# Progress report – Year 1
# July 2008 – April 2009
# "Provide HDF Support for the ESDIS Project and the EOSDIS Standard Data Format"

Mike Folk and Kent Yang
May 2009

Principal Investigator: Mike Folk
Project Manager: Kent Yang

Institution: The HDF Group, 1901 So. First St., Suite C-2, Champaign IL 61820
Period covered: April 2009–April 2010
NASA Award Number: NNX08AO77A

## Highlights

*User support*

- We received 600 requests; almost all ESDIS-related requests were addressed within two weeks.

*DAAC and NASA HDF user community*

- We addressed all requests collected from the NASA HDF annual briefing and the workshop.
- We evaluated some HDF third-party open source tools widely used by the NASA HDF user community.
- We investigated accessing HDF-EOS5 data via netCDF4 APIs.

*EOS support*

- We enhanced automatic configurations of HDF-EOS2 and HDF-EOS5 libraries.
- We enhanced the HDF4-to-HDF5 conversion tool so that an HDF-EOS2 file can be converted to an HDF5 file that conforms to the netCDF-4 data model.

*HDF4 and HDF5 libraries*

- The HDF Group released HDF5 1.8.2, HDF5 1.6.8, HDF 4.2r4 and h4h5tools 2.1. We have worked very hard to make sure the backward and forward compatibility issues of HDF5 libraries are addressed.

*Applications and vendors*

- We frequently communicated with NASA GES DISC developers to help them use the HDF5-OPeNDAP handler to serve Aura OMI and TES data.
- We also investigated the I/O performance of remotely accessing the subset of an HDF5 dataset via Secure File System (SSHFS) and File system in User Space (FUSE).

*Others*

- We co-hosted HDF and HDF-EOS workshop XII.
- We enhanced the HDF4-OPeNDAP handler so that NASA HDF4 and HDF-EOS2 data can be accessed by OPeNDAP clients via the handler.
- We finished the scoping study of independently mapping 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 July 2008 through April 2009
(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 and data users and for the EOSDIS Core System (ECS)
(2) Maintenance and quality assurance for HDF4 and HDF5 libraries and utilities
(3) Evolving 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 between July 2008 and April 2009

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

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

*Helpdesk support*

About 600 requests were received between July 15, 2008 and February 25, 2009. Almost all ESDIS-related requests were addressed within two weeks. Almost 85 percent of ESDIS-related requests were addressed within two days.

*Support for NASA data centers and NASA HDF user communities*

In addition to serving NASA data centers and NASA HDF user communities through helpdesk support, members of The HDF Group personally interacted with NASA HDF users at earth science conferences hosted by NASA. For example, we attended the Federation of Earth Science Information Partners (ESIPFED) summer meeting in July 2008 and the Earth Science Data Systems Working Group (ESDSWG) conference in October 2008. At both events, we provided help to NASA HDF users and collected valuable suggestions for improving HDF support for NASA data centers and HDF user communities.

We also evaluated the usage of some HDF third-party open source tools that are widely used by the NASA HDF user community. The final report is in review and should be published at The HDF Group website soon. We also published an RFC regarding the handling of special values inside the HDF5 library. Based on feedback from NASA Aura developers, we decided to assist in the support of special values for HDF5 applications. We also investigated accessing HDF-EOS5 data via netCDF4 APIs based on the request from an Aura developer. We wrote an investigation report and sent our recommendations to corresponding developers.

*EOS support*

The HDF Group enhanced automatic configuration supports to the HDF-EOS2 and HDF-EOS5 libraries. We added more configuration options to support szip compression. This feature has been sent to the EOS developer and will be integrated into new HDF-EOS2 and HDF-EOS5 releases. Developers at The HDF Group continue the automatic daily tests of HDF-EOS2 and HDF-EOS5 with HDF4 and HDF5 development snapshots.

We enhanced the HDF4-to-HDF5 conversion tool so that an HDF-EOS2 file can be converted to an HDF5 file that conforms to the netCDF-4 data model. This enhanced tool provides an alternative way for HDF-EOS2 users to access HDF-EOS2 data via netCDF tools.

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.

The HDF Group frequently communicated with HDF-EOS developers to diagnose possible bugs in the HDF products and tools used by HDF-EOS and data centers and to diagnose HDF-EOS bugs discovered during the process of testing HDF-EOS libraries with automatic tools.

We also worked very closely with vendors, such as The MathWorks and ITTVis, whose support of HDF is important to the EOS community. For example, The HDF Group helped ITTVis test the IDL-HDF5 module with the new HDF5 release so that the new IDL release can work with the latest release of the HDF5 library. We also helped The MathWorks with HDF5 1.8 support in MATLAB.

We co-hosted HDF and HDF-EOS Workshop XII on October 15–17, 2008. We gave 12 tutorials, presentations, and demos to help ESDIS users. We also participated in two discussions and invited speakers from other organizations. We also maintained the HDF-EOS website, which was used to announce the HDF and HDF-EOS Workshop and the release of EOS SCF ToolKit.

## 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 1.8.2, HDF5 1.6.8, HDF 4.2r4 and h4h5tools 2.1, which includes several new features and bug fixes. Support of compilers crucial to EOSDIS missions, such as Visual Studio 2008 and Mac Intel, was added for both HDF4 and HDF5. Backward/forward compatibility tests were enhanced to test several Aura HDF5 files that are known to have corrupted header messages.

## 3.3 Evolving the HDF5 library and utilities to meet new EOS requirements

*HDF5 library*

We worked on several improvements in the HDF5 library — such as managing free space in the file, metadata cache journaling, and FORTRAN 2003 features — to improve accessibility of HDF5 data for Fortran applications. A bug that caused significant performance degradation was fixed in HDF5 1.6 and 1.8. The HDF Group coordinated a wiki discussion on high-level language (Python, Perl, etc.) support for HDF5. This discussion addressed a request directly from a critical ESDIS user who wanted to use Python to access HDF5 files.

