Progress Report for ACCESS Project
Software to Access HDF5 Datasets via OPeNDAP's Data Access Protocol
October 2006 – August 2007

Cooperative Agreement Number NNX06AG75A between NASA and The HDF Group
Submitted by MuQun Yang
August 23, 2007

1 Highlights

In this report we highlight three major accomplishments for the ACCESS project:

- Fully functional HDF5-OPeNDAP data handler
- Use of OPeNDAP clients to read NASA HDF-EOS5 files
- HDF5-OPeNDAP Handler website

Each of these accomplishments is covered in subsequent sections of the report. Outreach activities are listed in the final section.

2 Development and Testing of the HDF5-OPenDAP Data Handler

2.1 Selection of Programming Language for Data Handler

The prototype data handler was implemented in C++. Because several third party HDF5 and OPeNDAP Python packages have gained popularity in the Earth Science community, some time was spent at the beginning of the project to evaluate whether or not Python should be adopted as the programming language for the final implementation of the new handler. After evaluation, the decision was made to do the implementation using C++ because of performance and maintenance considerations.

The discussion of the programming language topic is posted on the OPeNDAP TWiki page (http://wiki.opendap.org/twiki/bin/view/Developers/HDF5DataHandler). The discussion provides a good reference for future Python development of the handler.

2.2 Code Cleanup

Hundreds of lines of code in the original prototype handler were updated with current HDF5 APIs. Doxygen was used to extract documentation from source file comments. These activities will benefit future software maintainers as well as users of the HDF5-OPenDAP data handler.

2.3 Features Added to the HDF5-OPeNDAP Data Handler

2.3.1 Mapping Compound Datatype to DAP Structure

The HDF5 compound datatype can support very complex data structures. It is challenging to map the full range of supported complex datatypes to DAP. The addition of this support allows the handler to serve complex HDF5 data.

2.3.2 Mapping EOS Grid to DAP Grid

The ability to map EOS Grid to DAP Grid was added to allow users to visualize and analyze Aura EOS Grid files with existing OPeNDAP clients and the data handler. Although this feature cannot be applied to general HDF5 data, it is required to allow researchers to easily access Aura data and is critical for the NASA Earth Science mission. Progress on the use of the existing clients to access EOS Grid data is reported in Section 3 of this document.

2.3.3 Partial Support of CF Conventions [1]

Most existing OPeNDAP clients either loosely or closely follow CF conventions when accessing Earth Science data through OPeNDAP. Since it is impossible for the data handler to fully support CF conventions without providing extra information, CF conventions are only partially supported in the current implementation. The current level of support ensures that clients who loosely follow CF conventions can be used to visualize and analyze HDF5 data via OPeNDAP.

2.3.4 Configuration Options

A few configuration options have been added to support mapping EOS5 Grid to DAP Grid and to partially support CF conventions.

2.3.5 Relative Object Path Support

An HDF5 object name is generally represented with its full path from the root group. However, sometimes this makes the object name difficult to read when using OPeNDAP clients to access the data. To address this issue, an option has been added that allows OPeNDAP clients to represent the object name using a relative path. With this support DAP DAS and DDS are easy to read. The application is responsible for verifying that there are no naming ambiguities when using relative paths to represent HDF5 objects.

2.4 Test Suite

A comprehensive test suite has been added using DejaGNU. In addition, the data handler source code is checked into OPeNDAP's subversion repository so that the handler can be automatically tested every night. These actions help us find and fix bugs quickly.

2.5 Testing with NASA HDF-EOS5 Files

Several on-line sample Level 2 HDF-EOS5 data files from all four instruments in Aura: OMI, HIRDLS, MLS and TES, as well as several Level 3G HDF-EOS5 sample files obtained from Aura developers were used for testing. An internal test suite was created to test these data files with the data handler.

The tests uncovered a technical problem inside OpenDAP that occurs when it tries to parse a special character reserved for a DAP expression constraint in an Aura HDF5 dataset name. We are currently in the process of investigating a solution to this problem.

With the exception of the Aura EOS data that includes the special character in the dataset name; all Aura EOS data can be accessed successfully by the HDF5-OPeNDAP handler. An internal test suite has been implemented to periodically test the Aura data.

We are confident that the current HDF5-OPeNDAP handler robustly handles the current on-line NASA Aura data.

2.6 Documentation

Three documents are in the process of being written:

- Mapping HDF5 group to OPeNDAP
- Installation guide for the HDF5-OPeNDAP Data Handler
- Investigative report on the use of OPeNDAP clients to access HDF5 data

3 Using OPeNDAP Clients to Read NASA HDF-EOS5 Files

Five OPeNDAP clients were used to access Aura EOS Grid data with the HDF5-OPeNDAP handler. They are:

- Ferret[2]
- ncBrowse[3]
- IDV[4]
- ODC[5]
- GrADS[6]

All of these OPeNDAP clients either loosely or closely follow CF conventions. However, HDF-EOS5 does not follow CF conventions. Therefore, it is technically infeasible to make the HDF5-OPeNDAP handler fully support CF conventions without providing extra information to the handler when accessing HDF-EOS5 data.

The decision was made is to partially support CF conventions for the HDF5-OPeNDAP handler. That is, the handler adds the data structure and information necessary for CF conventions only from the served HDF-EOS5 file. If some information necessary for CF conventions is missing from the EOS5 file, the handler will continue executing without them.

After the described partial support for CF conventions was added, Ferret[2] and ncBrowse[3] can be used to visualize most datasets of sample Aura L3G data. See the demo section under http://hdfdap.hdfgroup.uiuc.edu:8080/ for examples.

Various problems accessing the data using other clients have been discovered. A report detailing these findings is currently being written.

4 HDF5-OPeNDAP Handler Website

An HDF5-OPeNDAP handler website has been created. The URL of the website is: http://hdfdap.hdfgroup.uiuc.edu:8080/

The HDF5-OPeNDAP website has been used to make several important documents available on-line. In addition, several demos have been posted demonstrating how the handler accesses Aura HDF-EOS5 files and other HDF5 files. Our hope is that by making this material available we will get timely feedback from Aura developers and potential OPENDAP HDF5 users.

5 Outreach

We have communicated actively with NASA Aura developers in the process of implementing the handler. All of the sample L3G Aura files were provided by Aura developers. The developers have also provided valuable feedback on the direction of the project.

Presentations reporting the progress of this project were given at the 10th HDF and HDF-EOS workshop[7] and at the 2007 NASA HDF annual briefing. A poster on the work was also presented at the 2007 ESIP Federation summer meeting. An abstract has been submitted to The 24th Conference on International Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology that will be held during 88th American Meteorological Society annual meeting.

References

1. CF conventions: http://www.cfconventions.org/
2. Ferret /OPeNDAP: http://ferret.wrc.noaa.gov/Ferret/DODS/ferret_dods.html
3. ncBrowse – A Graphical netCDF File Browser: http://www.epic.noaa.gov/java/ncBrowse/
4. IDV – Integrated Data Viewer: http://www.unidata.ucar.edu/software/IDV/index.html
5. ODC – OPeNDAP Data Connector: http://www.opendap.org/ODC/
6. GrADS – Grid Analysis and Display System: http://grads.iges.org/grads/grads.html
7. Introduction to HDF5-OPeNDAP Project – A presentation at 10th HDF workshop: http://hdfeos.org/workshops/ws10/presentations/day2/Introduction_to_HDF5_OPeNDAP_Project.ppt