

INTRODUCING THE HDF GROUP

The HDF Group is a non-profit 501(c)(3) company whose mission is to ensure the long-term accessibility of HDF data through sustainable development and support of HDF technologies.

Originally formed in 1987 as a working group at the National Center for Supercomputing Applications (NCSA) at the University of Illinois in Urbana-Champaign, the goal of the group was to develop a new file format and associated software library to move large, complex scientific data among disparate computing systems. The file format was called “Hierarchical Data Format”, or HDF, and its use quickly spread beyond NCSA. In 1998, HDF5 was introduced as a successor to the original format, now called “HDF4.” To better serve its growing user base, and to satisfy their requirements for long-term access to HDF-stored data, The HDF Group began full operations independent of NCSA and the University of Illinois in July 2006.

HDF technologies are essential for groups in academia, government, and industry, with over 200 different types of applications using HDF5 alone. The variety of data stored in the HDF formats and controlled through the HDF APIs is an indication of HDF’s flexibility and scalability. Examples of HDF applications include sensor data acquisition, management, and distribution; image interchanges and repositories; scalable computational meshes; high-performance data storage and retrieval on massively parallel systems; containers for heterogeneous collections of complex data; object stores for object relational databases; long-term archives; and special effects for “The Lord of the Rings.”

The adoption of HDF by a number of mission-critical projects is testimony to the quality and performance of the HDF technologies, as well as to The HDF Group’s commitment to supporting its customers. In the early 1990s, NASA’s Earth Observing System—the primary data gathering system for global climate research—selected HDF after investigating 15 different formats over a two-year period. The Department of Energy National Laboratories at Livermore, Sandia, and Los Alamos sponsored the development of HDF5 in 1998, and depend on the format for data management in many critical applications. Government organizations such as the Aberdeen Test Center rely on HDF for managing large complex data repositories. Companies such as Boeing use HDF as an internal standard for data processing, exchange, and archiving. Research institutions such as The National Center for Macromolecular Imaging (NCMI) use HDF for managing large multidimensional images. In 2006, MATLAB® began using an HDF5-based format to save data over 2 gigabytes in size, and has announced it will make that format the default MAT-file format in a future release.