*Tools*

The HDF Group fixed several bugs and enhanced specific features that were requested by data center users for ESDIS critical tools, such as h5dump, h5diff, and h5repack. HDFView 2.5 was released.

## 3.4 Integration with complementary technologies and application domains

We frequently communicated with NASA GES DISC developers to help them use the HDF5-OPeNDAP handler to serve OMI and TES data. The success of the HDF5-OPeNDAP handler led to the integration of HDF4 and OPeNDAP work. We are in the process of enhancing the current HDF4-OPeNDAP handler so that more NASA HDF4 data can be accessed by OPeNDAP clients via the enhanced HDF4-OPeNDAP handler.

We also investigated the I/O performance of remotely accessing the subset of an HDF5 dataset via Secure File System (SSHFS) and File system in User Space (FUSE). This investigation will be extremely useful for some ESDIS users to access subsets of an HDF5 dataset remotely without involving other tools.

We continued working closely with Unidata netCDF-4 developers. We provided information to Unidata Common Data Model developers on challenges and solutions to access HDF-EOS swaths and grids. We also provided a significant contribution to the netCDF-4 library so that it can read HDF-EOS5 data.

### 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 created 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 that HDF4 software may not be available in the distant future for accessing EOS data. This work was supported by extra funds under the previous cooperative agreement (NNX06AC83A). A proposal is currently in process for a second phase of this project that will productize and deploy the mapping technologies.

During the past year, the NPOESS Integrated Program Office (IPO) began to fund HDF5 support for the NPP and NPOESS programs. 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 the reporting period was supported through other funding sources, including the following:

### 4.1 HDF and OPeNDAP

The NASA ROSES ACCESS program that provided partial support to implement the HDF5-OPeNDAP handler and a prototyped DAP client library will end in August 2009 Because the HDF5-OPeNDAP handler has already been used to access Aura OMI and TES data, we will continue maintaining the HDF5-OPeNDAP handler and the prototyped DAP client library under the auspices of the cooperative agreement. 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, we are not able to support the easy access of all NASA HDF-EOS2 and HDF4 products via the HDF4-OPeNDAP handler by July 2009 Therefore, we hope to have the similar level of funding for another year to continue improving the HDF4-OPeNDAP handler so that it can be used with OPeNDAP clients to provide easy access to most NASA HDF-EOS2 and HDF4 data.

### 4.2 NOAA SDS

This one-year research project is funded by NOAA to investigate methods for storing NASA ECS data and metadata into HDF5 Archival Information Packages (AIP). To achieve this, the HDF4-to-HDF5 conversion tool has been enhanced so that converted HDF-EOS2 data can be read through the netCDF4 interface. We accumulated valuable knowledge and experience about handling HDF-EOS2 data in the course of working on this project. This experience and knowledge turned out to be extremely helpful for us in terms of providing better support to EOS users.

### 4.3 Improving long-term preservation of HDF4 EOS data by independently mapping HDF4 data objects – A scoping study

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. The normal way to access HDF-formatted data is through the HDF software libraries, either by using the HDF APIs directly or by using HDF tools that depend on the HDF libraries. However there is a risk in depending solely on the HDF libraries to access HDF-formatted data over the long term. It is possible, especially in the distant future, that the libraries may not be as readily available as they are today. To address this risk, it is desirable to have a way to retrieve the data *independently*.

This concept was explored in Phase 1 of the HDF4 Independent Mapping Project. Phase 1 included an assessment of the range of HDF4 formatted data held by NASA. Based on the results of this assessment, methods for producing a map of the layout of the HDF4 files held by NASA were prototyped using a XML-based HDF tool. The resulting maps allow a separate program to read the file without recourse to the HDF API. Two independent tools based solely on the map files were developed and tested. 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 with a cross-section of EOS files, a second phase for the project is being proposed, in which the mapping tools and XML schema are made production quality, and the mapping schema are integrated with existing XML metadata files in several data centers, with accompanying outreach activities to encourage and facilitate acceptance. In addition, now that the concept has been established for HDF4, a similar activity for HDF5 will also be proposed.

## 4.4 Hierarchical Data Format 5 Support for NPOESS

The purpose of this project is to provide HDF5 risk reduction support for the distribution of NPOESS VIIRS, OMPS, etc. 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 are as follows:

(1) Propose, develop, and implement approaches to access data pointed to by region references, rewrite region references in datasets and access individual quality flags.

(2) Determine approaches to view and export NPOESS quality flag values packed as bit fields into integers, view and export data pointed to by region references, display a rectangular region of data in spreadsheet format that is contained in a dataset corresponding to a set of region references and display bit fields packed into integers, corresponding to NPOESS quality flag values.

(3) Provide responsive and high-quality HDF5 data producer and data user support through a helpdesk to the NPOESS user community.

(4) Gather user needs and match to proposed plan, design, and design implementation. Address requirements, and inform the user communities on NPOESS HDF5 implementation and use.

## 4.5 Research on Data Management with HDF5 in Support of Military Operations

Funded by a U.S. Army Engineering Research and Development Center Broad Area Announcement program, the goals of this investigation are to identify the role that HDF5 can play as a data management platform for urban mission operations, to demonstrate the use of HDF5 visualization tools to present mission operational data, and to identify a research and development plan to develop a prototype data management system based on HDF5. This project deals a great deal with geospatial data management, and as such will generate enhancement to the HDF5 library and tools that will be of value to the EOS community.