Evaluation of Incremental Releases of ECS User Interfaces and the Development of HDF/HDF-EOS Tutorials

Final Report
Under
NASA Grant NAG5-7033

Covering the period of
February 1998 - September 2000

Prepared by

G.D. Emmitt
Principal Investigator
University of Virginia
Charlottesville, VA 22903

And

S. Greco
Simpson Weather Associates
Charlottesville, VA 22902

15 January 2001

Overview

During the reporting period, the PI has continued to serve on numerous review panels, task forces, and committees with the goal of providing input and guidance for the EOSDIS program at NASA Headquarters and NASA GSFC. In addition, the PI has worked together with personnel at Simpson Weather Associates (SWA) to help create an on-line HDF/HDF-EOS tutorial for beginning and non-expert users of both the HDF and HDF-EOS data format and software libraries. Finally, the PI has worked together with personnel at SWA and the Information Technology and Systems Center (ITSC) at the University of Alabama in Huntsville (UAH) on a feasibility study regarding the use of data mining software to ascertain features from the gridded output from numerical meteorological forecast models. A summary of these activities is provided below.

Participation on EOSDIS Panels and Committees

As a Research Associate Professor in the Department of Environmental Sciences at the University of Virginia, the PI directed and participated in a number of activities that supported the development of the EOSDIS. The PI was Chair of the EOSDIS Science Data Panel, Chair of the Goddard Space Flight Center's DAAC User Working Group, Chair of the MSFC DAAC User Working Group, and a member of several advisory panels (e.g. Science Executive Committee) to NASA on EOSDIS. His primary responsibilities included:

- Reviewing documents detailing the development and evaluation of the V0 interface to the EOSDIS.
- Planning workshops to improve the interactions between users and software designers.
- Assembling a group of scientists and other Earth science data users to test (tirekick) the user interfaces being designed for EOSDIS.
- Conducting surveys of the science data user communities to solicit input to the functionality of the user interfaces.
- Working with other researchers to design, prototype and evaluate data services, such as subsetting, coincidence searching, and data mining.

As the EOSDIS matured in its design and implementation, the PI directed his efforts to documenting "lessons learned" and serving on panels (NewDISS and ESDISSAS to provide NASA headquarters with visions for follow-on data information systems.

On-line HDF and HDF-EOS Tutorial

SWA has previously developed an on-line tutorial for beginning HDF users entitled "An HDF Tutorial for Beginners: EOSDIS Users and Small Data Providers". This tutorial was geared mainly for the beginner or non-expert HDF user as a guide to working with data using the HDF format. This tutorial, including both an on-line version and a Microsoft Word version, is available via the Internet at Simpson Weather Associates HDF page (http://cyclone.swa.com/swa/infosys/hdf/index.html).

Under the current subcontract, SWA has not only modified and updated the existing content, but has also expanded the depth of the HDF tutorial material to include additional information on data types such as raster images. However, the major component or thrust of this part of the proposed work was to combine the existing material of the HDF tutorial with new tutorial "chapters" dealing with HDF-EOS and the HDF-EOS library.

The HDF-EOS library is an extension of the base HDF library and is made up of a set of callable routines developed to deal specifically with the types of satellite and field campaign data products that are and will be routinely generated by EOS missions, but which are not fully supported by the base HDF library. Thus, to facilitate the exchange of data and data products generated by current (TERRA) and upcoming Earth Observing System (EOS) missions (such as AQUA and AURA), the Swath, Grid and Point APIs of the HDF-EOS library were developed for the EOSDIS Core System (ECS).

Similar to HDF, there are many advantages to the use of HDF-EOS. A key to it's success as a common data format and software library may be dependent upon expanding the general user and science communities' awareness, knowledge, and comfort with HDF and HDF-EOS. In particular, it is the individual investigators, academia (students through researchers), the educational community, and the general public that many times do not have the required knowledge, nor the resources to commit to obtaining this knowledge, to work with HDF-EOS data. The purpose of this tutorial is to provide the beginner and non-expert, particularly potential future users and producers of EOS data, with the necessary information to enable one to learn about both HDF and HDF-EOS and to successfully browse, read and write data sets in the HDF format.

