**Advances in
Geosciences**

A data delivery system for IMOS, the Australian Integrated Marine Observing System

R. Proctor, K. Roberts, and B. J. Ward

Integrated Marine Observing System, University of Tasmania, Hobart 7001, Tasmania, Australia

Abstract. The Integrated Marine Observing System (IMOS, www.imos.org.au), an AUD $150 m 7-year project (2007–2013), is a distributed set of equipment and data-information services which, among many applications, 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 through 11 facilities which effectively observe and measure the 4-dimensional ocean variability, and the physical and biological response of coastal and shelf seas around Australia. Through a national science rationale IMOS is organized as five regional nodes (Western Australia – WAIMOS, South Australian – SAIMOS, Tasmania – TASIMOS, New SouthWales – NSWIMOS and Queensland – QIMOS) surrounded by an oceanic node (Blue Water and Climate). Operationally IMOS is organized as 11 facilities (Argo Australia, Ships of Opportunity, Southern Ocean Automated Time Series Observations, Australian National Facility for Ocean Gliders, Autonomous Underwater Vehicle Facility, Australian National Mooring Network, Australian Coastal Ocean Radar Network, Australian Acoustic Tagging and Monitoring System, Facility for Automated Intelligent Monitoring of Marine Systems, eMarine Information Infrastructure and Satellite Remote Sensing) delivering data. IMOS data is freely available to the public.

The data, a combination of near real-time and delayed mode, are made available to researchers through the electronic Marine Information Infrastructure (eMII). eMII utilises the Australian Academic Research Network (AARNET) to support a distributed database on OPeN-DAP/THREDDS servers hosted by regional computing centres. IMOS instruments are described through the OGC Specification SensorML and where-ever possible data is in

CF compliant netCDF format. Metadata, conforming to standard ISO 19115, is automatically harvested from the netCDF files and the metadata records catalogued in the OGC GeoNetwork Metadata Entry and Search Tool (MEST). Data discovery, access and download occur via web services through the IMOS Ocean Portal (http://imos.aodn.org.au) and tools for the display and integration of near real-time data are in development.

1 Introduction

In recent years Australia has commenced the establishment of marine research-infrastructure around the country to systematically service Australia's significant requirements and responsibilities for one of the largest marine jurisdictions of any nation on earth (Myers, 2008). At over 14 million km^2 Australia's Exclusive Economic Zone (EEZ) is nearly twice the surface area of the Australian continent. It extends from the tropics to Antarctic waters and much of it is unexplored.

The surrounding Pacific, Southern and Indian Oceans strongly affect the continental climate-system at all time scales, from seasons to decades. The major ocean currents on its eastern, western, northern and southern boundaries, (best known of these being the East Australian Current and the Leeuwin Current in the west, Fig. 1), affect regional climatic conditions and marine ecosystems. There is evidence that these currents are changing on decadal time scales and have already impacted marine ecosystems, but the data are sparse and the currents and ecosystems have not been monitored in a systematic way. Management of climate impacts and sustainable use of the marine environment are major concerns in Australia, providing the rationale for large investments in the infrastructure to support relevant research. A challenge for IMOS from the beginning was to develop a broad consensus in the marine research community on a goal, national in scope that would provide the basis for a national approach to
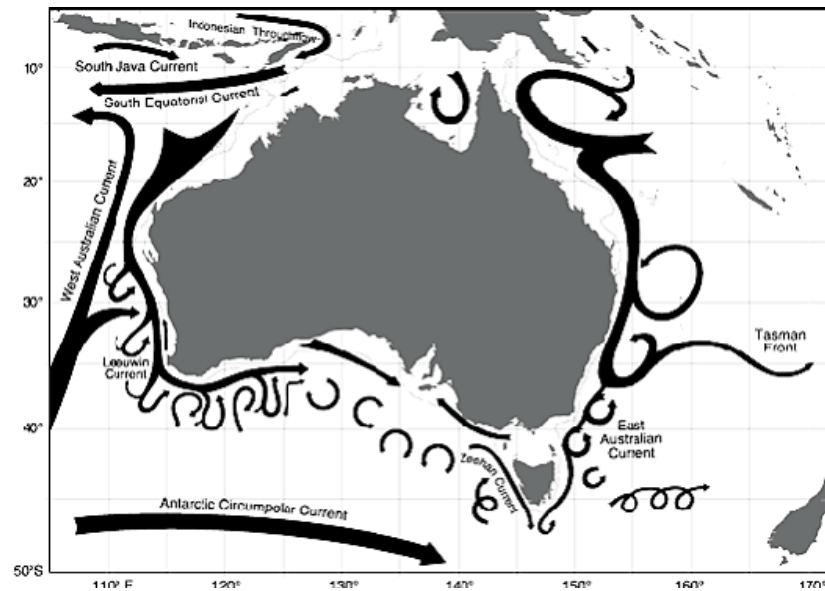
Fig. 1. Boundary currents around Australia.

marine observing. At the highest level that goal is to observe and support research on the impact of major boundary currents and regional ocean circulation on marine ecosystems and terrestrial climate (Hill, 2010).

