

Progress Report – Year 2

May 2009–April 2010

“Provide HDF Support for the ESDIS Project and the EOSDIS Standard Data Format”

Mike Folk and Kent Yang

May 2010

Principal Investigator: Mike Folk

Project Manager: Kent Yang

Institution: The HDF Group, 1901 So. First St., Suite C-2, Champaign, IL 61820

Period covered: May 2009–April 2010

NASA Award Number: NNX08AO77A

Highlights

User support

We received 813 requests in total; 80% of the ESDIS-related requests were addressed within two days.

DAAC and NASA HDF user community

- We updated the HDF-EOS website and added a user forum and several examples of programming and tools accessing HDF-EOS2 and HDF-EOS5 files.
- We implemented a tool to augment HDF-EOS5 files so that they can be accessed by netCDF-4.
- We maintained the HDF5-OPeNDAP data handlers.

EOS support

We continued maintaining the automatic configuration feature of HDF-EOS2 and HDF-EOS5 libraries.

HDF4 and HDF5 libraries

- The HDF Group released HDF4.2.5, HDF5 1.8.3, HDF5 1.8.4, HDF5 1.6.9 and HDF5 1.6.10. Substantial work was done in the HDF5 library to improve I/O performance with chunked datasets.
- We expanded daily regression tests of the HDF4 and HDF5 software packages. Regular snapshots of the source code distribution are provided at The HDF Group’s website.
- The updated *HDF4 User’s Guide* and *HDF4 Reference Manual* are now distributed in both HTML and PDF formats.

Applications and vendors

- We communicated frequently with NASA GES DISC developers to help them access HDF-EOS data via HDF OPeNDAP handlers and netCDF Java tools.
- We assisted Aura HIRDLS and MLS developers to help augment Aura HDF-EOS5 data.

Others

- We co-hosted HDF and HDF-EOS Workshop XIII and provided consultation to decadal mission teams.
- We enhanced the HDF4-OPeNDAP handler to access more NASA HDF4 and HDF-EOS2 data.
- We started the second phase of the project to support independent mapping of HDF4 data objects to improve the long-term preservation of HDF4 EOS data.

1 Introduction

This document has three parts:

- (1) Project goals
- (2) Progress report for May 2009 through April 2010
- (3) Related activities supported by other funding sources

2 Project goals

The primary goals of this cooperative agreement for the period between July 15, 2008 and July 14, 2011 are to provide the following:

- (1) Support activities for data providers, for data users and for the EOSDIS Core System (ECS)
- (2) Maintenance and quality assurance for HDF4 and HDF5 libraries and utilities
- (3) Evolution of the HDF5 library and utilities to meet new EOS requirements
- (4) Integration with complementary technologies and application domains
- (5) Support for the long-term preservation of HDF-EOS and the transition to NPOESS

3 Progress report for May 2009 through April 2010

This progress report is organized to correspond with the project goals listed in Section 2.

3.1 Support activities for data providers, for data users and for the EOSDIS Core System (ECS)

Helpdesk support

About 813 requests were received from May 1, 2009 through April 19, 2010. One hundred and six requests were from ESDIS-related users. About 80 percent of ESDIS-related requests were addressed within two days.

Support for NASA data centers and NASA HDF user communities

Many changes have been made to the HDF-EOS Tools and Information Center website (<http://hdfeos.org>) to help HDF and HDF-EOS users handle NASA HDF data. A new page entitled EXAMPLES was added to the website. It includes example code for accessing HDF-EOS Grid and Swath files using C, Fortran, IDL and MATLAB, as well as examples of using popular tools to access HDF-EOS and HDF files. The libraries and tools listed on the SOFTWARE page have all been reevaluated and updated where possible, and many new libraries and tools have been added. Lastly, a web forum, the FORUM page, was added to the website to enable users to share knowledge related to handling NASA HDF and HDF-EOS data. We are working on adding more sample NASA HDF and HDF-EOS files and providing corresponding example codes and screenshots for widely used tools such as NCL, MatLab and IDL. Members of The HDF Group personally interacted with NASA HDF users at earth science conferences hosted by NASA and at the AGU annual meeting. Two posters related to HDF-OPeNDAP support were given at the AGU annual meeting. We also visited the Jet Propulsion Laboratory (JPL) Earth Science division. We provided help to NASA HDF users and collected valuable suggestions for improving HDF support. For example, during the ESDSWG Conference, we provided consultations about netCDF-4 and HDF5 to decadal survey mission teams.