As with the earlier tutorial and mentioned above, the tutorial chapters dealing with HDF-EOS were geared mainly for HDF/HDF-EOS non-expert or beginner users. This tutorial was not developed for experienced HDF/HDF-EOS users or members of EOS Instrument teams, DAACs, and Science Computing Facilities. Information is provided in a clear and easy to understand form, including step-by-step directions on how to work with data in HDF-EOS (and HDF). Also included in the new tutorial chapters on HDF-EOS are sections on the basics of HDF-EOS and the HDF-EOS library; programming with HDF-EOS, ways of working with data in HDF-EOS, available tools for data in HDF-EOS (including links), and many other features

As an extension to the HDF Library, the HDF-EOS library requires the linking to a successfully installed and compiled HDF library. In addition, the commands and routines are built upon the similar commands and routines within the HDF library and APIs. As a result of this intermeshing and dependence of the HDF-EOS library on HDF, the components/chapters dealing with HDF have been combined with the information dealing with HDF-EOS to form one single tutorial entitled "An HDF and HDF-EOS Tutorial for Beginners: EOSDIS Users and Small Data Providers".

This HDF/HDF-EOS tutorial is now available via the Internet and may be accessed at Simpson Weather Associates HDF page (http://cyclone.swa.com/swa/infosys/hdf/index.html). A copy of portions of this HTML on-line version is found in Appendix A. In addition to the on-line version, a Microsoft Word version is also available for download at the same address.

Data Mining Feasibility Study

One of the objectives or tasks of the proposed work that was added to the work statement was the collaboration between UVA, SWA and the ITSC-UAH on topics such as subsetting, Internet searching and data mining. After many discussions and investigations into all three areas, it was felt that enough significant work was already being done on subsetting (for example, UAH's SPOT and HEW subsetting tools) and the development of Internet search tools (too numerous to mention). As a result, it was decided to focus upon the subject of data mining.

As concluded in the October, 1999 NASA Workshop on Issues in the Application of Data Mining to Scientific Data, the overwhelming volume of data generated by modern day satellites, models and other data acquisition systems is seen as a fundamental driver for the development of data mining techniques. In the Earth Sciences, especially the atmospheric discipline, a significant component of these growing data stores is gridded output from 3-D numerical forecast models that range from climate models to operational models on the global to mesoscale.

Although there has been limited usage of problem-specific data mining techniques in such areas as cyclone detection and seasonal-interannual prediction, these efforts have not supplied a general data mining tool to work with a wide range of earth or atmospheric science data sets, especially those from model output. Thus, a joint effort or feasibility study to mine the global model output using the Algorithm Development and Mining (ADaM) system was performed between by the ITSC at UAH and SWA. The ADaM system was developed by ITSC for processing and extracting information from large volumes of scientific data sets, especially earth science data. ADaM is a flexible and powerful advanced processing system with a rich set of capabilities, such as data mining, knowledge discovery, image processing and others. This system has been successfully utilized on other meteorological case studies such as detecting cyclones from AMSU-A, Mesoscale Convective Systems (MCS) from SSM/I, lightning from OLS; rainfall modeling for SSM/I data and creating cumulus cloud masks from GOES images.

The global data used for this pilot study was taken from the ECMWF T213 Global forecast model. The output parameters from this model include surface temperature, wind speed at the surface and pre-defined layers in the atmosphere. The model output was quality checked and reformatted for ingestion into the ADAM. As an initial step, we examined the feasibility of extracting upper atmosphere jet streams, a well-known atmospheric phenomenon, using wind speed parameters. The initial effort involved creating an ADaM reader for the ECMWF model data. Then a jet detection algorithm, based on thresholding provided by the SWA, was added to the mining system. This algorithm was used to identify the locations of jet streams. The jet streams at 7 levels between 500mb and 200mb are detected using the algorithm. The results can be seen in Fig 1B. The simple thresholding algorithm also picked out some isolated "spots" (noise) along with the detected jet stream features. To improve the quality of output image, the built-in morphological erosion and dilation functions were used to filter out these "spots". These image-processing operations improved the outcome of the detected jet streams. This initial effort demonstrated ADaM's flexibility in incorporating new algorithms for phenomena discovery.

