HDF-EOS, NASA's Standard Data Product Distribution Format for the Earth Observing System Data Information System

Richard E. Ullman,
NASA/Goddard Space Flight Center
Code 423
Greenbelt, MD 20771 USA
tel: +1.301.614.5228/ fax: +1.301.614.5270/ email: ullman@gsfc.nasa.gov

INTRODUCTION

HDF-EOS is the official data format standard for The Earth Observing System Data Information System (EOSDIS) data production, archive and distribution. The adoption of HDF-EOS as the data format standard will encourage and facilitate data and software exchange and interoperability of EOSDIS data. The scientific goal of EOS is to study our global environment and obtain sustained and continuous observations of the Earth's atmosphere, oceans, land surfaces, ice coverage, and land uses. The goal of EOSDIS is to enable wide use of the collected data. EOSDIS will receive and store over one terabyte of new data each day. EOSDIS stores and retrieves these data as "product collections" of "granules". These terms are roughly equivalent to "data sets" and "files" respectively. The users of these products will number in the thousands and will include scientists from many disciplines, educators, policy makers, commercial organizations and the public.

HDF

The Hierarchical Data Format (HDF), developed by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign, is a data format designed to facilitate sharing of scientific data. Features include platform independence, user extensibility, and embedded metadata for units, labels, and other descriptors. Standard data types include multidimensional array, text, table, raster image, and palette. Both Fortran and C functions access and process the data stored in HDF objects. Machines currently supported include the Cray, Convex, HP, DEC, Sun, IBM RS/6000, Silicon Graphics, Macintosh, and WINTEL PCs.

NCSA has recently released HDF5. HDF5 is substantially different from HDF 4, but several improvements in the format are critical to the future evolution of the EOSDIS standard. Among them are the ability to store larger files and more objects per file and improved support for parallel I/O, threads, and other requirements imposed by modern computing systems. HDF-EOS will eventually migrate to HDF5 and keep current with capability requirements.

HDF-EOS

HDF-EOS is a use of HDF developed by the EOSDIS Core System Project for NASA to enable EOS specific conventions, data types, and metadata for EOS data objects. An essential characteristic of all HDF-EOS granules distributed by EOSDIS is that they contain the inventory metadata needed for EOSDIS search services. There are also three defined HDF-EOS data objects: *Point*, *Swath*, and *Grid*. These three objects permit the data contents to be queried by earth coordinates and time. The *Point* interface supports data with associated geolocation information not organized in any well-defined spatial or temporal way. The *Swath* interface supports time-ordered data such as satellite scanner data that consist of a time-ordered series of scan lines, or satellite sounder data that consist of a time-ordered series of profiles. The *Grid* interface supports data that has been stored in a rectilinear array based on a well defined and explicitly supported projection.

HDF-EOS METADATA

EOSDIS standard metadata about the data products is stored in a "text annotation" object within the HDF-EOS granule. The metadata representation is Object Description Language (ODL) according to the semantic model defined by the EOSDIS data model. EOSDIS metadata is compliant with Federal Geographic Data Committee (FGDC) and Global Change Master Directory (GCMD) metadata standards. These standards facilitate exchange of data among disparate segments of the Earth Science and Geographic Science community. The metadata in an HDF-EOS granule serves several purposes. It stores information about the product author, production site and time, data geographic and temporal extent as well as indication of the quality of the granule as assessed by the responsible science team. Standard HDF calls cannot conveniently access the EOSDIS metadata parameters within the metadata object. However, a suite of specially developed HDF-EOS tools is available that can dump the ODL metadata annotation object into a text file and process the result into HTML for human readable display.

In addition to the EOSDIS metadata, each HDF-EOS granule contains "structural metadata" that are used to tie the individual HDF objects together into the more complex data

structures used in the HDF-EOS *Point*, *Swath*, and *Grid*. The details of HDF structural metadata are beyond the scope of this paper.

HDF-EOS POINT STRUCTURE

The HDF-EOS *Point* structure is for storage of like data that are discrete points in time and/or location. The *Point* design enables data subsetting by geographic location, time and altitude as appropriate. This subsetting uses standard names for the representation of these parameters. There are two major parts of *Point*.

First, a set of field labels: If geographic location information is included in the *Point* table, it must be using a field labeled "Latitude" or "Colatitude" and "Longitude". Time location information resides in the field labeled "Time". If altitude location is provided, it is in the parameter field labeled "altitude_units", where "units" is one of many common units of altitude or depth. Other field labels correspond to product metadata parameter values.

Second, hierarchical data tables store the data: *Point* supports up to eight levels of tables. Each successive level is an index into the level below. Table fields use the defined field labels and may be multivalued. The number of values defined for each field is the "order" of the field. All elements of the same field have the same C language data type.

