MOSAiC goes O2A - Arctic Expedition Data Flow from Observations to Archives

Antonia Immerz¹, Angela Schaefer¹, and the AWI Data Centre MOSAiC Team*
¹Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
*A full list of authors appears at the end of the abstract

During the largest polar expedition in history starting in September 2019, the German research icebreaker Polarstern spends a whole year drifting with the ice through the Arctic Ocean. The MOSAiC expedition takes the closest look ever at the Arctic even throughout the polar winter to gain fundamental insights and most unique on-site data for a better understanding of global climate change. Hundreds of researchers from 20 countries are involved. Scientists will use the in situ gathered data instantaneously in near-real time modus as well as long afterwards all around the globe taking climate research to a completely new level. Hence, proper data management, sampling strategies beforehand, and monitoring actual data flow as well as processing, analysis and sharing of data during and long after the MOSAiC expedition are the most essential tools for scientific gain and progress.

To prepare for that challenge we adapted and integrated the research data management framework O2A "Data flow from Observations to Archives" to the needs of the MOSAiC expedition on board Polarstern as well as on land for data storage and access at the Alfred Wegener Institute Computing and Data Center in Bremerhaven, Germany. Our O2A-framework assembles a modular research infrastructure comprising a collection of tools and services. These components allow researchers to register all necessary sensor metadata beforehand linked to automatized data ingestion and to ensure and monitor data flow as well as to process, analyze, and publish data to turn the most valuable and uniquely gained arctic data into scientific outcomes. The framework further allows for the integration of data obtained with discrete sampling devices into the data flow.

These requirements have led us to adapt the generic and cost-effective framework O2A to enable, control, and access the flow of sensor observations to archives in a cloud-like infrastructure on board Polarstern and later on to land based repositories for international availability.

Major roadblocks of the MOSAiC-O2A data flow framework are (i) the increasing number and complexity of research platforms, devices, and sensors, (ii) the heterogeneous interdisciplinary driven requirements towards, e. g., satellite data, sensor monitoring, in situ sample collection, quality assessment and control, processing, analysis and visualization, and (iii) the demand for near real time analyses on board as well as on land with limited satellite bandwidth.
The key modules of O2A’s digital research infrastructure established by AWI are implementing the FAIR principles:

- **SENSORWeb**, to register sensor applications and sampling devices and capture controlled meta data before and alongside any measurements in the field
- **Data ingest**, allowing researchers to feed data into storage systems and processing pipelines in a prepared and documented way, at best in controlled near real-time data streams
- **Dashboards** allowing researchers to find and access data and share and collaborate among partners
- **Workspace** enabling researchers to access and use data with research software utilizing a cloud-based virtualized infrastructure that allows researchers to analyze massive amounts of data on the spot
- **Archiving** and **publishing data** via repositories and Digital Object Identifiers (DOI)

AWI Data Centre MOSAiC Team:
A. Immerz (Antonia.Immerz@awi.de), M. Ajjan (Mohammad.Ajjan@awi.de), J. Bein (Jan.Bein@awi.de), B. Bräuer (Benny.Braeuer@awi.de), T. Dinter (Tilmann.Dinter@awi.de), A. Driemel (Amelie.Driemel@awi.de), T. Duede (Tobias.Duede@awi.de), J. Eilers (Janik.Eilers@awi.de), P. Gerchow (Peter.Gerchow@awi.de), F. O. Glöckner (Frank.Oliver.Gloeckner@awi.de), M. Günster (Michael.Guenster@awi.de), A. Haas (Antonie.Haas@awi.de), N. Harms (Nico.Harms@awi.de), S. Immoor (Sebastian.Immoor@awi.de), R. Koppe (Roland.Koppe@awi.de), H. Liegmahl-Pieper (Herbert.Liegmahl-Pieper@awi.de), A. Macario (Ana.Macario@awi.de), S. Makedanz (Siegfried.Makedanz@awi.de), J. Matthes (Joerg.Matthes@awi.de), M. Petri (Martin.Petri@awi.de), H. Pfeiffenberger (Hans.Pfeiffenberger@awi.de), S. Pinkernell (Stefan.Pinkernell@awi.de), D. Ransby (Daniela.Ransby@awi.de), S. Rehmcke (Steven.Rehmcke@awi.de), A. Schaefer (Angela.Schaefer@awi.de), C. Schäfer-Neth (Christian.Schaefer-Neth@awi.de), J. Schlüter (Jens-Michael.Schluter@awi.de), S. Schumacher (Stefanie.Schumacher@awi.de), R. Spettngel (Ralf.Spettngel@awi.de), A. Steinbach (Angelo.Steinbach@awi.de), A. Thiele (Andreas.Thiele@awi.de), F. Thiele-Wolff (Frauke.Thiele-Wolff@awi.de), M. Thoma (Malte.Thoma@awi.de), A. Walter (Andreas.Walter@awi.de), P. Weidinger (Philipp.Weidinger@awi.de), S. Frickenhaus (Stephan.Frickenhaus@awi.de)


**AWI Data Centre MOSAiC Team:** A. Immerz (Antonia.Immerz@awi.de), M. Ajjan (Mohammad.Ajjan@awi.de), J. Bein (Jan.Bein@awi.de), B. Bräuer (Benny.Braeuer@awi.de), T. Dinter (Tilmann.Dinter@awi.de), A. Driemel (Amelie.Driemel@awi.de), T. Duede (Tobias.Duede@awi.de), J. Eilers (Janik.Eilers@awi.de), P. Gerchow (Peter.Gerchow@awi.de), F. O. Glöckner (Frank.Oliver.Gloeckner@awi.de), M. Günster (Michael.Guenster@awi.de), A. Haas (Antonie.Haas@awi.de), N. Harms (Nico.Harms@awi.de), S. Immoor (Sebastian.Immoor@awi.de), R. Koppe (Roland.Koppe@awi.de), H. Liegmahl-Pieper (Herbert.Liegmahl-Pieper@awi.de), A. Macario (Ana.Macario@awi.de), S. Makedanz (Siegfried.Makedanz@awi.de), J. Matthes (Joerg.Matthes@awi.de), M. Petri (Martin.Petri@awi.de), H. Pfeiffenberger (Hans.Pfeiffenberger@awi.de), S. Pinkernell (Stefan.Pinkernell@awi.de), D. Ransby