

Outcomes of the Marine and Climate Data Discovery and Access Project (MACDDAP)

Peter Blain¹, Paola Petrelli², Nathan Bindoff⁴, Jason Lohrey⁵

¹University of Tasmania, Hobart, Australia, peter.blain@utas.edu.au

²University of Tasmania, Hobart, Australia, paola.petrelli@utas.edu.au

³University of Tasmania, Hobart, Australia, nathan.bindoff@utas.edu.au

⁴Arcitecta, Melbourne, Australia, jason.lohrey@arcitecta.com

Introduction

The Marine and Climate Data Discovery and Access Project (MACDDAP) was an e-Research project, funded by the National eResearch Architecture Taskforce (NeAT) under the National Collaborative Research Infrastructure Strategy (NCRIS). The project was completed in June 2011 and successfully delivered on its stated objective, which was to integrate large marine and climate data sets, and to deliver them through a wide range of data streams - thus engaging a broad community. The project built on web services technology to integrate marine and climate data sets distributed across Australian research institutions. The outputs delivered by MACDDAP facilitate knowledge discovery for marine and climate related applications by enabling researchers to collect, combine and analyse relevant data across scientific disciplines. MACDDAP has built on open scientific and geospatial data standards to enhance specialised web harvesters and search tools, to deliver large geospatial data-sets to users via web portals. MACDDAP also provides the functionality required to support these services, including an aggregator for combining geospatial data from distributed sources, and a translator for translating data sets into standard vocabularies used in meteorology and oceanography.

Data Streams

MACDDAP has made available, or enhanced the availability 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, marine data sets from the Australian Oceanographic Data Network (AODN), and others.

Sub-projects

The MACDDAP project comprises several main sub-projects:

The GeoNetwork MEST Sub-project.

GeoNetwork is a cataloging application for spatially referenced resources. A version of the software is deployed at the Integrated Marine Observing System (IMOS), and the Australian Ocean Data Network (AODN), to support initial data discovery and access for marine data in Australia. The MACDDAP project streamlined the cataloging of THREDDS, OGC WFS and OGC SOS services in GeoNetwork, which has led to improved data availability within the IMOS and AODN portals. The search capability was enhanced, allowing researchers to more easily find datasets relevant to their research.

The TPAC Digital Library Sub-project.

The digital library portal provides the marine and climate scientific communities with ready access to heterogeneous and geographically distributed ocean and climate datasets. It provides a user interface for viewing, searching, and accessing data collections, as well as enabling data services for user access. Files can be downloaded or accessed through OPeNDAP, GridFTP, WCS, Matlab and other methods. The digital library portal's administration console provides management functions for librarians of digital data collections. It allows interoperability with library service protocols, as well as other data archive organizations and scientific bodies, for data reference transparency. The MACDDAP project has provided access to large, and previously unmanageable, datasets within the TPAC digital library portal. Examples include Argo, WOCE and MODIS satellite images. This was achieved by performance improvements to the harvester and database, and also by improving the user interface, the search capability, and the management services for dataset owners.

The Aggregation Services Sub-project.

This MACDDAP sub-project has improved the usability of marine and terrestrial satellite data through the AO-DAAC aggregation service, which can now aggregate data subsets (that match OPeNDAP constraint URLs) into a single output

file with a common set of axes. The aggregation software is improving access to data at IMOS and TERN/AusCover. There are now 203 products available, and more will soon come online.

The OPeNDAP-OGC Integration Sub-project.

THREDDS Data Server (TDS) publishes self-describing scientific datasets through a suite of services, such as OPeNDAP and the THREDDS catalog. TDS supports over 20 file formats, including NetCDF, HDF and GRIB. The OGC Web Map Service (WMS) is a web-based ISO standard for requesting geospatial images. ncWMS is an implementation of a WMS server from the Reading e-Science Centre(ReSC) that can generate images from any of the file formats understood by TDS, as well as remote OPeNDAP resources. The MACDDAP project added ncWMS to THREDDS. The result is that any gridded dataset hosted on THREDDS, with an identifiable coordinate system, is potentially available to any WMS client.

The OPeNDAP Enhancements Sub-project.

This sub-project enhanced Hyrax and THREDDS OPeNDAP servers to improve reliability, security and operability with AAF, to facilitate the delivery of petabyte data stores to the research community. The enhancements included user authentication of programmatically generated data sets, trusted service communications between front-end, back-end and gateway servers and between the TPAC Digital Library and other OPeNDAP servers, as well as administrative monitoring, control, and configuration interfaces. The MACDDAP enhancements will provide access to datasets such as NMOG forecast and analysis products; global, regional, city atmospheric prediction products, ocean prediction products; wave prediction products; multiple years of ocean model analysis and observation assimilation analysis; multiple years of high & low resolution SST satellite products (L2P, L3, L4); new TERN satellite products; seasonal forecast products; Tsunami scenario database for inundation modelling.