HDF-EOS SWATH STRUCTURE

The design of HDF-EOS *Swath* accommodates classic satellite data collection patterns of scanning [Fig. 2] and sounding [Fig. 3]. An essential characteristic of *Swath* data is a regular relationship between the data observations and progression of time and location within the granule. *Swath* is useful for storage of geolocated radiances, typical of a Level 1 data set, and also for derived science parameters grouped along a satellite track, typical of Level 2 data sets. There are four primary parts to *Swath* and one optional index.

First, a set of named dimensions for each axis in the geolocation and data field elements defined below. The names of the dimensions map the relationship between the data observations and their geolocation. In EOS products designed for parameter subsetting, the data dimension names are the same as the named parameters listed in the EOSDIS metadata for the product collection. The standard geolocation dimension names are "Time", "Latitude", "Colatitude" and "Longitude".

Second, a set of one- or two-dimensional geolocation fields: These fields provide the set of geographic or time locations associated with the data field dimensions.

Third, a set of dimension maps: If a data field shares a dimension name with a geolocation field, no mapping is required. However, if the data field dimension name is

Fig. 1: Scenario for use of HDF-EOS POINT Structure

Fig. 2: Scenario for use of a Scanning SWATH Structure

Fig. 3: Scenario for use of a Sounding SWATH Structure

different from the geolocation field dimension name, an offset and an increment link each related pair. The n^{th} element of a geolocation field maps to the $(n \times \text{increment} + \text{offset})$ element of the data field.

Fourth, a group of data fields: A *Grid* object may contain many data fields, each with up to eight named dimensions and each sharing the same geolocation field dimensions. All elements in a single data field are of the same C language data type. Most often, the data fields are two or three-dimensional arrays.

Optionally, an index element: The *Swath* can be divided into predefined and possibly unequal "scenes" as with Landsat 7 products.

HDF-EOS GRID STRUCTURE

The design of HDF-EOS *Grid* accommodates geolocated data binned into a standard reference projection as is common with Level 3 and 4 data products [Fig. 4]. There are eleven named projections native to *Grid*. When EOS data providers require additional projections, HDF-EOS will be extended to support them also. There are three essential parts to *Grid*.

First, the projection, and its name, size and projection specific parameters: Defined projections are taken from the USGS General Coordinate Transformation Package (GCTP). GCTP is extended to include some additional projections such as the Intergerized Sinusoidal Projection from SeaWiFS. Projection specific parameters vary by projection. All projections result in a fixed number of divisions in the X and Y dimensions.

Second, a set of Grid parameter dimension names: Each name corresponds to an EOSDIS metadata defined parameter name. The two dimensions of the projection are "Xdim" and "Ydim". All other dimensions, such as height or parameter space take their names from the EOSDIS metadata for the product collection

Third, the set of data fields: A data field is an array of two or more dimensions. A *Grid* may have multiple data fields. All elements within a field have the same underlying C language data type. The same geolocation and projection relate each field within a single *Grid*. The names of the data fields enable subsetting when their names are from the EOSDIS metadata for the granule and collection.

TOOLS FOR READING HDF-EOS DATA

Any tool that processes standard HDF objects can read an HDF-EOS granule, although the semantic linkage between HDF objects that imparts the geolocation and temporal characteristics of the data is not apparent using standard HDF calls. NASA is actively encouraging the adoption of HDF-

Fig. 4: Scenario for use of a GRID Structure

EOS to facilitate data exchange among the many interrelated Earth science disciplines. These efforts are increasingly successful as the science community and software vendors come to understand the standard better. The HDF-EOS code is available for inclusion by vendors into tools and tool suites to support data analysis and visualization. A yearly workshop for vendors and users convenes to share experiences. Already, several free and commercial software packages support the HDF-EOS model including the geolocation and temporal attributes. Free software includes EOSview and DIAL from NASA's ESDIS Project, Fortner Software's HDF Browser and JPL's Webwinds. Commercial products include Fortner Software's Noesys now packaged with Research Systems' IDL, and The MathWorks Inc. Matlab and others.

BIBLIOGRAPHY

- Campbell, Janet, John Blaisdell and Michael Darzi, Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms, NASA TM 104566, GSFC Volume 32, Appendix A, January 1995.
- Carlisle, Candace and George Schwenke, HDF-EOS Workshop Home Page, <http://hdfeos.gsfc.nasa.gov>, ESDIS Project, NASA GSFC, September 1998.
- Ilg, Doug, David Wynne, Raj Gejjagaraguppe and Larry Klein, HDF-EOS Library User's Guide for the ECS Project, EOSDIS Core System Project, Landover, MD Technical Paper 170-TP-100, January 1999.
- Jet Propulsion Laboratory, Planetary Data System Standards Reference, Ver 3.2, Chap 12: Object Description Language (ODL) Specification and Usage, Pasadena, CA, July 1995. <http://pds.jpl.nasa.gov/stdref/stdref.htm>
- NCSA, NCSA HDF User's Guide, Version 4.1r2, University of Illinois at Urbana-Champaign, June 1998.
- NCSA, The NCSA HDF Home Page, <http://hdf.ncsa.uiuc.edu/>, University of Illinois at Urbana-Champaign, January 1999.