



by Laurie J. Schmidt  
December 22, 2000

On December 18, 1999, NASA and its international partners launched Terra, the first of the Earth Observing System (EOS) satellites planned for NASA's Earth Science Enterprise (ESE) program. Terra, along with future EOS satellites Aqua and Aura, will deliver more than two terabytes ( $2 \times 10^{12}$  bytes) of atmospheric, ocean, and land surface data each day.

The EOS Data and Information System (EOSDIS) distributes ESE data through the Distributed Active Archive Centers (DAACs), the institutions responsible for archiving and making data products readily available to anyone who wants them. Making these collections accessible is key to achieving EOS program goals. But the volume of data generated by EOS satellites and associated remote sensing instruments presents unprecedented processing and distribution challenges.

"If NASA's EOS project team created a different format for every instrument's data, they would have a 'Tower of Babel' situation on their hands," said Mike Folk, Senior Software Engineer and Manager of the HDF Group at the National Center for Supercomputing Applications (NCSA).

After conducting an extensive study of available formats, EOSDIS planners selected the Hierarchical Data Format (HDF) as the standard for Earth science data generated by EOS instruments. Initially developed in 1987 at the NCSA, HDF is a physical file format for storing scientific data. It features a collection of tools for writing, manipulating, viewing, and analyzing data across diverse computing platforms.

Earth-orbiting satellites transmit data at regular intervals to properly equipped ground stations. As raw instrument data are received, computers at the ground station translate them into scientific parameters, such as sea surface temperature and cloud classification, that are useful to researchers.

Computers understand the numeric language of binary numbers, or number sets consisting of 0s and 1s that determine which areas of an image should be dark and which should be bright. A file format is a way of organizing the 0s and 1s so that they can be stored and retrieved in a standard way.

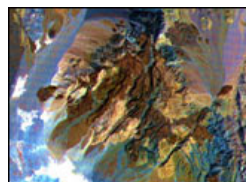
Distributing data in a standard file format has distinct advantages. First, a standard format ensures that researchers can access and easily combine and compare data from different instruments, facilitating cross-disciplinary collaboration.

**"If NASA's EOS project team created a different format for every instrument's data, they would have a 'Tower of Babel' situation on their hands."**



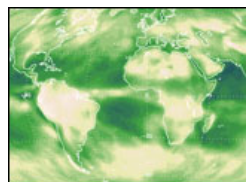
Atmospheric vortices seen by MISR.

[Click here for larger image.](#)



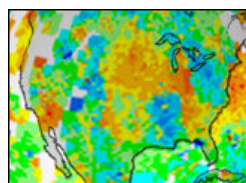
Prospecting from space using ASTER.

[Click here for larger image.](#)



First monthly CERES global longwave and shortwave radiation.

[Click here for larger image.](#)



MOPITT view of North America.

[Click here for larger image.](#)

Working with a single file format also has cost advantages. "By using one standard format, NASA has been able to put more of its resources into building tools, rather than into training people at the Distributed Active Archive Centers (DAACs) to be expert users of many different formats," said Larry Klein, project manager at Emergent Information Technologies, Inc. (EITI), the NASA contractor hired to develop and support HDF-EOS.

Although HDF meets many NASA specifications for accessing data, EOS applications required additional conventions and data types, which led to the development of HDF-EOS. "Our job was to develop a standard format for all data generated by instruments on EOS satellites," said Klein.

But developing a format to fit so many different data types has inherent problems. According to Klein, standardizing the way geolocation and temporal information is stored in files, a critical step for applications like re-projecting data, presents major difficulties.

HDF-EOS employs standard HDF objects, including images, tables, text, and data arrays. It also defines three additional data types based on HDF objects: grid, point, and swath. These data types allow the file contents to be referenced to Earth coordinates, such as latitude and longitude, and to time. Grid data types place the data on grids using one of many standard projections. Swath data types represent data that is ordered in time. Point data structures represent data that are irregularly spaced in time and geolocation, such as weather station data or instrument measurements taken on buoys.

HDF-EOS files are self-describing, aiding scientific data processing. For each data object in an HDF file, predefined tags identify information such as the data type, dimensions, and the locations within the file. The self-describing capability makes it possible to fully understand the structure and contents of a file from the information stored in the file alone. "If the information is packaged in another piece of documentation, such as a README file, it may get separated from the actual data file," said Klein. "In the history of packaging scientific data, that separation risk has always been a major problem."