In addition, a K-means Unsupervised Classifier, a well-known clustering algorithm was applied to the maximum wind speed data in order to identify the jets. The results from the K-means Classifier are shown in Fig. 1C and Fig. 1D for 3-clusters and 4-clusters, respectively. For 3 clusters, the mining system correctly groups data into jet, other (non-jet), and noise. The noise can be seen as dark specks northeast of India. The jet classified in Fig. 1C is fairly broad and the classification can be improved by adding one more cluster. In Fig. 1D, the classifier correctly extracts Jets as the fourth distinct group. This initial study shows that data mining technology can be a great help to atmospheric scientists in identifying and automatically locating meteorological features. Although the jet stream can be properly characterized by just the atmospheric wind parameters, other atmospheric features such as front may be more complex and the characterization may be quite non-linear. The vast set of data mining and image processing algorithms available in ADaM's toolkit allows ADaM to be an excellent tool that can be utilized for such scientific research.

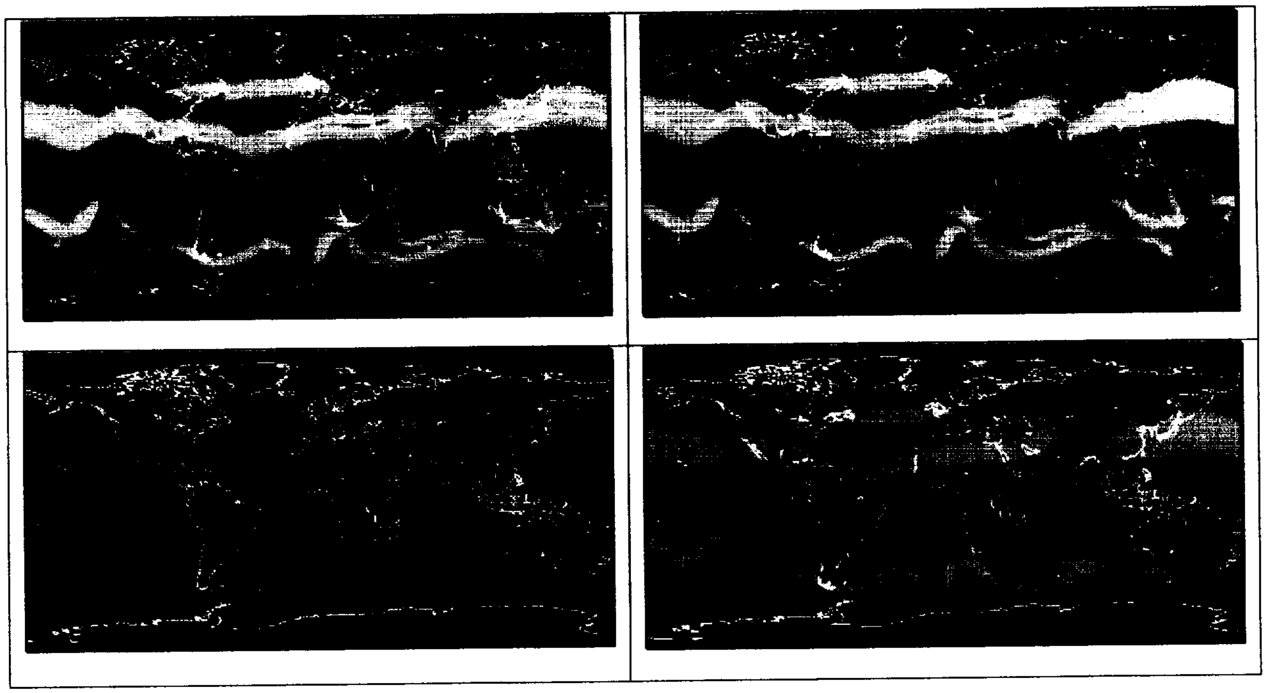

Figure 1: (A) Maximum Wind Speed Data (B) Applying Science Algorithm (C) Unsupervised Classification – 3 Clusters (D) Unsupervised Classification – 4 Clusters

APPENDIX A

An HDF and HDF-EOS Tutorial for Beginners: EOSDIS Users and Small Data Providers (On-Line HTML Version)

by

Mr. Steven Greco
Simpson Weather Associates
Charlottesville, VA

(In Fulfillment of NASA Contract NAG5-7033)