As requested by the Aura HIRDLS team, we implemented a tool to augment HDF-EOS5 files so that HDF-EOS5 data can be read by netCDF-4 APIs. The customized tool was used to generate new version HIRDLS data. We also provided C, Fortran and IDL example code on how one can handle multiple special values inside the HDF5 library. We also investigated accessing HDF-EOS2 data via a netCDF Java library, as requested by a GES DISC developer.

EOS support

The HDF Group continued the automatic configuration support to the HDF-EOS2 and HDF-EOS5 libraries. Developers at The HDF Group continued the automatic daily tests of HDF-EOS2 and HDF-EOS5 with HDF4 and HDF5 development snapshots.

The highest priority was assigned to the HDF libraries and tools bugs that were reported by the HDF-EOS development team; developers fixed the reported bugs in a timely manner and provided patches if necessary.

We also worked very closely with vendors, such as The MathWorks and ITTVis, whose support of HDF is important to the EOS community.

We co-hosted HDF and HDF-EOS Workshop XIII on November 3–5, 2009. We gave 12 tutorials, presentations, and posters to help ESDIS users.

3.2 Maintenance and quality assurance for HDF4 and HDF5 libraries and utilities

HDF4 and HDF5 libraries

During this period, The HDF Group released HDF5 1.8.3, HDF5 1.8.4, HDF5 1.6.9, HDF5 1.6.10, HDF 4.2.5, HDF Java 2.6, HDF Java 2.6.1 and H4toH5 2.1.1. A substantial effort was undertaken to improve the clarity and consistency of the entries in the *HDF5 Reference Manual*.

In HDF4.2r5 we included 64-bit Mac OS X Intel platform support and support for universal binaries that were requested by the ESDIS users. We also improved the structure of the HDF4 source code for easier maintenance and created more tests. HDF4 examples from the *HDF User's Guide* were added to the source and are regularly tested to assure correctness of the HDF4 documentation. To provide HDF4 users with timely access to the improvements and bug fixes in HDF4, we made periodic snapshots of the source code available at The HDF Group's website. Our updated *HDF4 User's Guide* and *HDF4 Reference Manual* are distributed now in both HTML and PDF formats.

During this period The HDF Group's developers made a substantial effort to improve the quality of the HDF5 libraries and command-line tools by using the code analysis tool "Coverity" to analyze the source code and to eliminate software problems. The HDF Group continues to expand and improve regression testing of HDF4 and HDF5. Every day more than 250 different configurations are tested on Linux, Unix, Mac OS X and Windows platforms. The configurations include testing HDF-EOS and netCDF-4 with the HDF software under development and testing HDF5 backward/forward compatibility with Aura files downloaded from NASA FTP servers to ensure the robustness and quality of HDF and its interoperability with software critical to ESDIS users.

3.3 Evolution of the HDF5 library and utilities to meet new EOS requirements

HDF4 and HDF5 libraries

We improved HDF4 source to handle files created by the customized HDF4 libraries. We provided new functions that can be used by applications such as MATLAB to detect and handle those files properly.

We assured forward and backward compatibility of the HDF5 file formats and library APIs. We also improved I/O performance for the chunking storage.

During this period, we worked on improving binary builds and distribution especially for Linux and Windows platforms. To facilitate installation of the HDF software on Linux we will distribute RPMs. To facilitate building, testing and packaging processes for the HDF software on Windows, we added Cmake to the HDF development branches. Both RPM and Cmake features will be available in the HDF5 1.8.5 release in June 2010 and in the HDF 4.2.6 release in February 2011.

Tools

The HDF Group fixed bugs and enhanced specific features that were requested by data center users for ESDIS critical tools, such as h5diff, h5repack, hdp, hdiff and hrepack.

3.4 Integration with complementary technologies and application domains

We communicated frequently with NASA GES DISC developers to help them use the HDF5-OPeNDAP handler to serve OMI, MLS and HIRDLS data. An enhanced HDF4-OPeNDAP handler implemented by The HDF Group was released in this period.

We also worked with a JPL developer to investigate ways to provide easy access to HDF data by using Web 2.0 technology. From this collaboration, a demo HDF5-Apache IO module was implemented.

We continued working closely with Unidata netCDF-4 developers. We tested the daily netCDF-4 snapshot against the HDF5 daily snapshot to ensure early identification of potential bugs.

3.5 Support for the long-term preservation of HDF-EOS and the transition to NPOESS

The HDF Group's efforts in the area of long-term preservation mostly involved a project to create maps of HDF4 data, which would enable future generations to access EOS data in HDF4 without needing the HDF4 library. The goal of this project is to mitigate the risk of EOS data inaccessibility should HDF4 software not be available in the distant future. This work was supported by extra funds under the previous cooperative agreement (NNX06AC83A) and current cooperative agreement (NNX08AO77A). This work is described in section 4.2.

