INSTRUCTIONS TO AUTHORS FOR THE PREPARATION OF PAPERS FOR THE OCEANOBS '09 CONFERENCE PROCEEDINGS

Pauline P. Mak(1, 4), Jon Blower(2), John Caron(3), Ethan Davis(3), Adit Santokhee(2), Nathaniel Bindoff(4,5,6,7,8)

[1] Australian Research Collaboration Services (ARCS), Hobart, Tas, 7001, pauline.mak@arcs.org.au
[2] Reading eScience Centre (ReSC), Environmental Systems Science Centre, University of Reading, Reading, RG6 6AL, UK, jdb@mail.nerc-essc.ac.uk
[3] Unidata, University Corporation for Atmospheric Research (UCAR), Boulder, CO, 80307-3000, USA, caron@unidata.ucar.edu, edavus@unidata.ucar.edu
[4] Tasmanian Partnership for Advanced Computing (TPAC), Hobart, Tas, 7001, Australia, N.Bindoff@utas.edu.au
[5] Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC), Hobart, Tas, 7001, Australia, N.Bindoff@utas.edu.au
[6] Commonwealth Scientific and Industrial Research Organisation Marine Research (CMAR), Hobart, Tas, 7000, Australia, N.Bindoff@utas.edu.au
[7] Centre for Australian Weather and Climate Research (CAWCR), Melbourne, Vic, 3001, Australia, N.Bindoff@utas.edu.au
[8] Institute of Antarctic and Southern Ocean Studies (IASOS), Hobart, Tas, 7001, Australia, N.Bindoff@utas.edu.au

ABSTRACT

THREDDS Data Server (TDS) is a highly flexible framework for publishing datasets over the internet. It has been integrated with the ncWMS server - an Open Geospatial Consortium Web Map Service (OGC WMS) for visualising datasets. This work is now fully integrated into TDS and forms part of the new 4.0 stable release. It is currently used to serving Integrated Marine Information System (IMOS) datasets. Furthermore, it is anticipated to be used by the MyOceans project.

1. THREDDS DATA SERVER (TDS)

TDS is a framework for serving and cataloguing heterogeneous data types through common protocols over HTTP. It is a middleware that simplifies the publication of and access to scientific data [1]. It has a significant global user base with many ocean, climate and modelling communities using this to share data. The main advantage of the TDS server (and also of other OPeNDAP servers) is its use of the Data Access Protocol (DAP) to harmonise the delivery across the internet of a whole suite of self-describing file formats (currently 20 types) commonly used in the these communities. Interoperability is enhanced with the use of NetCDF Markup Language (ncML) [2], where metadata views can be added to conform to a naming convention, while the underlying files remains unchanged. Additionally, ncML offers aggregation of datasets, where large datasets spanning multiple files can be seen as a single logical volume. These capabilities give TDS an enormous amount of flexibility to deliver heterogeneous files from legacy datasets and from diverse applications and sources to across the internet through a uniform interface with simple client applications. The DAP protocol also subset of a dataset to be downloaded, restricted by variable and index ranges. This powerful feature meant users only downloads what is needed.

However, sharing data across discipline, such as the GIS community has been difficult, as the underlying protocol, DAP does not allow data to be referenced in geospatial coordinates. This protocol depends on the structure of the underlying objects and uses exclusively indexes for referencing elements. This makes OPeNDAP extremely flexible, where subset can be retrieved using these indexes and therefore can be used the deliver almost any scientific datasets [3]. The flexibility does come at a cost to the users, as they themselves must map indexes to the real values. While some implementations of OPeNDAP, such as Hyrax, does provide server side function to subset by coordinates, this is not part of the DAP protocol. The lack of semantics for geo-referenced datasets is being filled by the specification of a suite of web services from the Open Geospatial Consortium (OGC). This suite includes data access - most commonly Web Feature Service (WFS) and Web Coverage Service (WCS) and visualisation - Web Map Service (WMS). The OGC standards are understood by many of the major GIS tools, such as ArcGIS. The addition of WMS in TDS will allow previously incompatible datasets to be visualised in GIS applications. This is also a web-based protocol and compliments DAP well. WCS has already been integrated into the TDS framework [4]. Adding WMS is a logical progression of features for TDS.