January 2001

An HDF and HDF-EOS Tutorial for Beginners: EOSDIS Users and Small Data Providers

Begin Tutorial

Tutorial Overview

The NASA ESDIS project selected the Hierarchical Data Format (HDF) as the common data format of choice for standard product exchange and distribution. As developed by the National Center for Supercomputer Applications (NCSA), the HDF format is supported by a collection of software routines and applications needed to work with data sets in HDF. This set of software, referred to as the HDF library, is available in the public domain. To facilitate the exchange of data and data products generated by current (TERRA) and upcoming Earth Observing System (EOS) missions (such as AQUA and AURA), as well as satellite missions such as TRMM, a sub-library or library extension of HDF, called HDF-EOS, has been developed for the EOSDIS Core System (ECS) by the Raytheon Systems Corporation (RSC). HDF-EOS was developed to specifically deal with the types of satellite and field campaign data products that are and will be routinely generated by EOS missions but which are not fully supported by the base HDF library.

While there are many advantages to the use of HDF and HDF-EOS, a key to their success as a common data format and software library may be dependent upon expanding the general user and science communities' awareness, knowledge, and comfort with HDF and HDF-EOS. In particular, it is the individual investigators, academia (students through researchers), the educational community, and the general public that many times do not have the required knowledge, nor the resources to commit to obtaining this knowledge, to work with HDF and HDF-EOS files.

In response to this need, the NASA ESDIS project has funded the creation of this on-line tutorial geared towards HDF and HDF-EOS beginners. The purpose of this tutorial is to provide the non-expert, particularly potential future users and producers of EOS data, with the necessary information to enable one to learn about HDF/HDF-EOS and to successfully browse, read and write data sets in HDF format. This information will include, but not be limited to, sections on the basics of HDF/HDF-EOS and HDF/HDF-EOS files, the required software/hardware, the various ways of working with the files, a review of HDF and HDF-EOS commands and operations, and instructions and advice for writing programs to work with HDF and HDF-EOS.

A portion of the information presented here can also be found in much further detail scattered throughout various other literature and several of the excellent reference guides and manuals (more on this in Section 2- An Introduction to HDF and Section 3- An Introduction to HDF-EOS) written by NCSA for HDF and by RSC for HDF-EOS. However, the goal of this tutorial is to present, in a concise and easy to understand form, only the information needed to help the HDF novice to browse, read and write basic HDF/HDF-EOS files. More importantly, as they are certainly linked, it is the goal of this tutorial to combine the information on both HDF and HDF-EOS under one encompassing document. Furthermore, the HDF library has been designed to work with many different types of data (arrays, images, etc.) and to carry out both simple and complex operations on data sets. As a teaching tool, this tutorial will concentrate on only several selected data types supported by HDF (images, palettes, scientific data arrays) and HDF-EOS (Point, grid, and swath) libraries, and the basic operations such as reading, writing, and browsing entire data sets. The user will be directed to the some of the most critical documentation to learn about all the various operations and data types supported by HDF and HDF-EOS.

Other excellent HDF tutorials have been developed by NCSA and GSFC. The current tutorial contains much more of the basic information of HDF and is geared to the HDF beginner or novice. Though there are many places of overlap between the tutorials, they seem to compliment each other in providing information for all types of HDF users.

Return to top

Versions of HDF and HDF-EOS

The current version of this HDF and HDF-EOS tutorial concentrates on the latest release of the HDF library (HDF 4.1r3 as of July 2000) and HDF-EOS sub-library (HDF-EOS 2.6 as of June 2000), and how to use them for browsing, reading and writing HDF files containing scientific data arrays, raster images with associated palettes; and point, swath, and gridded HDF-EOS data sets.

In addition to HDF4, a second version of HDF, called HDF5, has also been developed. This new library was designed to address the main drawbacks of HDF4, particularly the inability to deal with large data sets. Both the HDF4 and HDF5 Libraries are currently being used by data producers and being supported by NCSA, although all future development of the HDF library will be on HDF5.

As of the summer of 2000, investigators and science teams were making individual decisions regarding which HDF format/library (4 or 5) to use for their data. Data providers and instrument teams from the current TERRA platform (MODIS, MISR, etc..) and the upcoming AQUA platform (2001) are using the HDF4 library. Partly for this reason, as well as the fact that there has been little commercial tool development for HDF5, this tutorial will be focused upon HDF4. To learn more about HDF5, the readers are directed to the NCSA HDF5 page and the outstanding HDF5 tutorial that has been developed by NCSA.

