O$_2$A – Data Flow Framework from Sensor Observations to Archives

Roland Koppe, Peter Gerchow, Ana Macario, Antonie Haas, Christian Schäfer-Neth, Hans Pfeiffenberger, and Angela Schäfer

Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany (roland.koppe@awi.de)

The Alfred Wegener Institute for Polar and Marine Research (AWI) performing research from the Arctic to the Antarctic, which requires vessels and aircrafts, land-based stations and ocean-based stations as well as reams of other platforms, devices and sensor networks. The increasing number of hundreds of platforms, devices and sensors along with heterogeneous project-driven requirements towards satellite communication, monitoring, quality assessment and quality control, validation, processing algorithms, visualization and dissemination lead us to streamline data flows and reducing heterogeneity. To enable coherent data discovery, visualization and dissemination as well as archiving and data publication we are developing and sustaining a data infrastructure framework comprising several integrated components called O$_2$A – from observations to archives. O$_2$A facilitates the seamless flow of sensor observations to archives based on state-of-the-art technology. It supports international standards for metadata formats and interfaces assuring interoperability in international context (e.g. SOS/SWE, WPS, WMS, WFS, etc.).

Today O$_2$A comprises of several operational components. Scientists using our sensor.awi.de solution to describe their sensor systems aligned with the SensorML standard. These sensor descriptions also cover data source information, which enables automatic data harvesting and ingesting in for example into our near real-time databases and archives. The ingest component also allows automatic quality assessment and quality control methods applied to data streams. Our dashboard.awi.de solution enables scientists to monitor and share these data streams with e.g. plots, maps and statistics out of the box and is accompanied with an SOS interface. The data is seamlessly feed into our workspaces solutions (currently under development) proving analytic tools beyond others based on Hadoop and Spark. Services around geo-information systems complete the workspace for map-based visualization and analytics (maps.awi.de). Finally, we operate institutional and international repositories like PANGAEA for archiving and publishing our data.

Currently we are developing user-friendly private/hybrid cloud solutions for collaborative data management and analytics for our earth and environmental science community within the Helmholtz Data Federation (HDF) infrastructure project. This effort covers replicated storage on petabyte-scale level as well as state-of-the-art computing solutions with e.g. Hadoop, Spark, Jupyter/Zeppelin notebooks supporting Python, R and others and rasdaman with access to shared data pools based on near real-time data, workspace data and archives.