**The Hierarchical Data Format (HDF): A Data Foundation for Sustainable Software in Many Communities**

Sustainable science depends on intricate relationships between data and tools that access and analyze them. Software that cannot read data and data that can’t be read are both significant obstacles to sustainability. In the long-term, data must be preserved in well documented, self-describing formats that are accessible on multiple hardware platforms using a wide variety of programming languages. In addition, the formats must include mechanisms for including metadata that are required for using and understanding the data long after they are collected and processed.

In the short-term, general commercial, open-source, and community tools must support data storage and analysis needs of multiple communities. In addition, sustainable formats must address specific data needs of multiple scientific communities in order to achieve the breadth of usage and support required for on-going maintenance and development of new capabilities. In practice, this breadth of usage is achieved through the development of conventions that map community science data types to data objects in the file (computer science data types). These conventions ensure interoperability by providing consistency in the data layer and isolating science users from the details of the underlying storage organization and format.

The Hierarchical Data Format (HDF) addresses the needs described above and forms a foundation for data storage and access in many scientific communities. The flexibility of the format is achieved through simplicity. HDF is a container format that includes three fundamental objects: datasets, attributes, and groups. Datasets are multi-dimensional arrays of atomic or compound datatypes (computer science datatypes). They provide the flexibility required for storage of diverse scientific data objects while attributes allow annotation of these objects with metadata required for use and understanding. Groups provide overall organization of datasets and attributes into associations that define scientific data objects.

When combined with community specific conventions, HDF provides a robust and reliable foundation that scientific communities rely on for interoperability and high-performance storage of metadata and data. This combination naturally leads to a division of labor between the developers and maintainers of the format (The HDF Group) and the diverse communities that build conventional formats on top of HDF. This paper focuses on practices and experiences of the second group that will help others understand their approach and build on their success.