Geophysical Research Abstracts
Vol. 20, EGU2018-11016, 2018
EGU General Assembly 2018

Automated Data Ingestion for the Australian Ocean Data Network

Marton Hidas, Roger Proctor, Sebastien Mancini, Peter Blain, Leigh Gordon, Angus Scheibner, and Laurent Besnard

Integrated Marine Observing System, University of Tasmania, Hobart, Australia (marty.hidas@utas.edu.au)

The Australian Ocean Data Network (AODN) is an interoperable online network of marine and climate data resources. It is a collaboration between six major Australian Commonwealth Agencies, the Integrated Marine Observing System (IMOS), and a growing list of universities, state government offices and other organisations in Australia, New Zealand and the Pacific.

AODN data collections cover a large geographic area (from coast to open ocean, from equator to Antarctica), a wide range of observed parameters (physical, chemical, biological) and are obtained using a variety of platforms and technologies (e.g. ships, autonomous floats, gliders, moorings, satellites, animal tags, coastal radar). The end users include researchers, managers, policy makers, consultants, sailors and fishers. All data products contributed to the AODN are made freely and openly available to the public via the AODN Portal (https://portal.aodn.org.au/).

The development of the AODN Portal and its supporting infrastructure was guided by the diversity of data, the needs of its users, the goals of robustness, scalability, and interoperability with other programs and data sources. Wherever possible, we adopt and extend existing standards and open-source software. In particular, we create OGC-compliant Web Map Services and Web Feature Services using GeoServer, and describe data collections in ISO 19115 metadata records managed with GeoNetwork Opensource software. All software developed by the AODN is also open-source and accessible at https://github.com/aodn.

We focus on recent improvements in the data ingestion "pipeline" software that handle data submitted to the AODN for publication. At the core of this system is a new Python package that manages the common tasks of checking incoming files for compliance with conventions, moving files to public or archive storage, harvesting metadata and data into the database that forms the basis of our web services, and notifying data providers of the end result. Details of these tasks can easily be tailored for each data collection, and additional steps, such as data quality checks, the creation of derived data products or visualisations, can be added at various stages of the process. Robustness, transparency, and fail-safe operation were key design criteria. Every file received is either published successfully, or saved to an "error" directory, with details of the error logged and email notifications sent, so that the cause can be investigated and rectified. The benefits for end users are faster publication of new data, improved consistency within data collections, compliance (e.g. with the Climate and Forecasting conventions for NetCDF), and interoperability with other collections.