With the creation of a new HDF5 Library, there has simarlarly been the development of a new HDF-EOS sub-library which is totally based on HDF5. This is the new HDF-EOS 3.0 library, which has entirely new functionality, but it has yet to be offically accepted by ESDIS. This tutorial will focus on the contents and supporting software for the original HDF-EOS 2.6 library (based on HDF4).It should be noted that NASA and HDF-EOS have full intention of indefinitely supporting HDF4 and will expand significant effort on facilitating the interoperability and conversion between HDF4 and HDF5. This will be touched upon later on in the tutorial

Return to top

How to use the tutorial

In support of a contract to the NASA ESDIS project, this HDF tutorial has been designed by Simpson

Weather Associates, Inc. (SWA) with the goal of teaching the novice HDF and HDF-EOS users, especially potential users of EOSDIS and both current and future EOS data products, how to use the libraries to read and write HDF or HDF-EOS files. The tutorial has been constructed in two parts. The main part of the tutorial is what we call the "Lecture" component where we present what we think is the information necessary for a novice user to learn what HDF and HDF-EOS is, what it can be used for, and how to apply it in practice. Included in this "Lecture" material is a step-by-step outline detailing what the user must do to successfully read or write an HDF/HDF-EOS file.

This "lecture" component makes up most of the tutorial. The second and much smaller component of the tutorial is a question and answer section (what we call the "Laboratory") which tests the user's knowledge of HDF, concentrating on the information needed by the novice or beginner user to work independently with the HDF library to read and write files.

We realize that the familiarity and knowledge level of the users of this tutorial will span a wide range. As a result, we think it should be up to the users to decide how they wish to learn and navigate through the tutorial. However, we do advise that those with very little or no knowledge of HDF should first preview and study the lecture material before testing themselves with the Laboratory section.

Return to top

Future plans for the tutorial

Modifications and updates to the HDF4 and HDF-EOS2 libraries, as well as user feedback, may require changes to the tutorial content. As an example, new incremental versions of HDF (HDF4.1r4) and HDF-EOS (HDF-EOS 2.7) were released in December of 2000, after the completion of the contract. The tutorial will eventually be updated to indicate any neceesary changes from these versions of HDF and HDF-EOS. Although NCSA's HDF5 tutorial is very thorough, it is also hoped that, in the future, someone will take on the task of creating a tutorial specifically aimed at beginner users that covers the new HDF5 and HDF-EOS 3.0 libraries.

Return to top

An Introduction to HDF

Previous Main Topic Next Main Topic

Return to Main Topics

What is HDF?

HDF, which stands for Hierarchical Data Format, is a common data format that has been developed to aid scientists and programmers in the storing, transfer and distribution of data sets and products created on various machines and with different software. HDF has been selected by the NASA ESDIS project as the format of choice for the standard product distribution that will be part of the Earth Observing System Data and Informations System (EOSDIS).

In addition, HDF also refers to the collection of software, application interfaces, and utilities that comprise the HDF library and allows users to work with HDF files. The HDF library is discussed in detail in Section 3 - The HDF Library: Software and Hardware.

Features of HDF

HDF is a multi-object file format for the sharing and storing of scientific data. Some of the most important features of HDF are the following:

1. Self-describing: For each data object in an HDF file, there is also information (or metadata) about the data type, size, dimensions and location found within the file itself.
2. Extensibility: HDF is designed to accommodate future (new) data types and data models.
3. Versatility: Currently, HDF supports six different data types and provides software and applications to read and write these data types in HDF.
4. Flexibility: HDF lets the user group, store, and read/write different data types in the same file or in more than one file.
5. Portability: HDF software is mainly platform independent and can be shared across most computer platforms (all platforms have not been tested).
6. Standardization: HDF standardizes the formats and descriptions of many types of commonly-used data types (i.e., arrays, images, etc.).

7. HDF is available in the public domain.

Return to top

What types of data does HDF support?