During the past year, the NPOESS Integrated Program Office (IPO) began to fund HDF5 support for the NPP and NPOESS programs. This work is described in section 4.3. We anticipate this funding to continue for the coming year. As a result, we do not expect to need funds from this cooperative agreement for that work, and we plan to apply those funds to other activities of importance to EOS.

4 Related activities supported by other funding sources

Much of the work by The HDF Group during this reporting period was supported through other funding sources, as indicated in the following subsections.

4.1 HDF and OPeNDAP

Because of the success of HDF5-OPeNDAP work, NASA provided additional funding to enhance the HDF4-OPeNDAP handler so that OPeNDAP can access NASA HDF-EOS2 and HDF4 data. So far, we can successfully use the enhanced HDF4-OPeNDAP handler to visualize AIRS and some MODIS HDF-EOS2 swath and grid products. However, due to the much larger number and more complex data structures of NASA HDF-EOS2 and HDF4 data products, extra funds under the current cooperative agreement (NNX08AO77A) have been provided to support easy access to more NASA HDF-EOS2 and HDF4 products via the HDF4-OPeNDAP handler. During the past year, we greatly enhanced the HDF-EOS2 module of the HDF4-OPeNDAP handler and provided a customized version to NASA GES DISC. The customized handler can serve most HDF-EOS2 data distributed by GES DISC.

4.2 Improving long-term preservation of HDF4 EOS data by independently mapping HDF4 data objects

The NASA EOS has more than a petabyte of critical earth science data stored in the HDF4 format. This data comprises a core component of the long-term climate record, and hence it will be important to have access to

these data long into the future. There is a risk in depending solely on the HDF libraries to access HDF-formatted data over the long term. To address this risk, the HDF4 Independent Mapping Project was created to find a way to retrieve the data *independently* from HDF4 software.

In Phase 1 of the HDF4 Independent Mapping Project, we explored this concept. We assessed the range of HDF4-formatted data held by NASA and developed methods for producing XML-based maps of the layout of the HDF4 files held by NASA, including a tool for creating such maps and tools that could use those maps to read an HDF4 file without recourse to the HDF API. This work was carried out by The HDF Group, the National Snow and Ice Data Center (NSIDC) and Goddard Earth Sciences Data and Information Services Center (GES DISC).

As a result of this prototype test, a second phase for the project has begun with the purpose of creating production quality mapping tools and XML schema, and deploying these at NASA data centers. This work began in November 2009, and it is expected to be completed in July 2011. Thus far, the revision of XML schema is near completion, and work has begun on development of a tool that will create XML maps. In addition, the project has studied the need to include specific HDF-EOS data and the possibility of integrating the mapping schema with existing preservation technologies such as PREMIS and METS.

4.3 HDF5 Support for NPOESS

The purpose of this project is to provide HDF5 risk reduction support for the distribution of NPOESS VIIRS, OMPS, and other sensor and environmental data products. This important project has allowed The HDF Group to greatly improve its level of support for NPOESS, but also has freed up resources for important new ESDIS project activities.

Tasks for the current contract, which ends August 31, 2010 include maintaining NPOESS-specific software developed by The HDF Group, implementing periodic testing aimed at NPP/NPOESS needs, providing NPOESS user support, performing special maintenance projects as defined by the NPP and NPOESS communities, and performing special research projects as requested.

During the past year, library routines have been implemented in C and Fortran to greatly simplify access to region reference data and bitfield data, and HDF5 tools have been enhanced to make it possible to visualize this information. The HDF Group helpdesk also provided rapid response support to NPOESS users throughout the year. The HDF Group regularly tested HDF5 tools and libraries on systems of critical importance to NPOESS and held regular meetings with NPOESS developers and future users to help prioritize their needs and respond to their concerns.

4.4 Research on data management with HDF5 in support of military operations

Funded by a U.S. Army Engineering Research and Development Center (ERDC) Broad Area Announcement program, this investigation helped identify the role that HDF5 can play as a data management platform for mission operations and demonstrated how to adapt HDFView to present mission operational data. A paper describing the work is in the ERDC approval process. A poster was presented on this work at the HDF-EOS workshop in November 2009. The project also identified a number of research priorities for a second phase of work. Highest among these was to develop approaches to organizing and integrating spatio-temporal data in HDF5. Because this project deals with geospatial data management, it is expected to result in enhancements to the HDF5 library and tools that will be of value to the EOS community.