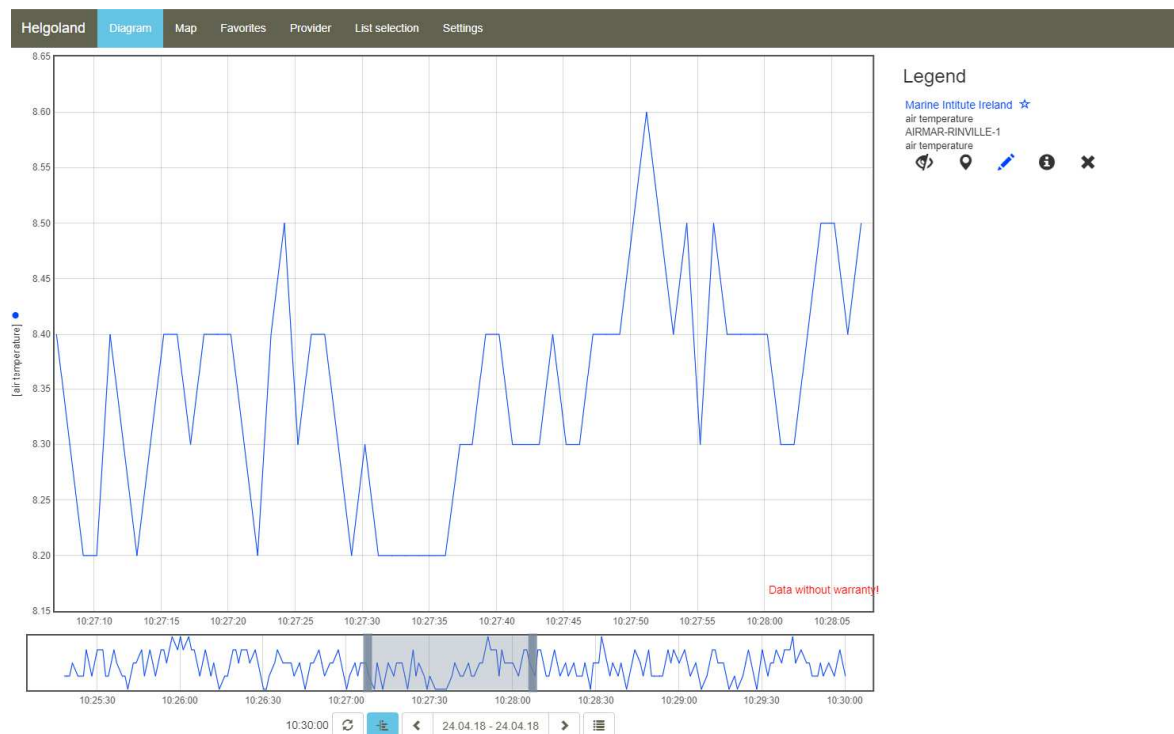


Figure 2: Published Observation Data Visualised in the Helgoland Sensor Web Viewer

In summary, the presented SeaDataCloud SWE Ingestion Service is a powerful tool to publish marine observation data from different sources on interoperable research data infrastructures. This is complemented by the SOS interface for facilitating data distribution and access, and integration into the SeaDataCloud CDI as well as a broad range of data analysis and viewing tools such as the Helgoland Sensor Web viewer.