

# Bridging the gap between data and sensor information

**Roland Koppe**, Alfred Wegener Institute (Germany), [Roland.Koppe@awi.de](mailto:Roland.Koppe@awi.de)  
**Ana Macario**, Alfred Wegener Institute (Germany), [Ana.Macario@awi.de](mailto:Ana.Macario@awi.de)

Over the last 2 decades, the Alfred Wegener Institute (AWI) Helmholtz Center for Polar and Marine Research has been continuously committed to develop and sustain an eResearch infrastructure for coherent discovery, view, dissemination, and archival of scientific data and related information in polar and marine regions. Most of the data collected by scientists originates from research activities carried out on a wide variety of research platforms operated by AWI (vessels, aircraft, stations, buoys, moorings, in-situ ocean floor stations, drones, and ocean floor crawling systems). Archival and publishing in the information system PANGAEA along with DOI assignment to individual datasets is a typical end-of-line for most data producers.

In order to address the increasing heterogeneity of research platforms and respective devices and sensors along with varying project-driven requirements, we built a generic and cost-effective virtual research infrastructure, hereafter named O2A, intended to support the flow of sensor observation to archives. O2A components are modular, extensible and in compliance with OGC standards and FAIR principles, ensuring interoperability at international level and data re-use.

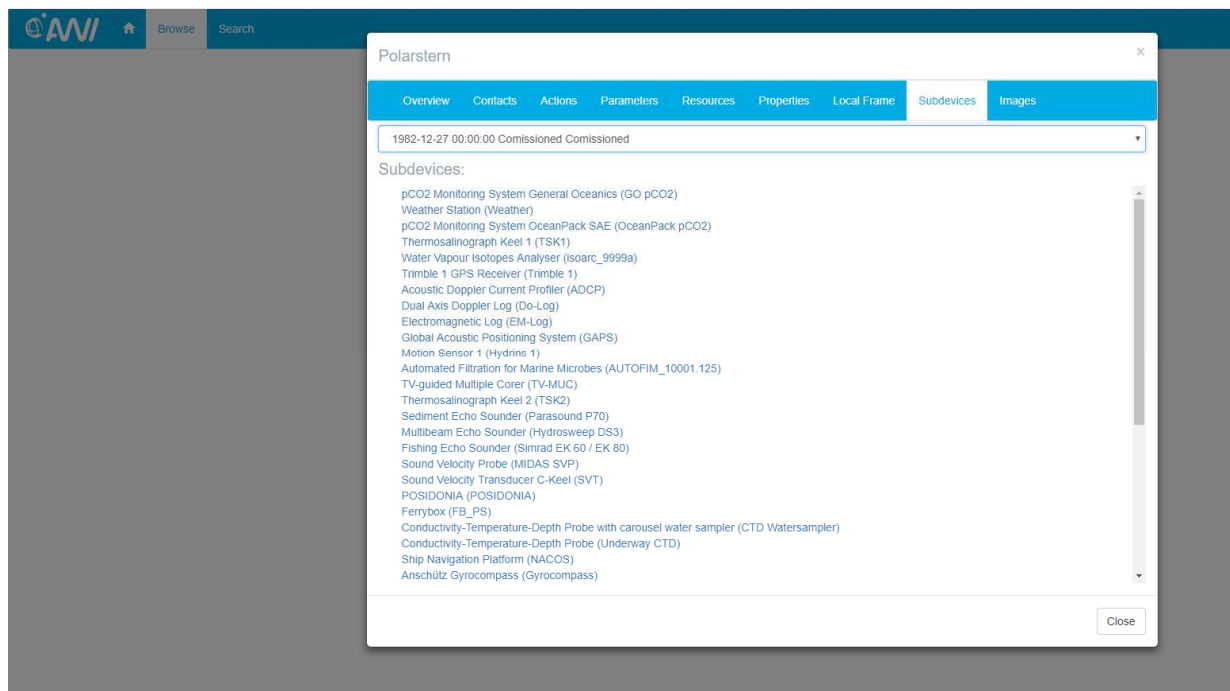


Figure 1: SENSOR <https://sensor.awi.de>, is one of the key components within O2A as a beginning-of-line for scientists. To date, over 1400 sensors are described. In this figure we show the shipborne instruments for RV Polarstern

The component SENSOR (<http://sensor.awi.de>), designed as repository for various types of information related to platforms, instruments and sensors, is intended to provide the following enhancements:

- Address specific requirements from scientists meant to support their scientific pipelines.  
Examples:
  - Integration of station lists from expeditions to action/event component (as part of the sensor provenance metadata)

- Harmonization of various vocabularies with existing data acquisition systems on board of our vessels and stations (device names, types, parameters, units)
- In-house inventory and management of platforms/devices/sensors
- Traceability of sensors in particular in regard to varying payloads for small vehicles (e.g. ROVs) and instrumentation of expeditions/landing parties and observatories

SENSOR has been publicly available since 2015 (incl. bi-annual release packages) and will be adopted as repository for the international Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAIC; start planned for Autumn 2019). Given that platforms and sensors evolve in time (sensors are being calibrated, mounted/unmounted, etc.), it is clearly important to trace changes applied to these over time. We have intensively discussed the usability scenarios with AWI scientists in particular in regard to which types of actions/events associated with a sensor are relevant in the long run from data use perspective. While tracking changes to platforms/instruments/sensors is technically feasible, it became clear to us early on that in practice one has to decide which type of action/event (e.g. sensor calibration) warrants the creation of an individual version of a sensor (= a new PID). Pragmatically, we have opted to (a) narrow down the scenarios in which metadata versioning takes place (including minting PIDs/handles), and, (b) “force” certain changes to sensors to trigger the creation of a new version.

Having successfully integrated AWI Handle Server services with SENSOR, we are currently able to support:

- Sensor versioning, i.e. audit trail of changes
- Minting of persistent identifiers, generating UUIDs in the handle syntax. Example: <https://hdl.handle.net/10013/sensor.bf472f3a-d236-47ee-972e-0e67c86085eb>
- Automated generation of a sensor citation

In this talk we will present our approach in detail and illustrate with concrete examples how sensor metadata enhances the quality of data archived in PANGAEA.