The in situ data when combined with satellite data, enables the modeling required to explain the role of the oceans in seasonal prediction and climate change. Sustaining the project will allow identification and management of climate change in the coastal marine environment. It will also provide an observational nexus to better understand and predict the fundamental connections between coastal biological processes and regional/oceanic phenomena that influence biodiversity. IMOS was primarily designed to support research; however, the data streams are also critical for societal, environmental and economic applications. Some of these include: management of marine natural resources and their associated ecosystems, support and management of coastal and offshore industries, safety at sea, marine tourism and defense.

Given the extent and challenge of addressing the broad range of marine issues in the Australian EEZ, IMOS is considered only the beginning of the observing system that Australia needs. The cost of an adequate observing system will be high due to the great length of coastline and the relatively small population and economy. Never-the-less, staged enhancements are being planned. The return from investing in ocean observations around Australia was estimated through an economic analysis undertaken in 2006 (ATSE, 2006). That study, based on only a limited set of benefiting industries concluded that the cost:benefit to the Australian economy of investing in ocean observations was better than 1:20.

IMOS at the present time has five regional Nodes covering the Queensland (QIMOS), New South Wales (southeastern Australia, NSWIMOS), Tasmania (TASIMOS), Southern Australia (SAIMOS) and Western Australia (WAIMOS) together with the oceanic Bluewater and Climate Node (Fig. 2), each with well-defined science objectives. The national observing Facilities delivering the necessary observations to meet these objectives include three for *bluewater and climate observations* (Argo Australia, Enhanced Measurements from Ships of Opportunity and Southern Ocean Time Series), three facilities for *coastal currents and water properties* (Moorings, Ocean Gliders and HF Coastal Radar), three for *coastal ecosystems* (Acoustic Tagging and Tracking, Autonomous Underwater Vehicle and a biophysical sensor network on the Great Barrier Reef) and an assembly centre for remote sensing data from satellites. The IMOS Facility concerned with data management, the eMarine Information Infrastructure (eMII), provides access to all IMOS data, and data services to all users.

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 focused on defining specific data streams and developing end-to-end protocols, standards and systems to join the related observing systems into a unified data storage and access framework.
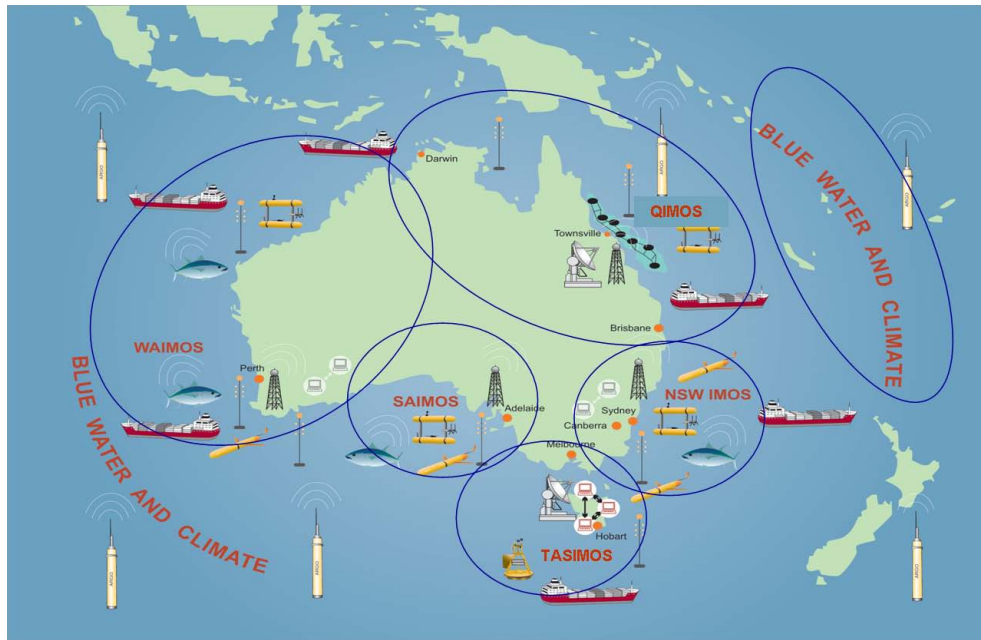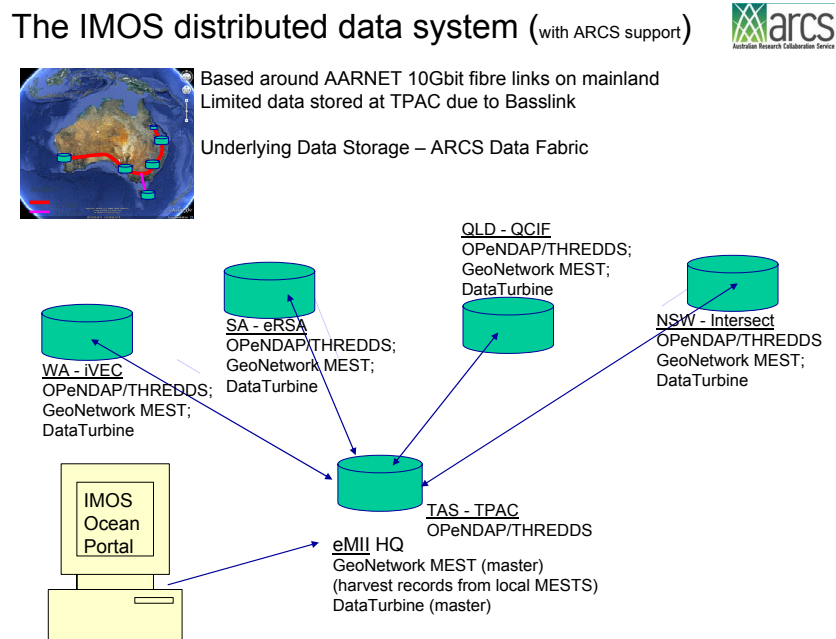
Fig. 2. IMOS Nodes and Facilities.