Despite the obvious advantages, however, resistance to HDF-EOS has stalled universal acceptance of the standard. The results of a 1999 user-preferences survey, conducted by the National Snow and Ice Data Center (NSIDC) for the Geoscience Laser Altimeter System (GLAS) project, indicate that about 50 percent of users prefer a flat binary format to HDF-EOS. A flat binary file is typically compact and can enable users to access data more quickly than a standard data format. "Some scientists have been using their own formats for a long time," said Klein. "They don't want to pay someone to write their data in a new format."

"Because of all the capabilities it offers, HDF is a complicated file format," said Folk. "Many scientists are used to getting data in flat binary files, so there is some resistance to working with a large, complex package."

According to Siri Jodha Singh Khalsa, ECS Science Outreach Liaison at the NSIDC DAAC, utilities included with the HDF software package can export data in an HDF-EOS file to a simple, binary format, which can then be accessed using any visualization software package. "The resistance is there if you don't already have the HDF software and you're used to reading data in a different format," said Khalsa. "But for the new data user, there seems to be a preference for HDF."



MODIS view of the Middle East. [Click here for larger image.](#)

For more information, visit:

[NASA's Distributed Active Archive Center Alliance](#)

[Goddard Space Flight Center DAAC \(now named the GSFC Earth Sciences DAAC\)](#)

[Oak Ridge National Laboratory DAAC National Snow and Ice Data Center DAAC](#)

[Langley Atmospheric Sciences Data Center DAAC](#)

[Jet Propulsion Laboratory DAAC \(now named the Physical Oceanography DAAC\)](#)

[EROS Data Center DAAC \(now named the Land Processes DAAC\)](#)  
(A new browser window will open for each.)

While resistance may stem partially from user reluctance to learn a new tool, some user objections represent legitimate concerns. First, the number of software programs that support HDF-EOS is limited. Since Terra was launched less than a year ago, archives are only now beginning to fill with data. "As the instrument calibration improves more users will appear, and the demand for applications will follow," said Klein. "But right now, there isn't much choice from commercial software companies."

Some users are concerned about the availability of tools that support HDF-EOS. Users need a spectrum of visualization and analysis capabilities, including browsing, viewing, image enhancement, mathematical and statistical analysis, graphics, and animation. "The number of tools available is still relatively small, simply because the program is not yet mature," said Klein.

Also, while most Terra data products are delivered in HDF-EOS format, some instrument data simply don't fit the three elements of HDF-EOS. For example, Clouds and the Earth's Radiant Energy System (CERES) instruments generate data that are stored in basic HDF format. "As a development team, we designed HDF-EOS to fit most of the data generated by the EOS satellites," said Klein. "But it's not a 'one size fits all' situation."

Some users also have problems converting older data into the HDF-EOS format. "The format can't accommodate all of the older data types," said Klein. "In some cases it works; in other cases, it's just not a good fit."

Future releases of HDF-EOS will address some of these user issues. A new version of HDF-EOS, released in September 2000, supports Terra and Aqua satellite data. "We didn't make any significant changes in the new version, but we added some additional functionality to support the newer instruments," said Klein. In January 2001, EITI plans to release a new version of HDF-EOS based on HDF 5, an updated version of HDF that contains a new user interface, simpler data packaging, and provisions for parallel processing.

But with new versions come new issues. Since HDF-EOS is currently based on the HDF 4 library, the release of HDF 5 means that users will now be working with both HDF 4-based files and HDF 5-based files. "We continue to introduce new functionality to the format, which then introduces a whole new generation of bugs," said Klein. "We have to be certain the system can handle both versions and that the interfaces work together."

According to Khalsa, both NCSA and NASA have implemented measures to address user issues. NCSA maintains an online help desk, and users can also subscribe to two user group list servers that provide answers to frequently asked questions. In addition, user services staff at the NASA DAACs responds to discipline-specific questions.

NASA's Earth Sciences Data and Information Systems (ESDIS) Project, responsible for providing users with access to EOS data, also offers annual HDF/HDF-EOS workshops. The fourth annual workshop was held September 19-21, 2000 at Raytheon's facility in Landover, MD. The workshops included educational sessions, hands-on training using HDF and HDF-EOS tools and utilities, a question and answer session led by an experts panel, and presentations by HDF-EOS users, software developers, and software vendors.

Terry Haran, Software Engineer at NSIDC, works with data in both flat binary files and HDF-EOS format. "There are pros and cons to both formats, but it's

important for scientists to be able to combine data from different sensors without having to write new programs, and that's the chief advantage of HDF-EOS," he said.

Although getting past the initial learning curve may prove a bit bumpy for some, the universal participation of software developers and users is key to supporting the HDF-EOS standard. "Any problems related to HDF-EOS stem mostly from the fact that it's an evolving standard. Bugs don't really get shaken out until a lot of data have been produced, and a lot of users are working on them," said Khalsa.