# The Marine and Climate Data Discovery and Access Project (MACDDAP)

**Raymond Williams[1,2], Pauline Mak[1,4], Peter Blain[1], Nathan Bindoff[1,3,5,6,7]**

[1]Tasmanian Partnership for Advanced Computing (TPAC), University of Tasmania, Hobart, Tasmania
[2]School of Computing and Information Systems, University of Tasmania, Hobart, Tasmania
[3]Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC), Hobart, Tasmania
[4]Australian Research Collaboration Service (ARCS)
[5]Institute for Antarctic and Southern Ocean Studies (IASOS), University of Tasmania, Hobart, Tasmania
[6]CSIRO Marine and Atmospheric Research, Hobart, Tasmania
[7]CAWCR

## Introduction

The Marine and Climate Data Discovery and Access Project (MACDDAP) is an e-Research project, funded by the National eResearch Architecture Taskforce (NeAT) under the National Collaborative Research Infrastructure Strategy (NCRIS). Its objective is to integrate the large marine and climate data sets, currently distributed across Australian research institutions, using web services technology. The ultimate aim is to create a "virtual database" enabling researchers to collect, combine and analyse relevant data across scientific disciplines to facilitate knowledge discovery for marine- and climate-related applications.

The project is building on open scientific and geospatial data standards to enhance specialised web harvesters and search tools, enabling them to handle large geospatial data-sets, and is creating web services to deliver scientific data to users via appropriate web portals. To provide the functionality required to support these services, MACDDAP is in the process of developing an aggregator for combining geospatial data from distributed sources and a translator for translating data sets into a single standard vocabulary widely used in meteorology and oceanography.

## Data Streams

Initially, MACDDAP will make available, or enhance the availability of, a number of Australian marine and climate datasets, including remote sensing and oceanographic data streams from the Integrated Marine Observing System (IMOS), ocean climate and weather data from the Bureau of Meteorology, earth system science modelling data from TPAC and marine data sets from the Australian Oceanographic Data Network (AODN). It is anticipated that other data sets will be made available, via MACDDAP services, in the future.

## Sub-projects

The MACDDAP project comprises several main sub-projects:

*The IMOS-GeoNetwork MEST Sub-project.*
The IMOS-GeoNetwork MEST [1] is a metadata catalogue and registry with facilities to harvest, translate, edit and store metadata. This sub-project is enhancing these to improve operability with OPeNDAP and Open Geospatial Consortium (OGC) data servers, by providing an enhanced harvest manager with update and version support, better support for translation between metadata standards and an improved editor. It will also incorporate Australian Access Federation (AAF) standards for user authentication and trusted communications with other services.

*The TPAC Digital Library Sub-project.*
The TPAC Digital Library [2] [3] comprises a web portal and data harvester, based on an OPeNDAP server, which brings together data resources from major ocean- and climate-related research organizations across Australia. This sub-project is providing enhanced functionality for searching over latitude, longitude and time, within data categories and across categories. The Digital Library will also be modified to provide AAF interoperability and the search capabilities of the data harvester will be improved so that it is able to handle datasets containing millions of files. Finally, an administrator's monitoring and notification service will be added to the portal to enable an administrator to regularly monitor usage of the Digital Library.

*The Aggregation Services Sub-project.*
This sub-project is developing a spatial data aggregator, capable of operating over an OPeNDAP network, which will allow researchers to combine distributed data without the need to understand the underlying data structures. It employs a geospatially-aware data harvester to create a PostGIS database of the available data, which can then be searched by the user, for any period and within any geographical bounding box, to provide an aggregated file which is then served via the OPeNDAP protocol.

*The OPeNDAP-OGC Integration Sub-project.*
The IMOS marine data management project has specified web services as the standard for delivery of scientific data and the IMOS-GeoNetwork MEST contains and connects with external OGC web services for this purpose. This sub-project has provided OGC Web Map Services for data viewing, together with OGC Web Feature and Web Coverage Services for data delivery, with the option to layer these services over OPeNDAP servers.

*The OpeNDAP Enhancements Sub-project.*
This sub-project is making modifications to the OPeNDAP servers to enhance their reliability, security and operability with AAF, so they are able to deliver petabyte data stores to the research community. These enhancements will include user authentication of programmatically generated data sets, trusted service communications between font-end, back-end and gateway servers and between the Digital Library and other OPeNDAP servers, as well as administrative monitoring, control, and configuration interfaces.

*The Translation Services Sub-project.*
Many valuable marine- and climate-related data sets have been accumulated over recent decades, but they have been stored in different file formats and described using various metadata schemes, severely restricting their general accessibility. The process of conforming millions of metadata records to required standards is extremely costly if undertaken manually. New contexts will arise with an increase in interdisciplinary research and invariably standards will evolve and new ones will emerge, so the cost of conformance is recurrent. Such costs are an impediment to the increased reuse of these data sets. The sub-project is creating a web service that will allow the translation of pre-existing data sets into a common standard widely used in meteorology and oceanography, greatly facilitating dataset discovery and search.

## Conclusion

The MACDDAP project aims to utilise international and national data standards, including the OPeNDAP standard protocol for scientific data exchange, the OGC standards for geo-spatial data exchange and the AAF national security standards for single sign-on across federated databases, to support the creation of aggregation and translation web services, which will then be made available to marine and climate researchers via the IMOS GeoNetwork MEST, the TPAC Digital Library Portal and other portals.

Customer satisfaction with the progress of the project and the quality of project outputs is being evaluated by seeking feedback from user organizations via meeting discussions, e-mail lists, website responses and training workshops, as well as by participation at national and international conferences associated with the relevant research communities. Portal use and data delivery will be monitored by obtaining usage statistics for the IMOS GeoNetwork and the TPAC Digital Library and by using tools such as Google-Analytics and Web-Analyser.

MACDDAP is closely linked with international standards organizations, including OPeNDAP Inc, the developers of the OPeNDAP protocol, and GeoNetwork, the developers of the underlying catalogue tool used in the project, as well as the University of Colorado Atmospheric Research and the University of Reading eScience Centre, who are developing integrated web services for the OPeNDAP-THREDDS environment.

The ultimate desired outcome for the MACDDAP project is for marine and climate data throughout Australia to be discoverable, searchable and conformable with standard vocabularies, enabling researchers to collect and aggregate data across disciplines for knowledge discovery.

## References


1. *IMOS-GeoNetwork MEST*. Available from:http://imosmest.emii.org.au/ accessed 5 June 2009.
2. Cumming, I. and G. Hyland, *Ocean and Climate Digital Library Portal*, in *Proceedings of the Australian Partnership for Advanced Computing Conference (APAC'05)*, Gold Coast, Australia, Sep 2005.
3. *TPAC Digital Library*. Available from: http://digitallibrary.tpac.org.au/ accessed 5 June 2009.