Fig. 3. The IMOS distributed data network.

2 Data management infrastructure

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 19115/19139 standard records (MEST)

– 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 through the portal or from the MEST

All components are built on open source software and to international standards.

2.1 The distributed data network

A distributed data storage system (Fig. 3) has been developed in association with the Australian Research Collaboration Service (ARCS). This has involved utilizing the ARCS Data Fabric, a "cloud" storage system, i.e. the location of data across multiple platforms is invisible to the user.

Working with ARCS eMII has established data storage facilities at the regional high performance computing centres within each of the four mainland IMOS regional nodes (WAIMOS, SAIMOS, NSWIMOS and QIMOS) with additional storage in Tasmania at the Tasmanian Partnership for Advanced Computing (TPAC), all linked through the AARNET fibre optic backbone (10GBit Bandwidth between mainland sites). This was necessary to overcome the potential problem of storing all data at TPAC (as originally intended) and experiencing a bottleneck in access to data due to the restricted bandwidth (Basslink, 310 MBit) across Bass Strait between Tasmania and mainland Australia.

Recognizing that a significant proportion of IMOS data is of either gridded (satellite, HF radar) or in timeseries form (Argo, ship of opportunity, gliders, moorings, networked sensors) 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 OPeN-DAP/THREDDS servers. Both netCDF format data and non-netCDF format data (e.g. Autonomous Underwater Vehicle 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. At each site a metadata catalog and metadata search and discovery tool, based on the GeoNetwork Metadata Entry and Search Tool (MEST), accumulates metadata records for data loaded at that site. 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 file-naming 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 19115/19139 standards, a record which can be automatically created from the netCDF file and uploaded to the MEST (see documentation

at http://www.imos.org.au/datadocpol). 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.

2.2 The IMOS MEST, the metadata catalog holding ISO standard records – the GeoNetwork Metadata Entry and Search Tool (MEST)

The IMOS MEST allows manual and automatic upload of metadata records (and data sets) and provides sophisticated options for data discovery, access and download. Work in eMII and a preceding project BlueNet (www.bluenet.org.au) has provided many enhancements to the GeoNetwork open source community trunk (http://geonetwork-opensource.org/). These 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 WMS Getcapabilities URLs

– addition of a "data parameters" metadata block to the Marine Community Profile (MCP)

– 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 CSW nodes. Also added fix to ensure metadata record's ownership information is maintained post-harvest

– modification of the classification of and searching for online resources, so that data on the ARCS system is retrieved in a "find data" search

– addition of the capacity to "select" metadata records of interest, and view, print or save this subset

– improved error-message wording and display

– harmonised SensorML with 1.0.121 schemas

– addition of CSW support for MCP, SensorML and World Meteorological Organisation (WMO) schemas

– integration of enhancements from the GeoNetwork trunk software into the IMOS MEST.

