INFORMATION INFRASTRUCTURE FOR THE AUSTRALIAN INTEGRATED MARINE OBSERVING SYSTEM

Roger Proctor[(1),(2)], Philip Bohm[(1)], Stephen Cameron[(1)], Jacqui Hope[(1)], Craig Jones[(1)], Sebastien Mancini[(1)], Kylie Pepper[(1)], Kate Roberts[(1)], Katherine Tattersall[(1)], Brendon Ward[(1)], Geoff Williams[(1)] and Pauline Mak[(3)], Florian Goessmann[(4)].

[(1)]University of Tasmania, Hobart, TAS 7001, Australia, roger.proctor@utas.edu.au, philip.bohm@utas.edu.au, stephen.cameron@utas.edu.au, jacqui.hope@utas.edu.au, craig.jones@utas.edu.au, sebastien.mancini@utas.edu.au, kbpepper@utas.edu.au, kate.roberts@utas.edu.au, katherine.tattersall@utas.edu.au, bjward@utas.edu.au, geoff.williams@utas.edu.au
[(2)]Proudman Oceanographic Laboratory, Liverpool, L3 5DA, UK, rp@pol.ac.uk
[(3)]Australian Research Collaboration Service, Hobart, TAS 7001, Australia, pauline.mak@arcs.org.au
[(4)]Australian Research Collaboration Service, Perth, WA, Australia, florian.goessmann@arcs.org.au

ABSTRACT

The Australian Integrated Marine Observing System (IMOS, www.imos.org.au), is a distributed set of equipment and data-information services which collectively contribute to meeting the needs of marine climate research in Australia. The observing system provides data in the open oceans around Australia out to a few thousand kilometres as well as the coastal oceans. The IMOS Office coordinates the deployment of a wide range of equipment and assembles the data through 11 Facilities distributed around the country. The data are made available to researchers through the electronic Marine Information Infrastructure (eMII) located at the University of Tasmania. This paper describes the information infrastructure.

1. BACKGROUND

IMOS was established as part of the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) with $50M AUD and more than equal co-investments from Universities and government agencies. Additional federal funds were secured in 2009, as part of the Education Investment Fund (EIF) Super Science Initiative, to enhance the existing observing system out to June 2013, and extend into Northern Australian and Southern Ocean waters.

IMOS is a nationally managed and distributed set of equipment providing streams of in situ oceanographic data and information services that collectively will contribute to meeting the needs of marine research in both open oceans and coastal oceans around Australia. Combined with satellite data, it provides essential in situ data to understand and model the role of the oceans in climate change, and data to initialize seasonal climate prediction models. If sustained in the long term, it will permit identification and management of climate change in the coastal marine environment. As an NCRIS project IMOS was designed to support research, but the data streams are also useful for many societal, environmental and economic applications, such as management of marine natural resources and their associated ecosystems, support and management of coastal and offshore industries, safety at sea, marine tourism and defense.

2. STRUCTURE OF IMOS

IMOS delivers data streams through ten national facilities (Argo floats, Ships of opportunity, Southern Ocean mooring timeseries, National ocean gliders, National mooring network, Coastal HF radar, Autonomous underwater vehicle, Acoustic tagging and monitoring of fish, Automated intelligent sensor networks, Satellite observing systems) and an eleventh facility, the eMarine Information Infrastructure (eMII) for IMOS data management and integration. Whilst facilities are national entities, the science rationale is delivered through five nodes - the ocean 'bluewater and climate' node, and four regional nodes Western Australia (WA-IMOS), Southern Australia (SA-IMOS), New South Wales (NSW-IMOS) and the Queensland (QLD) Great Barrier Reef (GBROOS), Fig. 1.

3. INFORMATION INFRASTRUCTURE

Marine data and information are the main products of IMOS and data management is therefore a central element to the project's success. eMII provides a single integrative framework for data and information management that will allow discovery and access of the data by scientists, managers and the public. The initial strategy has focussed on defining specific data streams and developing end-to-end protocols, standards and

Figure 1. Schematic of IMOS showing Nodes and Facilities

systems to join the related observing systems into a unified data storage and access framework.

IMOS data streams can be categorized as:

1) gridded data from satellites and HF radar systems
2) timeseries data from moorings, argo floats, gliders and ships of opportunity
3) image data from Autonomous Underwater Vehicles
4) biological data from continuous plankton recorders and acoustic tagging.

1) and 2) provide real-time and delayed-mode data sets whereas 3) and 4) are delayed mode delivery only.
The IMOS data management infrastructure employs Open Geospatial Consortium (OGC, www.opengeospatial.org) standards wherever possible.

A robust working infrastructure for end-to-end data management, search, discovery and access to IMOS data is now in existence. The components consist of:

- A distributed data storage network
- A metadata catalog holding ISO 19115 standard records (the GeoNetwork Metadata Entry and Search Tool (MEST)), incorporating SensorML, which provides standard models and an XML encoding for describing sensors and measurement processes
- An ocean portal for map-view access to the data. This provides
  - A national facility view of data
  - A regional node view of data
  - A real-time view of data
  - A search and discovery view of data

Data may be downloaded either from the portal or from the MEST. All components are built on open source software and international standards adhered to.