The Translation Services Sub-project.

The standards for describing research (and other) data are evolving constantly. For each new, or evolution of an existing standard, there are millions of data sets that need to be conformed to those standards. The cost for achieving conformance can be high, particularly when manual intervention is required. The Translation Services sub-project minimises the effort and cost for converting to different metadata standards. The Translation Services have initially been applied to conforming THREDDS catalogs to NetCDF Climate and Forecast (CF) Metadata Convention. In future, input and output “drivers” will be created for other standards such as ISO 19115, OGC, and others. The translation services are being made available (via the web) to the wider scientific data management community. They are currently providing standardised access to many (previously non-conforming) data collections available at the TPAC Digital Library Portal.

Conclusion

The MACDDAP project has built on 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 institutions, to support the creation of aggregation and translation web services, which are now available to marine and climate researchers via the IMOS GeoNetwork MEST, the TPAC Digital Library Portal and other portals. MACDDAP has been 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 MACDDAP project has delivered on its desired outcome, which 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.

About the Authors

Peter Blain.

Dr Blain is the software development manager and software architect at the Tasmanian Partnership for Advanced Computing (TPAC) at the University of Tasmania. He is the project manager for a number of eResearch projects at TPAC, including the Marine and Climate Data Discovery and Access Project (MACDDAP). Dr Blain received his PhD from the School of Computing and Information Technology at Griffith University in 2007, and a Bachelor of Engineering in computer systems from the University of Queensland in 1992. Prior to joining TPAC in 2008, he worked as a freelance software engineer/consultant for large Australian and International financial services companies including the Bank of Tokyo, HSBC, Westpac, the Commonwealth Bank of Australia, and the National Australia Bank.

Paola Petrelli.

Dr Petrelli is the earth system data librarian at the Tasmanian Partnership on Advanced Computing (TPAC) at the University of Tasmania. She is the data manager for the TPAC Oceans and Climate Digital Library Portal. Dr Petrelli received a PhD from the Department of Earth Science of the University of Siena (Italy) in 2005, and a Bachelor in Marine Environmental Sciences at the University of Venice (Italy) in 1999. Her research interests included modeling ocean and atmosphere interactions in Antarctica and sea ice processes. For the past 4 years she have been managing oceanographic and climatological datasets, acquiring extensive experience in web services and software used by the earth science research community.

Jason Lohrey.

Jason Lohrey is the CTO of Arcitecta and the conceiver and architect of Mediaflux™. Jason has a degree in Physics and Computer Science augmented with Fine Arts from the University of Tasmania and has worked in the IT industry for approximately 16 years. His background includes industrial, commercial, scientific, and creative applications for computing. For most of his working career, Jason has focused on research, design and development of digital asset management and database systems with companies including Kodak (with the Academy award winning non-linear editing and compositing system, Cineon), Discreet Logic and Silicon Graphics. Five years ago, while in residence at painter Arthur Boyd's property at Bundanon, New South Wales, he penned the first lines of software for Mediaflux™.

Nathan Bindoff.

Professor Bindoff is Professor of Physical Oceanography at the University of Tasmania, and CSIRO Marine Research Laboratories, Director of the Tasmanian Partnership for Advanced Computing.

Professor Bindoff is a physical oceanographer, specialising in ocean climate and the earth's climate system. He was the coordinating lead author for the ocean chapter in the Inter-Governmental Panel on Climate Change (IPCC) Fourth Assessment Report and for the Fifth Assessment Report for the detection and attribution chapter. His current interests are primarily in understanding how the changing ocean can be used to infer changes in atmosphere and whether these changes can be attributed to rising greenhouse gases and projecting future changes and its impacts on regional climates. Professor Bindoff has served on 13 international committees, six of which are still current, invited speaker at 18 conferences and workshops, co-chaired 2 workshops and was guest editor on two special volumes of Deep Sea Research, and convened Oceans session of the Climate Change Congress, Copenhagen March 2009. He has published more than 64 scientific papers and 35 reports. He established the programs and experiments that determined the total production of Adelie Land Bottom Water formation and its contribution Antarctic Bottom Water Formation, contributed to the development of some of the largest and highest resolution model simulations of the oceans and was deeply involved in oceanographic data and data management as the chairman of the Data Products Committee for the World Ocean Circulation Experiment and the International Polar Year.