For its metadata recording IMOS utilises the Marine Community Profile, a subset schema to the full ISO 19115 standard. The MCP was defined by the Australian Ocean Data Centre Joint Facility (AODCJF) (see www.aodc.org.au). Recent changes to the MCP by eMII include

– addition of a "Data parameter" block;

– modifications to the "Resource Constraints" section, to include a "U-nDP(Data Commons)" Licensing option, and re-arrangement/improvement to include distinct fields for each of "attribution", commercial" "derivatives" and "collective works" constraints.

– modifications to the following codelists

– the "onlineResource" protocol (to include "Data URL" and "Point-of-truth Metadata URL"

– the CitedResponsibleParty Role (to include co-investigator, research-assistant, moralRightsOwner, IP-owner, MetadataContact)

– the "Metadata information: HierarchyLevel" codelist (added 'Publication')

– modification to the list of inclusions for "Core" and "Minimum" (for instance "Core" now includes attached files);

– addition of 2 sensor-related fields: "sensor", and "sensorCalibration"

The inclusion of the U-nDP licencing option, whilst not required for IMOS data (IMOS data is freely available to all), is seen as a necessary attribute for future data acquisition of non-IMOS funded data.

2.3 The IMOS Ocean Portal

A critical factor in the design of an infrastructure for accessing IMOS data was the provision of an intuitive approach for rapid search and access to data. Certain constraints on the Portal development were established by the AODCJF Technical Committee. These were:

– 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.

Under these constraints the Portal was developed. In brief,

– the Portal is developed entirely using open source code on a Java-based AJAX framework ZK which is used by many large organizations

– it adopts a map-centred interface, based on OGC Openlayers, which is intuitive and in testing (e.g. data user workshops) has been used by people from a broad range of scientific and non-scientific backgrounds with ease

– it has shown excellent response times to groups of up to 75 con-current users and through installation on dedicated enterprise-scale servers no difficulties with up-scaling are expected

– the Portal can be deployed on any modern computing platform (Solaris, Linux, Windows, Apple OSX)

– the production system is currently running on Red Hat Enterprise Linux.

The Portal offers access to data in a number of ways:

– via the map interface (utilizing the OpenLayers approach) different datastream timeseries can be animated

– real-time data is available via DataTurbine (see below)

– 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, either from the OPeNDAP/THREDDS servers or from the Data Fabric

– 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, allowing users to gain access to further information and/or the data for download

– 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

– a Facility and a Regional view.

2.4 Access to real-time data

For accessing real-time or near real-time data, such as the frequently updated observations from the National Reference Stations and the networked sensors on the GBR, the middleware DataTurbine is used. This is "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.

2.5 User support

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 (telephone or email: info@emii.org.au). Use of the Portal and MEST are monitored by AWSTAT monitoring software which provides a comprehensive daily breakdown of use, yielding data on the range, scope and frequency of interest in IMOS data.

To facilitate scientific use of IMOS data, and improve efficiency of data delivery to the IMOS portal, a range of software tools are being (or planned to be) developed. For example: DataTurbine provides the Real-timeDataViewer (RDV), a Java powerful and flexible tool for panning through multiple timeseries; web-browser data entry tools; a Matlab toolbox for end-to-end (i.e. sensor-to-netCDF) processing of timeseries data (see http://code.google.com/p/imos-toolbox); visualisation tools for glider and HF radar. All tools to be freely available from the portal.

3 Summary

The Portal was publicly released on 29 June 2009. The Ocean Portal can be accessed via http://imos.aodn.org.au and the IMOS MEST via http://imosmest.aodn.org.au. 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 (argo.jcommops.org) and Oceansites (www.oceansites.org). Also, most of the real-time data of physical parameters will be exported to the Global Telecommunications System. Recent additional investment in IMOS by the Australian Education Investment Fund will lead to the establishment of a Tasmanian Node and enhancement of the IMOS infrastructure in line with the Commonwealth budget imperatives to place more emphasis on observing in Northern Australian waters and the Southern Ocean.

References

ATSE: Economics of Australia's sustained ocean observation system, benefits and rationale for public funding, Report for the Australian Academy of Technological Sciences and Engineering and the Western Australian Global Ocean Observing System Inc., August 2006, 44 pp., 2006.

Hill, K.: The Australian Integrated Marine Observing System (IMOS). Meteorological Technology International Annual Showcase, 114–118, http://viewer.zmags.com/publication/234fa028#/234fa028/1, 2010.

Meyers, G.: The Australian Integrated Marine Observing System, J. Ocean Technol., 3, 80–81, 2008.