3.1 The distributed data network

A distributed data storage system has been developed in association with the Australian Research Collaboration Service (ARCS). This has involved utilizing the ARCS Data Fabric, a 'cloud' storage system.

Storage facilities exist at supercomputing centres within each of the four IMOS regional nodes (WA – iVEC, SA – eResearchSA, NSW – Intersect, QLD – Queensland Cyber Infrastructure Foundation) and at TPAC, all linked through the Australian Academic and Research Network (AARNET) fibre optic backbone (10GBit Bandwidth between mainland sites). This was necessary to overcome the potential problem of storing all data at the Tasmanian Partnership for Advanced Computing (TPAC) (as originally intended) and experiencing a bottleneck in access to data due to the restricted bandwidth across Bass Strait.

Recognizing that a significant proportion of IMOS data is of either gridded (satellite, HF radar) or timeseries (Argo, ship of opportunity, gliders, moorings, networked sensors) form and could sensibly written into a self describing format (netCDF) meant that advantage could be taken of emerging web services to access these data through OPeNDAP/THREDDS servers. Both netCDF format data and non-netCDF format data (e.g. AUV imagery) can be accommodated within the ARCS Data Fabric.

At all sites a uniform data management system is installed ensuring consistency across the Data Fabric. A MEST instance at each site accumulates metadata records for data loaded at that site and routine harvesting of records from all components of the system to the TPAC MEST ensures a complete 'master' catalog of IMOS data is kept up to date. To ensure maximum machine functionality, eMII has produced a system of data management procedures, implemented by all IMOS facilities, which includes a) a procedure and filenaming convention for uploading, archiving and storing accessible data, b) a prescribed netCDF format for creating datasets which incorporates all necessary data to generate a metadata record conforming to ISO standards, a record which can be automatically created from the netCDF file and uploaded to the MEST. Manual creation of metadata records for non-netCDF formatted data is still required, and templates have been (or are in the process of being) created for these data types. A practice of mirroring of data between sites (for security and in case of link failures) is under development.

3.2 The metadata catalog holding ISO standard records - the GeoNetwork Metadata Entry and Search Tool (MEST)

The opensource GeoNetwork (http://geonetwork-opensource.org/) Metadata Entry and Search Tool (MEST) is used for metadata cataloguing. Much of the development work for this tool was carried out by the BlueNet project (www.bluenet.org.au), an Australian project intended to encourage data sharing across Australian universities. Enhancements to the IMOS MEST, via the GeoNetwork open source community trunk, have included

- Improved file upload and download functions, including logging of upload/download-related information, the capacity to overwrite an uploaded file (or not), the capacity to select and simultaneously download numerous files attached to one metadata record
- Enhanced "Advanced search" options (search for metadata that has data attached, search for data containing particular parameters. Improved keyword searching)
- Addition of a syntax for external webpages to link directly to a specific metadata record, or to a search-result
- Addition of the capacity to display map layers in the MEST's InterMap, for metadata records referring to Web Map Service (WMS) Getcapabilities URLs
- Addition of a 'data parameters' metadata block to the Marine Community Profile (MCP), the schema for marine metadata developed by the Australian Ocean Data Centre Joint Facility (AODCJF, www.aodc.org.au)
- Addition of an improved "Use constraints" metadata block, and inclusion of a "Terms of Use" agreement that can be set to download when a file is downloaded from the MEST
- Improved harvesting between MESTS, using Webdav, and from Catalog Services for the Web (CSW) nodes.
- Improved error-message wording and display
- Harmonised SensorML with 1.0.121 schemas
- Addition of CSW support for MCP, SensorML and World Meteorological Office (WMO) schemas.

3.3 The IMOS Ocean Portal

The Portal was developed under certain constraints, to:

- utilize open source code only; provide an easy-to-use interface for users (browser based)
- have the ability to manage multiple concurrent requests
- have scalability to cope with an increase in concurrent requests
- provide a good response at low bandwidth
- use a common and current development platform
- run in a Linux environment.

This resulted in the Portal being developed entirely using open source code on a Java-based AJAX framework ZK (www.zkoss.org). It adopts an intuitive map-centred interface, based on OGC Openlayers (http://openlayers.org/), Fig. 2, and can be deployed on any modern computing platform (Solaris, Linux, Windows, Apple OSX). The Portal offers access to data in a number of ways

- real-time data is available via DataTurbine (see below)
- via the map interface (utilizing the OpenLayers approach) different datastream timeseries can be animated
- users can change the style and opacity of layers non-IMOS data can be added via OGC Web Map Services
- data can quickly be accessed and downloaded
- a link to the IMOS MEST enables search and discovery by simple text search, by geographic area, by temporal extent, or by any combination of these. For more detailed search requirements a direct link to the IMOS MEST opens an additional window for complex searching; data with MEST records linked to a WMS can be added to the map.

