store the files instead of the original CDF Fortran method of using multiple files for one dataset on a VAX. CDF is not a good, transportable file format, but a very interesting calling interface for storage of scientific data. Its developers are reportedly working on methods for storing more than one array element at a time. There are only eight data types so far. These are the data types defined in VAX Fortran.

HDF can support many calling interfaces and the CDF interface may be a good one to support, for it has both C and Fortran interfaces.

FITS—Flexible Image Transfer System

The FITS file format is primarily used for astronomy; it was one of the first tagged formats. Each tag is an ASCII string in the header record. The data can be of several types and is stored in multi-dimensional arrays. The file is in Fortran records of a fixed length. For Fortran-based systems that store only scientific data, FITS is still applicable.

PHIGS and PHIGS+—Programmer's Hierarchical Integrated Graphics System

PHIGS and PHIGS+ are being adopted as standards for a calling interface to be used by high performance graphics programs. Most manufacturers of expensive graphics equipment will be providing PHIGS interfaces. They have not, as yet, defined a file storage format for PHIGS calls, so you have to send the source code with your image. NCSA is interested in adopting a PHIGS-compatible vector/polygon standard in HDF.

Chapter Overview

This chapter briefly describes the following file formats and compares their features with those of HDF: CGM, PICT2, TIFF, CDF, FITS, PHIGS, and PHIGS+.

File Formats

There are numerous data formats in use today, some designed with objectives very similar to those of HDF. We encounter many of these at NCSA, and expect to continue to work with them in the future. The following sections describe some of the important data formats in use and compare their usability to that of HDF.

CGM—Computer Graphics Metafile

CGM is an ANSI standard format for storing graphics in files, specifically graphics created with GKS libraries. It covers simple vectors, polygons, and raster images up to 24 bits deep. Many computing sites are using CGM, but they are quickly finding its limitations. For example, CGM does not allow simulation source data and an image of that data to be in the same file. Some implementations of CGM are not compatible with others, so there is now a SLATEC document that explicitly defines a subset of CGM for all DOE and NSF sites to implement.

PICT2—Macintosh Standard

The PICT2 format is extensible and already contains most of the primitives that are needed, including a one-to-one mapping with QuickDraw. Two problems with PICT2 are (1) it is a serial format and (2) it is proprietary. The Macintosh disk format includes what are called *resources*; the resource types are similar to HDF's tags, assigned for each new need that arises. This also is proprietary and is a difficult standard to use because MacOS defines most of its uses for operating system purposes, not generic data storage.

TIFF—Tag Image File Format

Microsoft and Aldus developed the TIFF format, which is very similar to HDF. TIFF defines a linked list of tag blocks in the file, with each tag block describing one image. There is no mechanism in place to help the standard grow in a regular fashion. One of the side effects is that each block of tags describes one frame which ties the file structure in with the data structure. The developers of TIFF defined only most of the tags necessary for scanner companies and then stopped. The irregular use of this standard by commercial companies makes it a moving target.

CDF—Common Data Format

This calling interface and format was developed at NASA for use with scientific datasets. It defines all datasets as multi-dimensional arrays of multiple variables and allows you to include any number of single variables. The NCAR UNIDATA project has implemented the Fortran interface and adapted it for use in C. At NCAR, they use XDR (from Sun) to

Chapter 5 Other Formats

Chapter Overview

File Formats

- CGM—Computer Graphics Metafile
- PICT2—Macintosh Standard
- TIFF—Tag Image File Format
- CDF—Common Data Format
- FITS—Flexible Image Transfer System
- PHIGS and PHIGS+—Programmer's Hierarchical Integrated Graphics System