As of the latest release of HDF (HDF 4.1r3 release as of July 2000), the HDF library supports the working with raster images, color or gray scale palettes, multi-dimensional arrays, text strings, and statistical data (in the form of tables). The HDF library supports the following data types:

1. Scientific Data sets -- Multi-dimensional integer or floating point arrays
2. Vertex Data (Vdata and Vgroups) -- Multi-variate data stored as records in a table
3. General Raster (Gr) -- Raster images
4. Annotation -- Text strings to describe files and parts of files (metadata)
5. 8-bit Raster images
6. 24-bit Raster images
7. Palette -- 8-bit color palettes (accompany images)

In addition to these data types supported by the base HDF library, a sub-library called HDF-EOS has been developed to support various data types from the Earth Observing System (EOS) and other satellite missions. The HDF-EOS data models include point data, satellite swath data, and gridded data. HDF-EOS files are already being routinley generated by the instruments aboard the EOS TERA platform, as well as the TRMM satellite. More information on HDF-EOS and the differences between HDF and HDF-EOS will be provided in following sections.

As mentioned in the Welcome section, this HDF component of the tutorial will concentrate on the Scientific Data and raster image Data Models as a means of teaching the essentials of HDF. More information on the other data models can be obtained in the various documents (particularly the HDF User's Guide) provided by NCSA through their anonymous ftp server or World Wide Web home page.

Return to top

Which version of HDF should I use?

The most current version or release of HDF is the best place to begin. The current version of the HDF library as of summer 2000 is HDF 4.1r3, but with a new release slated for December 2000. An extension of the HDF library, called HDF-EOS (HDF-EOS2.6), is based on this version of HDF and is designed specifically to work with data from EOS satellite missions. The current tutorial will focus on the releases (i.e., r1, r2 or r3) of HDF4.1. One feature of HDF4 that is important, especially to experienced users of HDF, is the backwards compatibility of HDF. That is, HDF4.1r3 is compatible with earleir versions such as HDF4.1r1 and the data sets that were generated.

It should be noted that a second version of HDF, called HDF5, has also recently been developed to address the shortcomings of HDF4. This new HDF library includes simpler source codes, more consistent and fewer data models, and the ability to work with large data sets (> 2GB). Although HDF5 and associated software will not be covered in this tutorial, we want to say a few words about the transition and compatability between HDF4 and HDF5. For complete information on HDF5, the user is directed to NCSA's HDF5 Page.

Return to top

HDF4 vs HDF5

Although HDF4 is the basic underpinning of HDF-EOS and will continue to be supported by both NCSA and HDF-EOS, it is the new HDF5 Library that is slowly emerging as the new standard and is the HDF library that will be developed in the future. However, the transition will take many years as investigators and science teams are making individual decisions regarding which HDF format/library (4 or 5) to use for their data. Data providers and instrument teams from the current TERRA platform (MODIS, MISR, etc..) and the upcoming AQUA platform (2001) are using the HDF4 library.

Due to the existence of both HDF libraries and the anticipation of significant scientific data sets being create in both HDF4 and HDF5, a major task facing NASA and NCSA is to facilitate the interoperability and conversion between HDF4 and HDF5 data sets. Source code, documentation, and software being (or has been) written for this purpose. The reader is directed to a NCSA HDF4 to HDF5 White Paper for further information.

Where can I get additional and detailed information on HDF?

The best site or location to find detailed information on all aspects of HDF is the NCSA HDF Information Server available through the Internet. Another good place to start is the HDF FAQ (Frequently Asked Questions) document. For complete documentation on HDF, the user is directed to the NCSA HDF Documentation page (most current) and the NCSA anonymous ftp server . Inquiries and further questions should be sent to hdfhelp@ncsa.uiuc.edu.

The following documents and information can be obtained through the sources mentioned above:

1. The most current HDF 4.1 Reference Manual (4.1r3 as of Summer 2000)
2. The most current HDF 4.1 Users Guide (4.1r3 as of Summer 2000)
3. HDF Specifications and Developers Guide v3.2 (mainly for the programmers/developers)
4. HDF Newsletters
5. HDF Frequently Asked Questions (FAQ)
6. Java Products
7. Frequently Asked Questions about Java and HDF
8. Release Notes and Man Pages provide information on items that are not covered in the above documents
9. HDF software contributions from non-NCSA users

In addition, users may wish to join the hdfnews mailing list (by emailing ncsalist@ncsa.uiuc.edu and placing subscribe hdfnews in the body of the message) for discussions and updates on HDF.