Access to real-time or near real-time data, such as the frequently updated observations from the National Reference Stations and the networked sensors on the

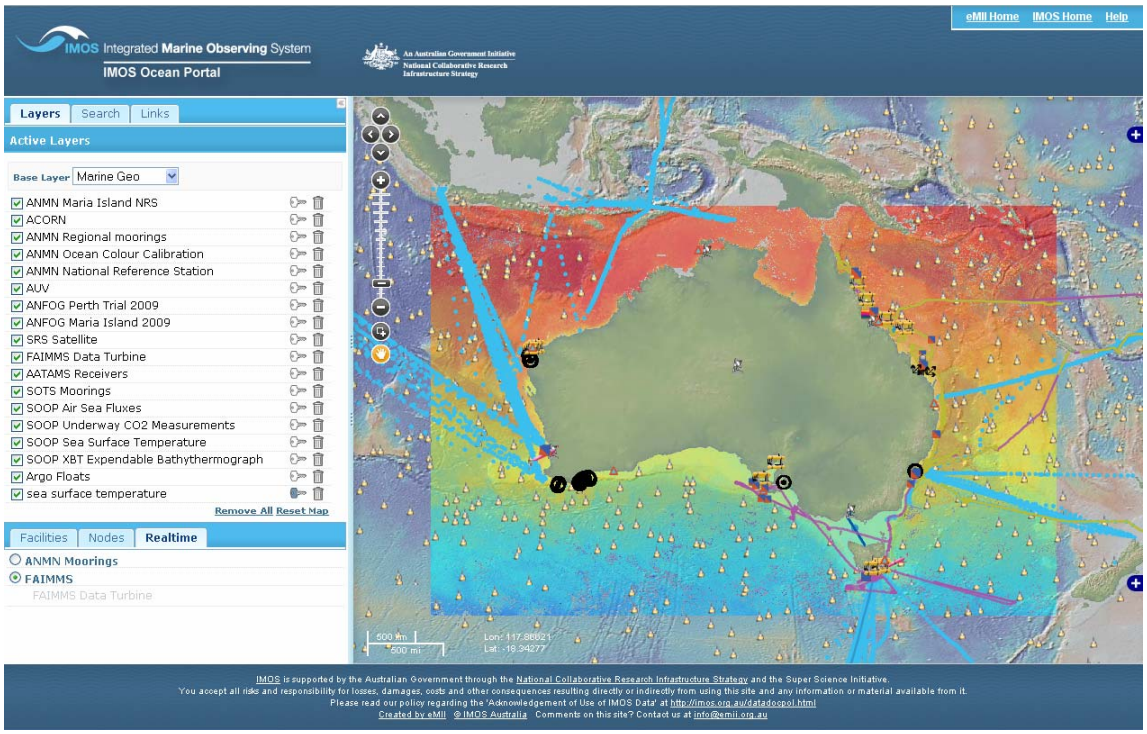

Figure 2. The IMOS Ocean Portal

Great Barrier Reef (from the IMOS Facility for Automated Intelligent Monitoring of Marine Systems (FAIMMS)) is provided through DataTurbine, 'a robust open-source streaming data middleware system that satisfies the core requirements for sensor-based environmental observing systems' (quote from www.dataturbine.org). The core of DataTurbine is a 'ring-buffer'- a circular buffer which stores a predefined amount of the latest data (say, for one month) after which the oldest data is overwritten by new data. It can be used for numeric data, still images or video and data can also be written to a database or archived on disk. This flexibility makes it ideal for the multi-variate data collected by IMOS.

Within the Portal a 'How do I?' help facility is provided and there is a publicly accessible Trac site with Wiki to allow discussion. Feedback on the system (e.g. suggestions, bugs, criticism) is handled through the eMII Helpdesk (info@emii.org.au). Use of the Portal and MEST are monitored by AWSTAT monitoring software, providing a comprehensive daily breakdown of activity.

The Ocean Portal can be accessed via http://imos.aodn.org.au and the IMOS MEST via http://imosmest.aodn.org.au. All IMOS data are freely available without constraints and are obtainable through a simple self registration process.

Data storage and retrieval in IMOS is designed to be interoperable with other national and international programs. Thus, it will be possible to integrate data from sources outside IMOS into IMOS data products, and IMOS data will also be exported to international programs such as Argo, Oceansites. The real-time physical parameters data will be exported to the WMO's Global Telecommunications System (GTS).

4. CONCLUSION

As the IMOS program gains momentum the concept of data sharing and its value is spreading across Australia. The long-term view of the data management infrastructure developed for IMOS is that it will become the infrastructure of the Australian Oceans Data Network (AODN), exposing marine data from Commonwealth and State agencies, universities and the community at large.