



Implementing FAIR in a Collaborative Data Management Framework

Angela Schäfer, Norbert Anselm, Janik Eilers, Stephan Frickenhaus, Peter Gerchow, Frank Oliver Glöckner, Antonie Haas, Isabel Herrarte, Roland Koppe, Ana Macario, Christian Schäfer-Neth, Brenner Silva, and Philipp Fischer
Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Computing and Data Centre, Bremerhaven, Germany

Today's fast digital growth made data the most essential tool for scientific progress in Earth Systems Science. Hence, we strive to assemble a modular research infrastructure comprising a collection of tools and services that allow researchers to turn big data into scientific outcomes.

Major roadblocks are (i) the increasing number and complexity of research platforms, devices, and sensors, (ii) the heterogeneous project-driven requirements towards, e. g., satellite data, sensor monitoring, quality assessment and control, processing, analysis and visualization, and (iii) the demand for near real time analyses.

These requirements have led us to build a generic and cost-effective framework **O2A (Observation to Archive)** to enable, control, and access the flow of sensor observations to archives and repositories.

By establishing O2A within major cooperative projects like **MOSES** and **Digital Earth** in the research field Earth and Environment of the German Helmholtz Association, we extend research data management services, computing powers, and skills to connect with the evolving software and storage services for data science. This fully supports the typical scientific workflow from its very beginning to its very end, that is, from data acquisition to final data publication.

The key modules of O2A's digital research infrastructure established by AWI to enable Digital Earth Science are implementing the **FAIR** principles:

- **Sensor Web**, to register sensor applications and capture controlled meta data before and alongside any measurement 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 NRT 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 in a cloud-based virtualized infrastructure that allows researchers to analyse massive amounts of data on

the spot

- **Archiving** and **publishing data** via repositories and Digital Object Identifiers (DOI).