Return to top

An Introduction to HDF-EOS

What is HDF-EOS?

What types of data does HDF-EOS support?

Which version of HDF and HDF-EOS should I use?

Where can I get additional and detailed information on HDF-EOS?

Where can I get HDF-EOS data?

Previous Main Topic Next Main Topic

Return to Main Topics

What is HDF-EOS?

As mentioned in Section 2 - An Introduction to HDF, HDF as developed by NCSA is the common data format selected by NASA and ESDIS as the format standard for data exchange of EOS data. The HDF software library provides the application routines and code to work with (create, read, write, etc...) various data types in HDF.

As it stands, however, many of the data products and data types that result from the EOS missions (current and future) are not fully supported by the base HDF library nor map directly to NCSA/HDF data types. These data types mostly include geolocation data types, or data sets with associated geolocation or temporal information. Examples of these types of data sets include point data sets from field observation studies, data sets from the swaths of satellite borne sensors, and data sets on gridded map projections. These type of measurements or data, called Point, Swath, and Grid data sets, are major components and products of the EOS missions, which began in late 1999 with the launch of TERRA, and will continue with the future launches of AQUA (2001) and AURA. As a result, the HDF-EOS (Hierarchical Data Format-Earth Observing System) software library was created for the EOSDIS Core System (ECS) by the Raytheon Systems Company (RSC) to fully and optimally support these types of data.

The HDF-EOS library is not a stand alone software package, rather it is an extension of the base HDF library that requires the installation of the base HDF library and the use of its' basic function calls as an underlying basis to successfully work with HDF-EOS data types like Point, Swath, and Grid. Therefore, in all actuality, HDF-EOS data sets or objects are basically HDF objects with the added feature of being able to support geolocation information. This geolocation information is included in the HDF files through the automatic inclusion of structual metadata.

In particular, HDF-EOS allows for the subsetting of point, swath, and gridded data sets in terms of time and geolocation (Latitude, longitude, altitude) and not just in rows and columns (i and j) like in HDF. One other feature that specifically defines an HDF-EOS data set is the mandatory inclusion of ECS Core metadata. This is not required in ordinary HDF files, although it may be used as an option. It should also be noted that an HDF file may include both HDF AND HDF-EOS data objects and that HDF-EOS data sets can indeed be read by HDF tools and routines but with the geolocation information not accessible.

The HDF-EOS format shares many of the same important features of HDF (i.e., self-description) as mentioned in Section 2 - An Introduction to HDF. In addition, HDF-EOS provides the following:

- Geolocation information coupled to data
- Standardized geographical subsetting
- ECS metadata

Return to top

What types of data does HDF-EOS support?

As of Summer 2000, the current version of the HDF-EOS library (HDF-EOS 2.6) supports three types of geolocation data types:

1. **Point Data Types** - Data, such as ship observation reports, that is irregularly spaced in time and/or space.
2. **Swath Data Types** - Time-ordered satellite data which represents time sequences of scan lines, profiles, or other array data.
3. **Grid Data Types** - Data that has been stored or can be represented on a regular grid and is based on certain set eath/map projection (i.e., Mercator, Lambert Conformal, etc..).

Similar to HDF, each data type is supported by a collection of software called an Application Programming Interface (API) that allows the user to work with the data, particularly in reference to both time and space. The Swath, Point, and Grid APIs will be discussed later in the tutorial.

Return to top

Which version of HDF and HDF-EOS should I use?

The most current version or release of HDF and HDF-EOS is the best place to begin. As of July 2000, the current version of the HDF library is HDF 4.1r3. The current version of the HDF-EOS library, HDF-EOS2.6, is built upon and based on this version of HDF (HDF4.1).

As mentioned previously, a new version of HDF, called HDF5, has also recently been developed to address the shortcomings of HDF4. This new HDF library, which will eventually become the new standard over time, includes simpler source codes, more consistent and fewer data models, and the ability to work with large data sets (> 2GB). A new version of the HDF-EOS Library, called HDF-EOS3.0, has been developed which is totally based on the new HDF5 Library and will have an entirely different functionality. The HDF-EOS3.0 Library will will soon officially become "avaialble".