Datasets are typically served through OPeNDAP using TDS with additional servers installed and configured to enable visualisation. It requires managing multiple servers and essentially doubling the amount of administration workload. The tight integration of ncWMS allows the visualisation service to be toggled like any other services in TDS. It also means only a single server has to be administrated for data access and visualisation.

Instead of implementing from scratch another WMS server, an existing server, ncWMS was chosen to integrate into TDS. The ncWMS server was developed by the Reading eScience Centre (ReSC) as part of the UK e-Science initiative to enable commonly developed meteorological and oceanographic data sets that were available in the NetCDF file type to be delivered to the geographical information systems community using internationally recognised standards, such as WMS. This application allowed the visualisation of the NetCDF data into this standard protocol, thus creating a bridging from NetCDF data types to the WMS standard.

The previous version of TDS server (3.17) already has the capacity to deliver data in WCS and OPeNDAP across the internet. The server is built on top of the core NetCDF-Java library - an implementation of Unidata's Common Data Model (CDM). CDM creates an abstraction layer over file formats and metadata convention, such that, it is possible to access data using temporal-spatial referencing systems through a single interface.

2. ncWMS

ncWMS is a visualisation server that is also using the same NetCDF-Java library. It contains an interactive web interface, Godiva2 that allows users to select and view configured layers. The majority of the integration work was to port the core WMS methods: GetCapabilities, GetMap and GetFeatureInfo. Enabling these features allows any WMS clients to add configured datasets as layers. Legends was also ported across, where each layer would dynamically generate the colour range, based on values in the file. An additional ncWMS specific method, GetMetadata, was also implemented. By adding this method, other Godiva2 instances can be configured to access layers served by TDS.

The WMS service point was kept consistent with the TDS WCS server, where each dataset is a WMS server. Another advantage of keeping the service end point on a dataset level is that a smaller GetCapabilities XML document is generated. As there is no way to "drill down" into sub layers in WMS, the GetCapabilites XML must show available layers must be shown in this document. For a server with hundreds of layers, this document can become too large to handle by servers and clients alike. Using datasets as service points means that the likelihood of overloading servers and clients is reduced, however, users are required to do more work to find layers to add to their clients.

3. INTEGRATING ncWMS INTO TDS

As TDS and ncWMS share many common libraries, integration could proceed without major changes to the code. Furthermore, the introduction of the Spring framework within TDS meant that the additional resources such as JSPs from ncWMS can be dropped into TDS with only minor changes. However, due to differences between how ncWMS and TDS loads metadata and datasets, some of the sources from ncWMS cannot be used directly in TDS. This has not been addressed in the current implementation and is causing lag between changes in ncWMS code being updated in TDS. It is envisaged that future work would involve refactoring ncWMS and TDS by using common interfaces to ensure a more streamlined process to merging ncWMS changes into TDS.

The work is now included as part of the TDS 4.0 stable release and is expected to form part of the infrastructure for the MyOcean project (http://www.myocean.eu.org). It is also used by the eMarine Information Infrastructure (eMII) to serve Integrated Marine Information System (IMOS) datasets (http://www.imos.org.au/).

4. ACKNOWLEDGEMENTS

This project is made possible through the support of the NERC Knowledge Exchange Funding Scheme, Unidata, the Australian Research Collaboration Services and Australian National Data Services (ANDS).

5. REFERENCES

1. Caron, John, Davis, E. R., Ho, Y. and Kambic, R. P. 2006, *Unidata's THREDDS Data Server*, 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography and Hydrology.

2. Nativi, Stefano, Caron, J., Davis, E., Domenico, B., November 2005, *Design and implementation of netCDF markup language (NcML) and its GML-based extension (NcML-GML)*, Computers & Geosciences, vol 31, issue 9, pp 1104-1118.

3. Gallagher, James, Potter, N., Sgouros, T., Hankin, S., Flierl, G. 10th October 2007, *The Data Access Protocol – DAP 2.0*, http://opendap.org/pdf/ESE-RFC-004v1.1.pdf, last accessed 30th June

4. Nativi, S. and Domenico, B. and Caron, J. and Davis, E. and Bigagli, L, 2006, *Extending THREDDS Middleware to Server OGC Community*, Advances in Geosciences, v8, pp 57-62.