# Merging the Data Models of NetCDF and DAP: Design Choices and Benefits

James Gallagher*, Russ Rew**, Rikki McQueary*, Dennis Heimbigner** and Ed Hartnett**
*OPeNDAP, Inc., **Unidata

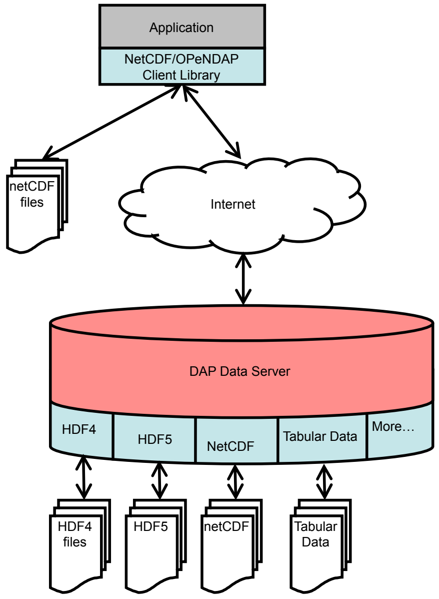
## Abstract

Beginning in 2008 OPeNDAP and Unidata have been working on an ambitious project to merge the functionality of two different implementations of the netCDF API into a single body of code. Unidata's implementation reads and writes to disk files while OPeNDAP's reads from data servers that support its Data Access Protocol (DAP). The reasons for combining the two are principally to reduce maintenance costs and delays for the introduction of new features, but a side affect has been to focus both groups on the issues of data model flexibility and simplicity. The netCDF format/API have been used in a wide range of contexts spanning the gamut of earth science disciplines including meteorology and oceanography, as well as GIS applications. The Data Access Protocol has seen a similar breadth of use. Both of these software systems employ general data structuring technology based on well-understood information science principals such as data typing and grouping. However, the actual data models of DAP 2.0 and netCDF 3 are different in some significant ways. In merging the two both will require significant changes. We will discuss the on-going process of deciding which changes should be made, where they should be made and how to implement them without breaking software that uses the existing software and data models. In addition we will discuss the exciting prospects that combining these two libraries will provide, particularly how the combination of hierarchical, relational and array data types can facilitate data fusion.

## New Capabilities for Existing Applications

✦ Access to thousands of remotely-served data sets

✦ Access to data regardless of the way they are stored



## Background

Building on a foundation laid in the inception of the Distributed Oceanographic Data System project initially funded by NASA, this project will improve the effectiveness of the NetCDF client library developed by OPeNDAP. The resulting rewritten software will be supported by Unidata and will be written entirely in C.

## Access to Thousands of Remotely Served Data Sets

Applications that can read from netCDF files can now read from not only local files, but also from hundreds to thousands of remotely served data sets. These data sets range from those served from NASA and NOAA, ones from other government agencies around the world, universities, and to non-profits. Running a DAP data server is no harder than running a web server. Several organizations in addition to OPeNDAP make DAP 2.0 compatible data servers.

## Access to data regardless of the way they are stored

Adding support for new formats and remote access protocols in an application is complex and time consuming. Using a data server transfers that complexity to the data server. An application that uses the netCDF/OPeNDAP client library can read from any data set served by a DAP server – whether it's one OPeNDAP has developed or one developed by another group. Furthermore, since the data servers are a community developed and maintained resource, the effort to support the myriad of new formats is transferred from a single developer to the community at large. The network effect is leveraged to develop support for new formats in existing applications without actually modifying the applications themselves.

## Current Status

Unidata has integrated the OPeNDAP C++ implementation of its NetCDF client library into the Unidata code base.

OPeNDAP has completed the first round of changes to the DAP.

OPeNDAP and Unidata will coordinate on the next round of changes – described at right – needed for CDM-DAP interoperability.

Updated versions of OPeNDAP's C library for clients and Unidata's netCDF library have been released.

### Bibliography

Gallagher, J., Potter, N., Sgouros, T., Hankin, S., and Flierl, G., "The Data Access Protocol: DAP 2.0," NASA ESE-RFC-004.0.08, 2005. URL: http://spg.gsfc.nasa.gov/rfc/ese-rfc-004.

Caron, John, "CDM overview (Nov 2008)", 2008 URL: http://www.unidata.ucar.edu/staff/caron/presentations/CDMoverview2008.ppt

### Contact

Use the opendap-tech@opendap.org mail list (join at www.opendap.org/mailLists)
Or, send a message to support@opendap.org.

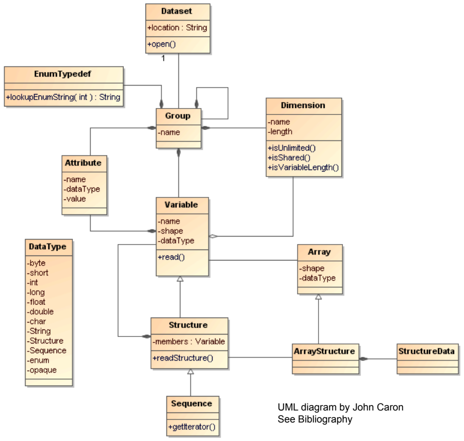
Presented at the Fall 2008 AGU meeting, San Francisco, CA.
Development of the netCDF Client Library by Unidata and OPeNDAP funded by NSF grant Number OCI-0721628.

## Interoperability between the NetCDF and DAP data models is critical

✦ In order for the netCDF library to read data from DAP servers, both the library and the data servers must represent data in a way that is compatible.

✦ While DAP was, in part, based on netCDF, important differences exist between the netCDF 3 and DAP 2 data models.

✦ Solution: Evolve the data models in concert with interoperability as an explicit goal.



UML diagram by John Caron
See Bibliography

### The Common Data Model Data Access Layer and DAP 4

Unidata has developed the Common Data Model (CDM) as a way of unifying the DAP and NetCDF data models (as well as the HDF5 data model). The above UML diagram is the CDM which is implemented by NetCDF4.

New aspects of the CDM that DAP must support are:

◆ **Data type definitions**: DAP will add support probably using the current attribute scheme it contains, with data type definitions expressed in a special global attribute container. The use of a special container for type definitions moves this solution out of the realm of a 'convention.'

◆ **Groups**: Similar to namespaces in a programming language, DAP will support this by adding the group/namespace concept as a basic component of DAP4.

◆ **Shared dimensions**: DAP will support this by relaxing the constraints on its existing 'Grid' data type.

◆ **Enumerations**: Support for this type will likely be added using DAP's attributes.