Data producers and instrument teams for the TERRA and AQUA sensors are using the HDF4 Library (and, thus, HDF-EOS 2.6 if HDF-EOS is used) to create many data sets. Users of this data should continue to use the HDF4/HDF-EOS2.6 libraries. However, as time evolves, data sets created by or together with the HDF5 Library will employ routines from HDF-EOS3.0. Tool developers and data providers have been working with all the various HDF and HDF-EOS libraries and have found that, due to naming convention conflicts, both the HDF-EOS2.6 and HDF-EOS3,0 libraries can not be opened and accessed at the same time. Only one can be used.

One of the bigger differences between HDF-EOS2.6 and HDF-EOS3.0 is that with HDF-EOS2.6 the user must access the structual metadata to map geolocation data with the scientific data while this dependency has been been renoved in HDF-EOS3.0 as the structual Metadata/geolocation information will be contained within HDF5 objects.

Return to top

Where can I get additional and detailed information on HDF-EOS?

There are several locations where the user may find detailed documentation and information on HDF-EOS. One of the most helpful sites for HDF-EOS documentation is the ECS Data Handling System (EDHS). Doing a quick search for "HDF-EOS" at this location, the user will find links to download many HDF-EOS documets, including the following:

- HDF-EOS 2.6 Version Description Document (from 3/2000)
- HDF-EOS Users Guide, Volume 1: Overview and Examples (from 2/2000)
- HDF-EOS Users Guide, Volume 2: Reference Guide (from 2/2000)
- Thoughts on HDF-EOS Metadata
- Writing HDF-EOS Grid Products for Optimum Subsetting Services
- Writing HDF-EOS Point Products for Optimum Subsetting Services
- Writing HDF-EOS Swath Products for Optimum Subsetting Services
- An HDF-EOS Primer
- The HDF-EOS Swath Concept
- Release 5B SDP Toolkit Users Guide (from 4/2000)
- EOSView Version 3.1 Users Guide for the ECS Project (from 1/2000)
- HDF-EOS Interface Based on HDF5, Volume 1: Overview and Examples
- HDF-EOS Interface Based on HDF5, Volume : Function Reference Guide

Another informative site that contains detailed HDF-EOS documentation is the HDF-EOS Standards and Tools Information Center home page. The HDF-EOS Information Resources page at this site also provides links to the latest documentation on HDF-EOS including:

- HDF-EOS Version Description Document
- HDF-EOS Users Guide, Volume 1: Overview and Examples
- HDF-EOS Users Guide, Volume 2: Reference Guide

In addition to containing HDF-EOS documentation, this site also provides links to or information on HDF-EOS source code, how to order an HDF-EOS data sampler, and the latest version of the HDF-EOS FAQ (Frequently Asked Questions). This site also provides links to documentation on the HDF Configuration Record (HCR).

The NCSA HCR Page provides complete documentation on HCR. The HCR is a free set of software tools created by NCSA that provides an interface for the user to create, read or write HDF (including HDF-EOS) files and is a formal description of the HDF-EOS data. The HCR will be discussed in later sections.

Return to top

Where can I get HDF-EOS data?

An HDF-EOS Data Sampler CD can be ordered from the HDF-EOS Standards and Tools Information Center web site. In addition, some of the HDF-EOS data contained on the CD sampler may also be viewed or downloaded from the DIAL server at NASA Goddard Space FLight Center.

With the launch of TERRA in late 1999, as well as the data being generated by Landsat-7, earth science data in HDF/HDF-EOS format is being archived at the various Distributes Active Archive Centers (DAACs). The user may wish to check the EROS Data Center (EDC) for Landsat-7, ASTER, and MODIS-Land data; the Langley Atmospheric Sciences Data Center for MISR and MOPITT data products; the National Snow and Ice Data Center (NSIDC) for MODIS ice-related products; and the Goddard Space Flight Center Earth Science (GES) Data Center for MODIS atmospheric and ocean data products. Sample MODIS products may also be obtained from GSFC. The above-mentioned DAACs can also provide the user with HDF and HDF-EOS reader software to work with the data.

As mentioned earlier, many of the files containing HDF-EOS objects such as swaths and grids have been placed together with HDF objects in "hybrid" HDF files using the HDF library rather then using the HDF-EOS library to create pure HDF-EOS files with only HDF-EOS objects.

Return to top