TPAC Digital Library Portal Development and MACDDAP

TPAC hosts a digital library web portal that provides the marine and climate scientific communities with ready access to a large number of heterogeneous and geographically distributed ocean and climate datasets. There is an associated data harvester that populates the portal with new datasets, and checks for updates. The portal and harvester have recently been enhanced to cater for larger amounts of data, and to extend the search and access functionality available to researchers. The development work is being carried out as part of the Marine and Climate Data Discovery and Access Project (MACDDAP) – which is funded by the National eResearch Architecture Taskforce (NeAT) under the National Collaborative Research Infrastructure Strategy (NCRIS).

The digital library portal's presentation tier has been redeveloped to provide an enhanced and more interactive user experience. Figure 1 illustrates the user interface, which includes a file-manager-style tree navigation panel to the left. The panel to the right allows users to search across datasets using criteria such as data type, source, and variables. Users can also search for datasets based on geospatial extents by drawing a search area on the map, or by selecting a search area from the drop-down list. These features help researchers to quickly locate and access the data they require.

Figure 1. TPAC Digital Library Portal – search capability
The portal includes other new features such as the ability to auto-generate Matlab scripts for data access. The re-developed portal retains the functionality that existed in the previous version – such as the shopping cart feature, which allows users to select and manage multiple files. The existing features have been re-implemented under the more interactive interface (see Figure 2.)

![TPAC Digital Library Portal](image)

**Figure 2.** TPAC Digital Library Portal – navigation capability

The redeveloped portal is based on a modern Spring/GWT architecture. In addition to the improved user interface, the migration to Spring/GWT has yielded a hardened enterprise application that is easily scalable and maintainable. Spring is a light-weight open source framework for Java that comprises a number of modules including transaction management, security, and other enterprise-grade features. The Google Web Toolkit (GWT) is a java framework for developing desktop-like web applications (using AJAX). The integrated Spring/GWT architecture reduces the programming effort required to maintain the digital library portal, and improves code quality and reliability through automated unit testing.

Other development has been carried out on the harvester, which scans datasets and updates the portal's database. Bottlenecks at the database were beginning to limit the number of datasets that could be maintained. Performance improvement strategies were implemented that increased the speed of the harvester. The strategies included the use of memory tables and caching. The resulting speed increase has allowed significantly more, and larger, datasets to be harvested and made accessible through the portal. The digital library portal now contains 15 categories, 130 datasets, and 1128978 files. These numbers are set to
expand even further.

Other work being carried out under the broader MACDDAP project is also adding value to the portal. The translation services sub-project, for example, is creating an application that will facilitate the translation of pre-existing data sets and metadata into common standards used in meteorology and oceanography, such as COARDS, ISO 19115 and RIF-CS. Translated datasets will be harvestable and published through the portal in standardised formats. Other work in progress will add security features to the portal – enabling dataset owners to provide information about the dataset and to configure data access rights.

Conclusion

The TPAC digital library web portal is an increasingly powerful e-research tool available to the marine and climate scientific communities. Enhancements under the MACDDAP project have expanded the amount of data that can be provided through the portal. Other MACDDAP enhancements have modernised the architecture, and have enhanced the user interface - making it even easier for researchers to locate and access the data they require.